A Review of Hydrodynamic and Energy-Transport Models for Semiconductor Device Simulation

TIBOR GRASSER, TING-WEI TANG, FELLOW, IEEE, HANS KOSINA, MEMBER, IEEE, AND SIEGFRIED SELBERHERR, FELLOW, IEEE

Contributed Paper

Since Stratton published his famous paper four decades ago, various transport models have been proposed which account for the average carrier energy or temperature in one way or another. The need for such transport models arose because the traditionally used drift-diffusion model cannot capture nonlocal effects which gained increasing importance in modern miniaturized semiconductor devices. In the derivation of these models from Boltzmann's transport equation, several assumptions have to be made in order to obtain a tractable equation set. Although these assumptions may differ significantly, the resulting final models show various similarities, which has frequently led to confusion. We give a detailed review on this subject, highlighting the differences and similarities between the models, and we shed some light on the critical issues associated with higher order transport models.

Keywords*—Numerical analysis, semiconductor device modeling, simulation.*

I. INTRODUCTION

At the very beginnings of semiconductor technology, the electrical device characteristics could be estimated using simple analytical models relying on the drift-diffusion (DD) formalism. Various approximations had to be made to obtain closed-form solutions, but the resulting models captured the basic features of the devices. These approximations include simplified doping profiles and device geometries. With the ongoing refinements and improvements in technology, these approximations lost their basis and a more accurate description was required. This goal could be achieved by solving the DD equations numerically. Numerical simulation of carrier transport in semiconductor devices dates back to the famous work of Scharfetter and Gummel [1], who proposed a robust discretization of the DD equations which is still in use today.

However, as semiconductor devices were scaled into the submicrometer regime, the assumptions underlying the DD model lost their validity. Therefore, the transport models have been continously refined and extended to more accurately capture transport phenomena occurring in these submicrometer devices. The need for refinement and extension is primarily caused 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 field inside the devices has increased. A large electric field which rapidly changes over small length scales gives rise to nonlocal and hot-carrier effects which begin to dominate device performance. An accurate description of these phenomena is required and is becoming a primary concern for industrial applications.

To overcome some of the 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. However, a vast number of these models exists, and there is a considerable amount of confusion as to their relation to each other. The aim of this paper is to clarify the important differences and similarities between the various models. To master this task, a closer look at the various derivations is required, where the important points are highlighted. Then the most important models are summarized and assessed, followed by a critical discussion of the assumptions made in the derivation.

Manuscript received January 3, 2002; revised July 2, 2002. This work was supported in part by the Christian Doppler Forschungsgesellschaft, Vienna, Austria, and in part by the National Science Foundation under Grant ECS-0120128.

T. Grasser, H. Kosina, and S. Selberherr are with the Institute for Microelectronics, Technical University Vienna, A-1040 Vienna, Austria (e-mail: grasser@iue.tuwien.ac.at).

T.-W. Tang is with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003 USA (e-mail: ttang@ecs.umass.edu).

Digital Object Identifier 10.1109/JPROC.2002.808150

II. The Drift-Diffusion Model

The DD model is the simplest current transport model which can be derived from Boltzmann's transport equation (BTE) by the method of moments [4] or from basic principles of irreversible thermodynamics [5]. For many decades, the DD model has been the backbone of semiconductor device simulation. In this model, the electron current density is phenomenologically expressed as consisting of two components. The drift component is driven by the electric field and the diffusion component by the electron density gradient. It is given by

$$\mathbf{J} = q\left(n\mu\mathbf{E} + D_n\nabla n\right) \tag{1}$$

where μ and D_n are the mobility and the diffusivity of the electron gas, respectively, and are related to each other by the Einstein relation

$$D_n = \frac{k_B T_n}{q}\mu \tag{2}$$

where k_B is the Boltzmann constant. The current relation (1) is inserted into the continuity equation

$$\nabla\cdot\mathbf{J} = q\partial_t n \tag{3}$$

to give a second-order parabolic differential equation, which is then solved together with Poisson's equation. Note that generation/recombination effects were neglected in (3).

In the DD approach, the electron gas is assumed to be in thermal equilibrium with the lattice temperature ($T_n = T_L$). However, in the presence of a strong electric field, electrons gain energy from the field and the temperature of the electron gas T_n is elevated. Since the pressure of the electron gas is proportional to nk_BT_n, the driving force now becomes the pressure gradient rather then merely the density gradient. This introduces an additional driving force, namely, the temperature gradient besides the electric field and the density gradient. Phenomenologically, one can write

$$\mathbf{J} = q\left(n\mu\mathbf{E} + D_n\nabla n + nD_T\nabla T_n\right) \tag{4}$$

where D_T is the thermal diffusivity.

Although the DD equations are based on the assumption that the electron gas is in thermal equilibrium with the lattice, an estimation for the local temperature can be calculated with the local energy balance equation [6]

$$T_n = T_L + \frac{2}{3}\frac{q}{k_B}\tau_{\mathcal{E}}\mu\mathbf{E}^2 = T_L\left(1 + \left(\frac{E}{E_c}\right)^2\right) \tag{5}$$

where $\tau_{\mathcal{E}}$ is the energy relaxation time. Equation (5) is obtained under the assumption of a local energy balance. At the critical electric field E_c, which depends on the electric field via the mobility, the carrier temperature reaches twice the value of the lattice temperature. E_c is in the order of 10 kV/cm, a value easily exceeded even in not very advanced devices where values higher than 1 MV/cm can be observed [7]. Note too that the temperature obtained from (5) introduces an inconsistency with the assumptions made during the derivation of the DD model where the electron gas has been assumed to be in equilibrium with the lattice temperature.

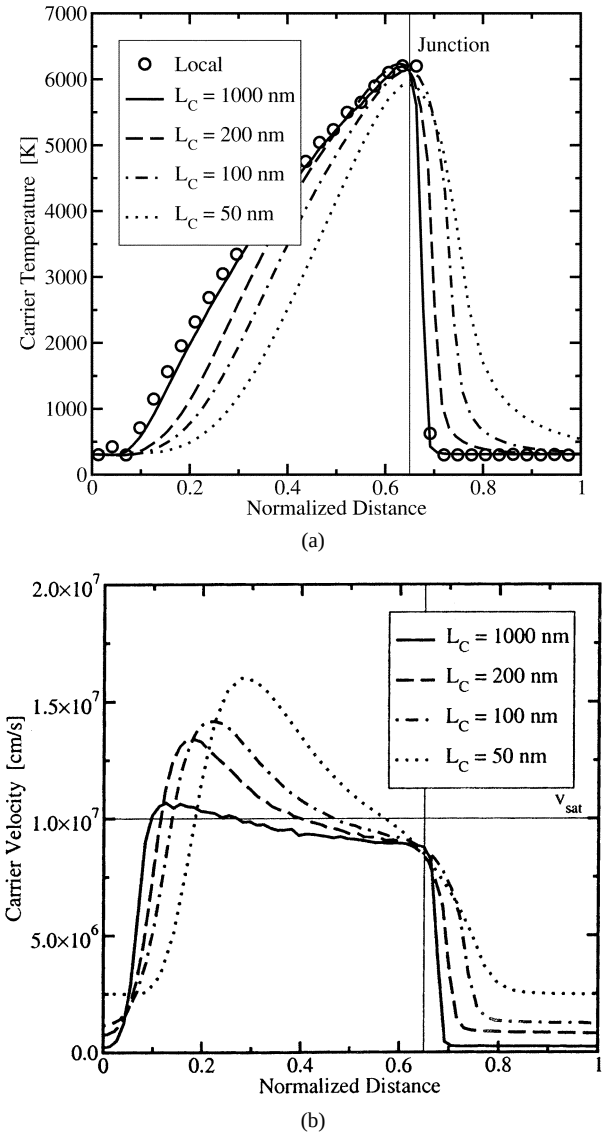


Fig. 1 (a) The carrier temperature of comparable n^+- n - n^+ structures with varying channel lengths where the spatial coordinates have been normalized to get an overlapping electric field. (b) The average carrier velocity where the velocity overshoot is caused by the nonlocality of the carrier temperature.

For a rapidly increasing electric field, however, the average energy lags behind the electric field, and the assumption of local equilibrium becomes invalid [8]. A consequence of the lag is that the maximum energy can be considerably smaller than the one predicted by the local energy balance equation. This nonlocality of the carrier temperature is shown in Fig. 1(a) for n^+- n - n^+ structures with varying channel lengths where the spatial coordinate has been normalized to make the electric field distribution of all devices overlap. The bias has been chosen to give a maximum electric field of 300 kV/cm in all devices. An important consequence of this behavior is that the lag of the average energy gives rise to an overshoot in the carrier velocity, as shown in Fig. 1(b).

Also shown is the saturation velocity v_{sat}, which is the maximum velocity observed in stationary bulk measurements. The reason for the velocity overshoot is that the mobility depends to first order on the average energy and not on the electric field. As the mobility μ 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.

Similar to the carrier mobility, many other physical processes like impact ionization are more accurately described by a local energy model rather than a local electric field model, because these processes depend on the distribution function rather than on the electric field.

Altogether, it can be noted that modeling of deep-submicrometer devices with the DD model is becoming more and more problematic. Although successful reproduction of terminal characteristics of nanoscale MOS transistors has been reported with the DD model [9], the values of the physical parameters used significantly violate basic physical principles. In particular, the saturation velocity v_{sat} had to be set to more than twice the value observed in bulk measurements. This implies that the model is no longer capable of reproducing the results of bulk measurements and as such looses its consistency. Furthermore, the model can hardly be used for predictive simulations. These solutions may provide short-term fixes to available models, but obtaining "correct" results from the wrong physics is unsatisfactory in the long run.

In the following, the derivation and properties of higher order transport models from Boltzmann's equation is given.

III. Boltzmann's Transport Equation

Transport equations used in semiconductor device simulation are normally derived from the BTE, a semiclassical kinetic equation, which reads [10]

$$\partial_t f + \mathbf{u}\cdot\nabla_{\mathbf{r}} f + \frac{\mathbf{F}}{\hbar}\cdot\nabla_{\mathbf{k}} f = C[f] \tag{6}$$

where $f(\mathbf{k},\mathbf{r},t)$ represents the carrier distribution function in the six-dimensional phase space, and the term on the right side represents the rate of change of f due to collisions. The BTE is valid for general inhomogeneous materials with arbitrary band structure [11]. To account for quantum effects, equations based on the Wigner–Boltzmann equation have been considered [12]. The group velocity $\mathbf{u}$ is defined as

$$\mathbf{u}(\mathbf{k},\mathbf{r}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k},\mathbf{r}) \tag{7}$$

where $\mathcal{E}$ represents the carrier kinetic energy. The inverse effective mass tensor is defined as

$$\hat{m}^{-1}(\mathbf{k},\mathbf{r}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\otimes\mathbf{u}(\mathbf{k},\mathbf{r}) = \frac{1}{\hbar^2}\nabla_{\mathbf{k}}\otimes\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k},\mathbf{r}) \tag{8}$$

where $\otimes$ denotes the tensor product [11]. The force $\mathbf{F}$ exerted on electrons in the presence of electric and magnetic fields and inhomogeneous material properties is generally given as

$$\mathbf{F}(\mathbf{k},\mathbf{r}) = -\nabla_{\mathbf{r}}\left(E_{c,0}(\mathbf{r}) + \mathcal{E}(\mathbf{k},\mathbf{r})\right) - q\left(\mathbf{E}(\mathbf{r}) + \mathbf{u}(\mathbf{k},\mathbf{r})\times\mathbf{B}(\mathbf{r})\right) \tag{9}$$

and depends both on $\mathbf{k}$ and $\mathbf{r}$. The two position-dependent terms $(\nabla_{\mathbf{r}})$ account for changes in the bottom of the conduction band edge $E_{c,0}$ and the shape of the band structure. Omitting the influence of the magnetic field (see [6] for a treatment of this term) and assuming homogeneous materials, $\mathbf{F}$ simplifies to the electrostatic force

$$\mathbf{F}(\mathbf{r}) = -q\mathbf{E}(\mathbf{r}). \tag{10}$$

In the following, we will consider only position-independent masses, but permit energy-dependent masses. Generalizations to position-dependent band structures will be given in Section VIII.

The BTE represents an integro-differential equation in the seven-dimensional space $(\mathbf{k},\mathbf{r},t)$. To solve this equation numerically by discretization of the differential and integral operators is computationally very expensive. A widely used numerical method for solving the BTE is the Monte Carlo (MC) method. This method has been proven to give accurate results, but is still computationally expensive. 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. Another approach, which is based on an expansion of the distribution function in momentum space into a series of spherical harmonics, has been successfully used to solve the BTE [13], [14]. In contrast to the MC method, the spherical harmonic expansion method is deterministic, and the computational effort, while still high, is significantly reduced. However, by considering only the lower order terms of the expansion, approximations are introduced whose influence on the accuracy of the simulation results is still not fully clarified. Particularly in the ballistic regime, numerically calculated full-band structures cannot be included, as opposed to the MC method, where usage of such band structures is a solved problem.

A common simplification, which will be the subject of this paper, is to investigate only some moments of the distribution function, such as the carrier concentration and the carrier temperature. A moment is obtained by multiplying the distribution function with a suitable weight function $\Phi = \Phi(\mathbf{k})$ and integrating over $\mathbf{k}$ space

$$\langle\Phi\rangle = \frac{1}{4\pi^3}\int \Phi f d^3\mathbf{k}. \tag{11}$$

Thus, three coordinates are eliminated at the expense of information loss concerning the details of the distribution function.

In the following, we will separate the group velocity $\mathbf{u}$ into a random part $\mathbf{c}$ and the mean value $\mathbf{v} = \langle\mathbf{u}\rangle/\langle 1\rangle$ as $\mathbf{u} = \mathbf{c} + \mathbf{v}$. We will assign the symbols given in Table 1 to

the moments of the distribution function [15]. Furthermore, we will employ an isotropic effective mass approximation via the trace of the mass tensor [16]

$$m^{*-1} = \frac{1}{3}\mathrm{tr}\langle \hat{m}^{-1}\rangle. \tag{12}$$

IV. Band Structure

A common assumption in macroscopic transport models is that the band structure is isotropic; that is, the kinetic energy depends only on the magnitude of the wave vector $\mathbf{k}$. With this assumption, the dispersion relation can be written in terms of the band form function γ

$$\gamma(\mathcal{E}) = \frac{\hbar^2 k^2}{2m^*}. \tag{13}$$

The simplest approximation for the real band structure is a parabolic relationship between the energy and the carrier momentum $\hbar\mathbf{k}$

$$\gamma(\mathcal{E}) = \mathcal{E} \tag{14}$$

which is assumed to be valid for energies close to the band minimum. A first-order nonparabolic relationship was given by Kane [17]

$$\gamma(\mathcal{E}) = \mathcal{E}(1+\alpha\mathcal{E}) \tag{15}$$

with α being the nonparabolicity correction factor. Kane's dispersion relation gives the following relationship between momentum and velocity:

$$\mathbf{u} = \frac{\hbar\mathbf{k}}{m^*}\frac{1}{1+2\alpha\mathcal{E}} = \frac{\hbar\mathbf{k}}{m^*}\sum_{i=0}^{\infty}(-2\alpha\mathcal{E})^i. \tag{16}$$

Therefore, the average velocity contains an infinite number of higher order terms which are not necessarily negligible. This is problematic because these quantities are additional unknowns representing higher order moments of the velocity distribution which prohibits closed-form solutions.

To obtain a more tractable expression, Cassi and Riccò [18] approximated Kane's dispersion relation as

$$\gamma(\mathcal{E}) = x\mathcal{E}^y \tag{17}$$

and fitted the parameters x and y for different energy ranges. For $y = 1$, the conventional parabolic dispersion relation is obtained. As pointed out in [19], this expression must be used with care. In particular, physically meaningful results could be obtained only by fitting (17) to the energy range $[0, 0.2\,\mathrm{eV}]$. This can be explained by looking more closely at the density of states, which is obtained as

$$g(\mathcal{E}) = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\gamma^{1/2}\frac{d\gamma}{d\mathcal{E}} \tag{18}$$

and in the particular case of Cassi's model

$$g(\mathcal{E}) = \frac{8\pi}{h^3}(2m^*x)^{3/2}y\mathcal{E}^{(3/2)y-1} = g_0\mathcal{E}^{\lambda}. \tag{19}$$

A comparison of different fits to the Kane expression is shown in Fig. 2, together with the numerical density of states used by Fischetti and Laux [20]. The fitted values were taken from [19], and are $x = 1.365$ and $y = 1.52$ when fitted to the low-energy range $[0, 0.2\ \mathrm{eV}]$ and $x = 1.185$ and $y = 1.052$ when fitted to the high-energy range $[1.5, 3.0\ \mathrm{eV}]$, where x has the unusual dimension of eV^{1-y}. For $\lambda = 1$, the shape of $g(\mathcal{E})$ changes from convex to concave; thus, either the low- or the high-energy range can be fitted properly, but not both simultaneously. As the dispersion relation is needed for the evaluation of the moments of the distribution function, which requires an integration over the whole energy range, a value smaller than one is required for λ to accurately fit

Table 1
Definition of the Most Important Quantities

Quantity	Symbol	Definition
carrier density	n	$\langle 1\rangle$
average crystal momentum	$\mathbf{p}$	$\frac{1}{n}\langle\hbar\mathbf{k}\rangle$
average velocity	$\mathbf{v}$	$\frac{1}{n}\langle\mathbf{u}\rangle = -\frac{\mathbf{J}}{\mathrm{q}n}$
average energy	w	$\frac{1}{n}\langle\mathcal{E}\rangle$
energy flux	$\mathbf{S}$	$\frac{1}{n}\langle\mathbf{u}\mathcal{E}\rangle$
heat flux	$\mathbf{Q}$	$\frac{1}{2n}\langle m(\mathbf{k})\mathbf{c}^2\mathbf{c}\rangle$
temperature tensor	$\hat{\mathrm{T}}$	$\frac{1}{\mathrm{k_B}n}\langle m(\mathbf{k})\mathbf{c}\otimes\mathbf{c}\rangle$
energy tensor	$\hat{\mathrm{U}}$	$\frac{1}{n}\langle\hbar\mathbf{u}\otimes\mathbf{k}\rangle$
fourth order tensor	$\hat{\mathrm{R}}$	$\frac{1}{n}\langle\hbar\mathbf{u}\otimes\mathbf{k}\mathcal{E}\rangle$
group velocity	$\mathbf{u}$	$\frac{1}{\hbar}\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k},\mathbf{r})$
inverse effective mass tensor	$\hat{\mathrm{m}}^{-1}$	$\frac{1}{\hbar^2}\nabla_{\mathbf{k}}\otimes\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k},\mathbf{r})$

Bold symbols denote vector quantities, whereas tensors are written as $\hat{T}$.

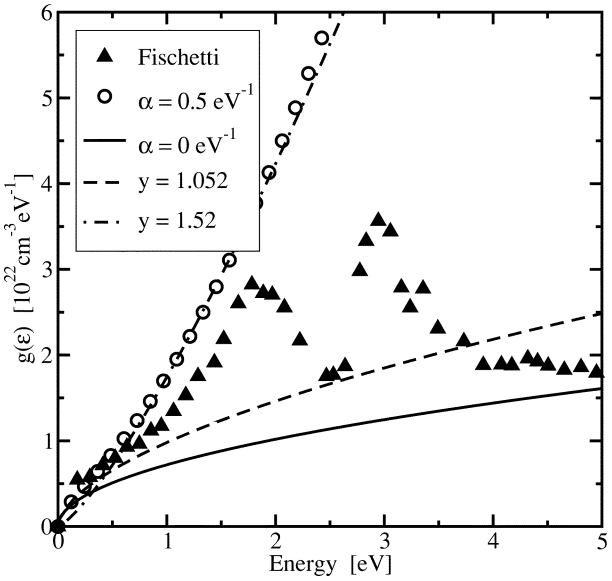


Fig. 2 Comparison of different expressions for the density of states.

the low-energy region because this is where $f(\mathcal{E})$ has its maximum. However, the resulting density of states shows a "parabolic-like" behavior; therefore, it is of limited value for the description of nonparabolic transport phenomena.

V. Derivation of Higher Order Transport Models

In this section, the basic steps for transforming BTE into a macroscopic transport model will be reviewed to point out the principal differences among the models. Generation and recombination processes are not considered in the following.

A. Stratton's Approach

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

$$f(\mathbf{k},\mathbf{r}) = f_0(\mathbf{k},\mathbf{r}) + f_1(\mathbf{k},\mathbf{r}). \tag{20}$$

From $f_1(-\mathbf{k},\mathbf{r}) = -f_1(\mathbf{k},\mathbf{r})$, 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

$$C[f] = -\frac{f - f_{\text{eq}}}{\tau(\mathcal{E},\mathbf{r})} \tag{21}$$

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

$$f_1 = -\tau(\mathcal{E},\mathbf{r})\left(\mathbf{u}\nabla_{\mathbf{r}} f_0 - \frac{q}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}} f_0\right). \tag{22}$$

The microscopic relaxation time is then expressed by a power law

$$\tau(\mathcal{E}) = \tau_0\left(\frac{\mathcal{E}}{k_B T_L}\right)^{-p}. \tag{23}$$

When f_0 is assumed to be a heated Maxwellian distribution, the following equation system is obtained:

$$\nabla\cdot\mathbf{J} = q\partial_t n \tag{24}$$

$$\mathbf{J} = q\mu n\mathbf{E} + k_B\nabla(n\mu T_n) \tag{25}$$

$$\nabla\cdot(n\mathbf{S}) = -\frac{3}{2}k_B\partial(nT_n) + \mathbf{E}\cdot\mathbf{J} - \frac{3}{2}k_B n\frac{T_n - T_L}{\tau_{\mathcal{E}}} \tag{26}$$

$$n\mathbf{S} = -\left(\frac{5}{2} - p\right) \times\left(\mu n k_B T_n\mathbf{E} + \frac{k_B^2}{q}\nabla(n\mu T_n)\right) \tag{27}$$

Equation (25) can be rewritten as

$$\mathbf{J} = q\mu\left(n\mathbf{E} + \frac{k_B}{q}T_n\nabla n + \frac{k_B}{q}n(1+\nu_n)\nabla T_n\right) \tag{28}$$

with

$$\nu_n = \frac{T_n}{\mu}\frac{\partial\mu}{\partial T_n} = \frac{\partial\ln\mu}{\partial\ln T_n} \tag{29}$$

which is commonly used as a fit parameter with values in the range $[-0.5,-1.0]$. For $\nu_n = -1.0$, the thermal diffusion term disappears. Under certain assumptions [2], [21], the coefficient p equals $-\nu_n$. The problem with expression (23) for τ is that p must be approximated by an average value to cover the relevant scattering processes. In the particular case of impurity scattering, p can be in the range $[-1.5, 0.5]$, depending on charge screening [22]. Therefore, this average depends on the doping profile and the applied field; thus, no unique value for p can be given.

Note that the temperature T_n which appears in (24)–(27) is a parameter of the heated Maxwellian distribution, which has been assumed in the derivation. Only for parabolic bands and a Maxwellian distribution, this parameter is equivalent to the normalized second-order moment.

B. Bløtekjær's Approach

Bløtekjær [3] derived conservation equations by taking the moments of the BTE using the weight functions one, $\hbar\mathbf{k}$, and $\mathcal{E}$ without imposing any assumptions on the form of the distribution function. These weight functions Φ define the moments of zeroth, first, and second order. The resulting moment equations can be written as follows [15]:

$$\partial_t n + \nabla\cdot(n\mathbf{v}) = nC_n \tag{30}$$

$$\partial_t(n\mathbf{p}) + \nabla\cdot(n\hat{U}) - n\mathbf{F} = n\mathbf{C}_p \tag{31}$$

$$\partial_t(nw) + \nabla\cdot(n\mathbf{S}) - n\mathbf{v}\cdot\mathbf{F} = nC_{\mathcal{E}}. \tag{32}$$

These expressions are valid for arbitrary band structures, provided that the carrier mass is position independent. When $\mathbf{F}$ is allowed to be position dependent, additional force terms appear in (30)–(32) [23]. The collision terms are usually modeled employing a macroscopic relaxation time approximation as

$$C_n = 0 \tag{33}$$

$$\mathbf{C}_p = -\frac{\mathbf{p}}{\tau_p} \tag{34}$$

$$C_{\mathcal{E}} = -\frac{w - w_0}{\tau_{\mathcal{E}}} \tag{35}$$

which introduces the momentum and energy relaxation times τ_p and $\tau_{\mathcal{E}}$, respectively. A discussion on this approximation is given in [24]. One should note the difference between the *microscopic* relaxation time approximation as used by Stratton, where the whole scattering operator is approximated by a single relaxation time, and the *macroscopic* relaxation time approximation, where a separate relaxation time is introduced for every moment of the scattering operator. The latter is assumed to be more accurate.

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. Traditionally, parabolic bands are assumed, which gives the following closure relations for $\mathbf{p}$, $\hat{U}$, and w:

$$\mathbf{p} = m^*\mathbf{v} \tag{36}$$

$$\hat{U} = \frac{m^*}{n}\langle\mathbf{u}\otimes\mathbf{u}\rangle = \frac{m^*}{n}\langle\mathbf{c}\otimes\mathbf{c}\rangle + m^*\mathbf{v}\otimes\mathbf{v} \tag{37}$$

$$w = \frac{m^*}{2n}\langle\mathbf{u}^2\rangle = \frac{m^*}{2n}\langle\mathbf{c}^2\rangle + \frac{1}{2}m^* v^2. \tag{38}$$

The random component of the velocity has zero average, that is, $\langle \mathbf{c} \rangle = 0$. Under the assumption that the distribution function is isotropic, the following relation can be derived:

$$\langle \mathbf{c}^2 \rangle \hat{I} = \frac{1}{3} \langle \mathbf{c} \otimes \mathbf{c} \rangle. \tag{39}$$

This assumption is frequently considered to be justified because of the strong influence of scattering. Approximating $\hat{T}$ by a scalar temperature T_n as

$$\hat{T} \approx T_n \hat{I} = \frac{1}{3}(T_{xx} + T_{yy} + T_{zz})\hat{I} \tag{40}$$

(37) and (38) become coupled

$$\hat{U} = k_B T_n \hat{I} + m^* \mathbf{v} \otimes \mathbf{v} \tag{41}$$

$$w = \frac{3}{2} k_B T_n + \frac{1}{2} m^* v^2. \tag{42}$$

With (34) and (36), one obtains the following formulation for $\mathbf{C}_p$:

$$\mathbf{C}_p = -\frac{q\mathbf{v}}{\mu} \tag{43}$$

where m^* and τ_p have been lumped into one new parameter, the mobility

$$\mu = \frac{q\tau_p}{m^*}. \tag{44}$$

Relationship (44) is valid for parabolic bands only, whereas for arbitrary bands the mobility can be defined directly via the collision term. As signal frequencies are well below $1/(2\pi\tau_p) \approx 10^{12}$ Hz, the time derivative in (31) can safely be neglected [25].

Furthermore, a suitable approximation for the energy flux density $n\mathbf{S}$ has to be found, and different approaches have been published. Bløtekjær used

$$n\mathbf{S} = (w + k_B T_n) n\mathbf{v} + n\mathbf{Q} \tag{45}$$

and approximated the heat flux $n\mathbf{Q}$ by Fourier's law as

$$n\mathbf{Q} = -\kappa(T_n) \nabla T_n \tag{46}$$

in which the thermal conductivity is given by the Wiedemann–Franz law as

$$\kappa(T_n) = \left(\frac{5}{2} - p\right) \left(\frac{k_B}{q}\right)^2 q\mu n T_n \tag{47}$$

where p is a correction factor. The artificial introduction of $\mathbf{Q}$ was based on physical reasoning only because a shifted and heated Maxwellian distribution gives $\mathbf{Q} = 0$. This is a major inconsistency of the model. In addition, as has been pointed out in [15], this expression is problematic, as (46) only approximates the diffusive component of $n\mathbf{Q}$. For a uniform temperature, $\nabla T_n = 0$; thus, $\mathbf{Q} = 0$, which contradicts with MC simulations. The convective component $\mathbf{Q}_{\text{conv}}$ must be included to obtain physical results when the current flow is not negligible.

With these approximations, (30)–(32) can be written in the usual variables as [26]

$$\nabla \cdot \mathbf{J} = q\partial_t n \tag{48}$$

$$\mathbf{J} - \frac{\tau_p}{q} \nabla \cdot \left(\mathbf{J} \otimes \frac{\mathbf{J}}{n} \right) = \mu k_B \nabla (n T_n) + qn\mu \mathbf{E} - \tau_p \partial_t \mathbf{J} \tag{49}$$

$$\nabla \cdot (n\mathbf{S}) = -\partial_t (nw) + \mathbf{E} \cdot \mathbf{J} - n \frac{w - w_0}{\tau_{\mathcal{E}}} \tag{50}$$

$$n\mathbf{S} = -\frac{1}{q}(w + k_B T_n)\mathbf{J} - \kappa(T_n) \nabla T_n \tag{51}$$

to give the full hydrodynamic model (FHD) for parabolic band structures, which has been supplemented by the phenomenological constitutive relation (51) to close the system. This equation system is similar to the Euler equations of fluid 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 $v_c = \sqrt{k_B T_n / m^*}$, and the electron flow may be either subsonic or supersonic. With $T_n = \xi T_L$ and $T_L = 300$ K, $v_c = \sqrt{\xi} 1.3 \cdot 10^7$ cm/s, while for $T_L = 77$ K, $v_c = \sqrt{\xi} 6.6 \cdot 10^6$ cm/s [27]. 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 low temperatures. As the equation system is hyperbolic in the supersonic regions, special numerical methods have to be used (see Section XIII) which are not compatible to the methods employed for the parabolic convection-diffusion type of equations.

When the convective term in the current relation (49)

$$\frac{\tau_p}{q} \nabla \cdot \left(\mathbf{J} \otimes \frac{\mathbf{J}}{n} \right) \tag{52}$$

is neglected, a parabolic equation system is obtained which covers only 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, resulting in

$$w \approx \frac{3}{2} k_B T_n. \tag{53}$$

The two assumptions made previously can be justified from a mathematical point of view because they follow consistently from appropriately scaling the BTE. The Knudsen number appears as a scaling parameter, which represents the mean free path $\tau_0 v_0$ relative to the device dimension [28]

$$\lambda = \frac{\tau_0 v_0}{x_0} \tag{54}$$

where τ_0 is the characteristic time between scattering events, v_0 denotes the velocity scale, and x_0 is given by the size of the simulation domain. Carriers in a semiconductor at room temperature can be considered a collision-dominated system, for which $\lambda \ll 1$. Diffusion scaling assumes the time scale of the system to be

$$t_0 = \frac{\tau_0}{\lambda^2}. \tag{55}$$

In the limit of vanishing Knudsen number, $\lambda \to 0$, one obtains that convective terms of the form $\langle \mathbf{u} \rangle \otimes \langle \mathbf{u} \rangle$ are neglected against $\langle \mathbf{u} \otimes \mathbf{u} \rangle$. The consequences are that the drift

kinetic energy $m^*\langle \mathbf{u} \rangle^2/2$ is neglected against $k_B T_n$, and that in the flux equations the time derivatives vanish.

The resulting simplified equations are

$$\nabla \cdot \mathbf{J} = q\partial_t n \tag{56}$$

$$\mathbf{J} = \mu k_B \nabla(nT_n) + qn\mu\mathbf{E} \tag{57}$$

$$\nabla \cdot (n\mathbf{S}) = -\frac{3}{2}k_B\partial_t(nT_n) + \mathbf{E}\cdot\mathbf{J} - n\frac{3}{2}k_B\frac{T_n - T_L}{\tau_{\mathcal{E}}} \tag{58}$$

$$n\mathbf{S} = -\frac{5k_BT_n}{2q}\mathbf{J} - \kappa(T_n)\nabla T_n. \tag{59}$$

Equations (56) and (59) form a typical three-moment model which has been closed using Fourier's law and is commonly known in the literature as the energy-transport (ET) model. Actually, this name is misleading, as the model consists of the two conservation equations (56) and (58) and the two constitutive equations (57) and (59).

To overcome the difficulties associated with the Fourier law closure (46), the fourth moment of the BTE has been taken into account, [15] which gives

$$\nabla \cdot (n\hat{R}) - n(w\hat{I} + \hat{U}) \cdot \mathbf{F} = n\mathbf{C}_{p\mathcal{E}}. \tag{60}$$

The time derivative is ignored using the same scaling argument that led to neglecting the time derivative in (31). The collision term in (60) can be modeled in analogy to (43) as

$$\mathbf{C}_{p\mathcal{E}} = -\frac{q\mathbf{S}}{\mu_S} \tag{61}$$

which gives

$$\mathbf{S} = \frac{\mu_S}{\mu}(w\hat{I} + \hat{U}) \cdot \mathbf{v} + \frac{\mu_S}{qn}\left((w\hat{I} + \hat{U}) \cdot \nabla\cdot(n\hat{U}) - \nabla\cdot(n\hat{R})\right). \tag{62}$$

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

$$\hat{R} = \frac{5}{2}k_B^2 T_n^2 \hat{I}. \tag{63}$$

Using closure (63) and the same approximations that led to the three-moments ET model (56)–(59), a more accurate expression for $n\mathbf{S}$ is obtained from the fourth moment of the BTE

$$n\mathbf{S} = -\frac{\mu_S}{\mu}\frac{5}{2}\frac{k_BT_n}{q}\mathbf{J} - \frac{\mu_S}{\mu}\frac{5}{2}\left(\frac{k_B}{q}\right)^2 q\mu n T_n \nabla T_n \tag{64}$$

which should be used to replace (59) to give a four-moments ET model. Comparing (64) with (59) reveals the inconsistency of the three-moments ET model. In the four-moments model, we have the factor

$$\frac{5}{2}\frac{\mu_S}{\mu} \tag{65}$$

for both terms. In the three-moments model, however, the factors are $5/2$ and $5/2 - p$, which means that the heat flux can be adjusted independently. This inconsistency can be avoided only if $\mu_S/\mu = 1$ and $p = 0$. However, the ratio

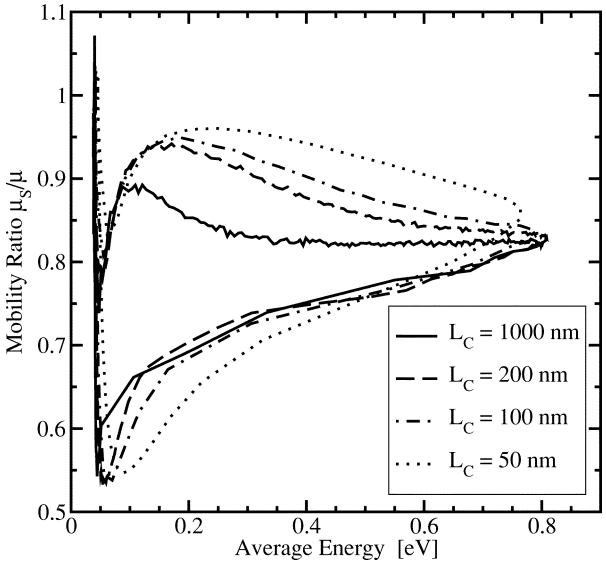


Fig. 3 Ratio of μ_S and μ as a function of the carrier temperature inside the n^+- n - n^+ structures obtained from MC simulations.

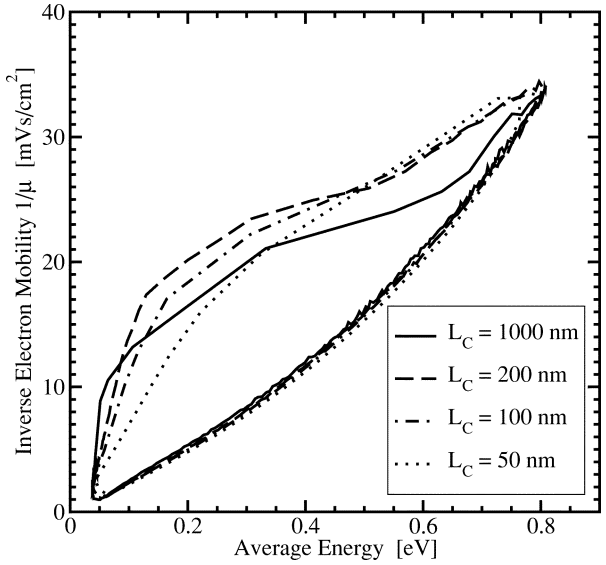


Fig. 4 Electron mobility inside n^+- n - n^+ structures obtained from MC simulations.

μ_S/μ strongly depends on the carrier temperature and shows a pronounced hysteresis when plotted over the average energy as shown in Fig. 3. Similar observations can be made for the electron mobility [15] which is shown in Fig. 4.

C. Methods Based on an Ansatz for the Distribution Function

Another approach which is frequently used in the derivation of macroscopic transport models appears to be similar to Bløtekjær's method. It relies on an Ansatz for the distribution function, which is then used to derive the closure relations. The difference from Bløtekjær's moment method will be outlined in the following, using the average energy w as an example.

During the derivation of (42) only general assumptions about the distribution function are made, in particular that the

distribution function is isotropic, that is, $f(\mathbf{k}) = f(|\mathbf{k}|)$. Furthermore, parabolic bands have been assumed. For the simplified expression (53), the convective energy is neglected.

When a certain form of the distribution function is assumed, normally a Maxwellian shape

$$f(\mathcal{E}) = A \exp\left(-\frac{\mathcal{E}}{k_B T_M}\right) \tag{66}$$

the average energy can be directly calculated as

$$w = \frac{\langle \mathcal{E} \rangle}{\langle 1 \rangle} = \frac{3}{2} k_B T_M \tag{67}$$

which is structurally equivalent to (53). However, the temperature T_n appearing in (53) is defined over averages of a generally unknown distribution function, whereas T_M is the parameter of (66). Using (66) for the calculation of the other closure relations, an ET model can be derived which has the same structure as the previously given four-moment ET model. This is not the case for different Ansatz functions, and becomes obvious when, for instance, a Fermi–Dirac distribution is assumed instead of a Maxwellian distribution [19].

Mixed procedures have been considered as well. In [29], for example, the author uses Bløtekjær's moments method to derive moment equations of arbitrary order. As Bløtekjær's method produces more unknowns than equations, closure relations are derived via an analytical expression for the distribution function.

Unfortunately, the analytical Ansatz functions, which are normally used for the distribution function, give only poor approximations in realistic devices [30]. As these closure relations are crucial for the accuracy of the transport model, methods based on analytic expressions have been shown to perform poorly [31], [32] and require a large number of moment equations. This has been interpreted as a failure of the moments method. We believe, however, that this is not a failure of the moments method but rather a failure of the Ansatz for the distribution function used in obtaining the closure relations.

D. Comparison

One of the extensively discussed differences between Bløtekjær's and Stratton's approach is that in Stratton's model, the mobility is *inside* the gradient $\nabla(n\mu_2 T_n)$, whereas in Bløtekjær's it stands *in front* of the gradient $\mu_1 \nabla(nT_n)$ in the current relation.

This issue was addressed by Stratton himself [33] and by Landsberg [34], [35]. Other comparisons of the two approaches can be found in [21], [36]–[39]. It is important to note that although the parameters μ_1 and μ_2 are called mobilities in both approaches, their definition differs significantly. Both approaches are compared in [36] and both appear to be justified, provided that the respective mobilities are modeled accordingly. For bulk simulations, the mobilities are equal [36], [40] and can be properly modeled using conventional energy-dependent expressions [41], [42]. 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 μ_1 formulation lies in the fact that for increasing values of the electric field, it can be roughly approximated by its bulk value, whereas μ_2 is always different. Thus, μ_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 $\mathbf{C}_p$ empirically as

$$\mathbf{C}_p = \mathbf{C}_p^* + \lambda_p \nabla \cdot \hat{U} \tag{68}$$

where $\mathbf{C}_p^*$ is the homogeneous component and λ_p a dimensionless transport coefficient, Tang *et al.* [43] showed that Stratton's model can be obtained from Bløtekjær's model with $\lambda_p = \eta_n$ where

$$\eta_n = -\frac{T_n}{\mu^*} \frac{\partial \mu^*}{\partial T_n}. \tag{69}$$

Here, $\mu^* = \mu^*(T_n)$ is the bulk mobility, which can be exactly represented as a function of the temperature. η_n given by (69) is similar to $-\nu_n$ given by (29) in Stratton's model. However, a comparison with MC data shows that $\lambda_p \neq \eta_n$ in inhomogeneous samples. In [44], it is proposed to use

$$\lambda_p = \frac{\eta_n}{1 + \eta_n}. \tag{70}$$

VI. Extensions for Degenerate Semiconductors

The previously given equations predict a Maxwell–Boltzmann distribution for the equilibrium case. This can be seen by solving the current relation for $\mathbf{J} = 0$, which gives

$$n = N_c \exp\left(-\frac{q(\psi - \psi_n) - E_c}{k_B T_L}\right). \tag{71}$$

This is problematic for low temperatures and degenerate semiconductors [45], [46]. In the case of the DD model, extended models can be derived by putting Fermi–Dirac statistics into the BTE [47]. Basically, the same procedure as for extended DD models is applicable for hydrodynamic models, and a heated Fermi–Dirac distribution has been used [11], [19], [46]. However, even though a Fermi–Dirac distribution correctly reproduces the equilibrium case, it becomes questionable for the description of nonequilibrium transport.

VII. Nonparabolicity Extensions

The general hydrodynamic equations (30)–(32) are valid for any band structure. However, parabolicity assumptions are invoked to derive the closure relations (36)–(42). Furthermore, nonparabolicity effects enter the hydrodymamic equations through the models used for the collision terms. A good example is the mobility whose homogeneous values are frequently obtained through measured $\mathbf{v}(\mathbf{E})$ characteristics. This mobility contains the full information of a real band structure, something which is much more difficult to obtain with MC simulations where the mobility has to be modeled using microscopic scattering rates [20].

As pointed out in the discussion of (16), there is no simple relationship between the average energy $\langle \mathcal{E} \rangle$ and the average

velocity $\mathbf{v}$ in the general case. For parabolic bands, the carrier temperature is normally defined via the average carrier energy as

$$T_n = \frac{2}{3k_B}\left(\langle\mathcal{E}\rangle - \frac{m^*\mathbf{v}^2}{2}\right). \tag{72}$$

Unfortunately, there is no similar equation for nonparabolic bands. Another possibility is to define the temperature via the variance of the velocity as [16]

$$T^* = \frac{m^*}{3k_B}\langle\mathbf{u}^2\rangle. \tag{73}$$

Definitions (72) and (73) are consistent with the thermodynamic definition of the carrier temperature in thermodynamic equilibrium, and both are identical for nonequilibrium cases when a constant carrier mass is assumed, which in turn corresponds to the assumption of parabolic energy bands. However, large differences are observed when a more realistic band structure is considered [16], [48].

A. The Generalized Hydrodynamic Model

Thoma *et al.* [16], [48] proposed a model which they termed the generalized hydrodynamic model. Instead of using the average energy *and* the temperature as variables in their formulation, they opted for a temperature-only description. To obtain a form similar to standard models, they defined the temperature according to (73), which differs significantly from (72) for nonparabolic bands. Instead of the momentum-based weight functions $\hbar\mathbf{k}$ and $\hbar\mathbf{k}\mathcal{E}$, they used the velocity-based weight functions $\mathbf{u}$ and $\mathbf{u}\mathcal{E}$ to derive the moment equations of order one and three. Without assuming any particular dispersion relation, they derived the following equations for the current and energy flux density:

$$\mathbf{J} = \frac{\tau_i}{\tau_i^*}\mu^* k_B\nabla(nT^*) + q\mu^* n\mathbf{E} \tag{74}$$

$$\nabla\cdot(n\mathbf{S}) = -\frac{3}{2}k_B\partial_t(nT^*) + \mathbf{E}\cdot\mathbf{J} - n\frac{3}{2}k_B\frac{T^*-T_L}{\tau_{\mathcal{E}}^*} \tag{75}$$

$$n\mathbf{S} = -\frac{\mu_S^*}{\mu^*}\frac{5}{2}\frac{k_BT^*}{q}\left(\mathbf{J} + \mu^*\frac{\tau_i}{\tau_i^*}nk_B\nabla T^*\right) \tag{76}$$

All relaxation times and mobilities are modeled as a function of T^*, and explicit formulas were given in [49]. The advantage of this formulation is that it can be applied to arbitrary band structures. Thoma *et al.*, however, used parameters extracted from MC simulations employing Kane's dispersion relation. As these parameters are extracted from homogeneous MC simulations, their validity for realistic devices is still an open issue. In a recent investigation [50], the relaxation times were calculated from full-band MC bulk simulations. Good agreement of transit-times of SiGe heterostructure bipolar transistors was observed in comparison with full-band MC simulations.

B. Model of Bordelon et al.

Bordelon *et al.* [51], [52] proposed a nonparabolic model based on Kane's dispersion relation. As weight functions, they used one, $\hbar\mathbf{k}$, and $\mathcal{E}$, and closed the system by ignoring the heat flux. To avoid the problem with the missing energy-temperature relation, they formulated their equation system solely in w. Their model is based on two assumptions: first, they assumed that the diffusion approximation holds. Regarding the energy tensor, this allows for the following approximation:

$$n\hat{U} = \langle\mathbf{u}\otimes\hbar\mathbf{k}\rangle \approx \frac{1}{3}\langle\mathbf{u}\cdot\hbar\mathbf{k}\rangle\hat{I} = \frac{2}{3}\left\langle\mathcal{E}\frac{1+\alpha\mathcal{E}}{1+2\alpha\mathcal{E}}\right\rangle \tag{77}$$

where Kane's relation has been used. The problem is now to express the right side of (77) by available moments without introducing new unknowns. This is not exactly possible, and the approximation

$$\left\langle\frac{1}{1+2\alpha\mathcal{E}}\right\rangle \approx \frac{1}{1+2\alpha\langle\mathcal{E}\rangle} \tag{78}$$

has been introduced. Defining $H(w) = (1+\alpha w)/(1+2\alpha w)$, they obtained

$$\mathbf{J} = \mu\frac{2}{3}\nabla(nwH(w)) + q\mu n\mathbf{E} \tag{79}$$

$$n\mathbf{S} = -\Omega(w)w\frac{\mathbf{J}}{q} \tag{80}$$

with $\Omega(w) \approx 1.3$. In the comparison made in [38], the predicted $\mathbf{v}$ and $\mathbf{S}$ curves agree quite well with the MC data, even with this simplified model for $\mathbf{S}$.

C. Model of Chen et al.

In [53], Chen *et al.* published a model which they termed the improved energy transport model. They tried to include nonparabolic and non-Maxwellian effects to a first order. Their approach is based on Stratton's, the use of Kane's dispersion relation, and an Ansatz for the distribution function

$$f(\mathcal{E}) = \left(1+\gamma\frac{\mathcal{E}}{k_BT_e}\right)\exp\left(-\frac{\mathcal{E}}{k_BT_e}\right) \tag{81}$$

which contains a non-Maxwellian factor γ. They give the following equations:

$$\mathbf{J} = k_B\nabla(n\mu T_m) + q\mu n\mathbf{E} \tag{82}$$

$$n\mathbf{S} = -C_e\left(\mu nk_BT_m\mathbf{E} + \frac{k_B^2}{q}\nabla\left(n\mu T_m^2\right)\right) \tag{83}$$

with

$$C_e = \left(\frac{5}{2}-p\right)\left(1-\frac{1}{2}\alpha k_BT_m\right) \tag{84}$$

$$\langle\mathcal{E}\rangle = \left(1+\frac{5}{2}\alpha k_BT_m\right)\frac{3}{2}k_BT_m. \tag{85}$$

Interestingly, the non-Maxwellian factor γ does not show up in the final equations. Sadovnikov *et al.* [49] showed that Chen's model shows some weakness in predicting proper velocity profiles and is not consistent with homogeneous simulation results.

D. Model of Tang et al.

This model is based on Kane's dispersion relation and takes particular care of correctly handling the inhomogeneity effects, which are commonly ignored [43]. By observing that $(\hat{U} - \mathbf{v}\otimes\mathbf{p})$ and $(\hat{R} + 2.5w\mathbf{v}\otimes\mathbf{p})$ show nearly no hysteresis

when plotted against the average energy for several n^+- n - n^+ structures, Tang *et al.* proposed the following closure relations:

$$\hat{U} = \frac{2}{3} w \hat{I} + \mathbf{v} \otimes \mathbf{p} + u(w) \hat{I} \tag{86}$$

$$\mathbf{p} = m^* \mathbf{v} + 2\alpha m^* \mathbf{S} \tag{87}$$

$$\hat{R} = \frac{10}{9} w^2 \hat{I} - \frac{5}{2} w \mathbf{v} \otimes \mathbf{p} + r(w) \hat{I} \tag{88}$$

with $u(w)$ and $r(w)$ being single-valued fit functions. The use of Kane's dispersion relation for the MC simulation might limit the validity of the preceding expressions. Unfortunately, the additional convective terms are difficult to handle and cause numerical problems [54]. Therefore, simplified expressions have been given in [55].

E. Model of Smith and Brennan

Smith and Brennan [19] derived two nonparabolic equation sets for inhomogeneous and degenerate semiconductors (see also [56], [57]). They used both Kane's dispersion relation and the power-law approximation after Cassi and Riccò [18] because the former cannot be integrated analytically. Furthermore, they used Fermi–Dirac statistics to include degeneracy effects. They showed that the typically employed binomial expansion of the Kane integrands loses its validity, and physically inconsistent results are obtained. The power-law approximation, on the other hand, approaches the parabolic limit and has a larger range of validity.

Their approach has two drawbacks: first, as pointed out previously, Cassi's density of states (19) cannot capture the nonparabolic nature of the bands; therefore, it is of limited use for a nonparabolic transport model. The authors themselves noted that for $\lambda > 1$, incorrect transport coefficients are obtained. Therefore, they obtained the parameter λ by a fit to the low energy range $[0, 0.2\,\mathrm{eV}]$ of Kane's dispersion relation. As can be seen in Fig. 2, this gives near-parabolic behavior of the density of states. Second, a heated Fermi–Dirac statistics provides no improvement over a heated Maxwell statistic in terms of hot-carrier transport.

F. Model of Anile et al.

Anile and Romano [58] and Muscato [59] derived closed-form expressions for the closure $\hat{U}$ and $\hat{R}$ using an Ansatz for the distribution function based on the maximum entropy principle. In addition, they were able to derive expressions for the collision terms. They found that their model fulfills Onsager's reciprocity principle and gave a comparison with other hydrodynamic models which violate the principle. Although Anile's model has a sound physical basis, it is of limited practical use. Despite its complicated nature, the model is based on an analytical expression for the distribution function, which was assumed to be of Maxwellian shape. Extended models were given in, e.g., [60]. Note, however, that Onsager's reciprocity theorem is valid only near equilibrium, a condition significantly violated in today's semiconductor devices [61], [62].

G. Comparison

The use of more realistic band structure models than the parabolic band approximation adds severe complications to macroscopic transport models. Even for the rather simple Kane dispersion relation, no closed-form equations can be given. Therefore, all models rely on more or less severe approximations. Whereas Thoma's model is applicable for general band structures, Tang's models attempt to capture nonlocal effects. A comparison of the simple ET model with the expressions given by Thoma, Lee, Chen, and Tang for silicon bipolar transistors is given by Sadovnikov *et al.* [49], where good agreement with MC data was obtained for the models of Thoma and Tang.

VIII. Extensions for Semiconductor Alloys

The derivations given previously are restricted to homogeneous materials where the effective carrier masses and the band edge energies do not depend on position. Over the last few years, extensive research has been made concerning compound semiconductors where the inclusion of the carrier temperature in the transport equations is generally considered a must. To properly account for the additional driving forces due to changes in the effective masses and the band edge energies, the ET models have been extended accordingly. The foundation for these extensions was laid in the pioneering work by Marshak for the DD equations [47], [63]. These concepts have been applied to the ET models by Azoff [11], [23], [64]. In the case of a position-dependent parabolic band structure, the force exerted on an electron is given as

$$\mathbf{F} = -\nabla E_c + \mathcal{E} \frac{\nabla m^*}{m^*}. \tag{89}$$

These additional forces give rise to an additional component in the current relation, and the electric field is replaced by an effective electric field which also contains the gradient of the band edges

$$\mathbf{J} = \mu n \nabla E_c + \frac{2}{3} \mu \nabla (nw) - \mu n w \nabla \ln(m^*) \tag{90}$$

$$n\mathbf{S} = -\frac{5}{3} \frac{\mathbf{J}}{q} w - \frac{10}{9} \mu k_B n w \nabla w. \tag{91}$$

An extension to nonparabolic band structures has been presented by Smith *et al.* [19], [56].

IX. Multiple Band Models

Bløtekjær's [3] equations were originally devised for semiconductors with multiple bands. Woolard *et al.* [65]–[67] extended these expressions for multiple nonparabolic bands in GaAs. Other models for compound semiconductors can be found in [68], [69]. Wilson [70] gave an alternate form of the hydrodynamic model, which he claims to be more accurate than that reported in [3]. Another multivalley nonparabolic ET model was proposed

in [71]. Due to the complicated band structure of III–V semiconductors, the effective electron gas approximation is even more critical than for silicon. This manifests itself in large hysteresis loops in the relaxation times [72], [73].

X. Band-Splitting Models and Higher Order Models

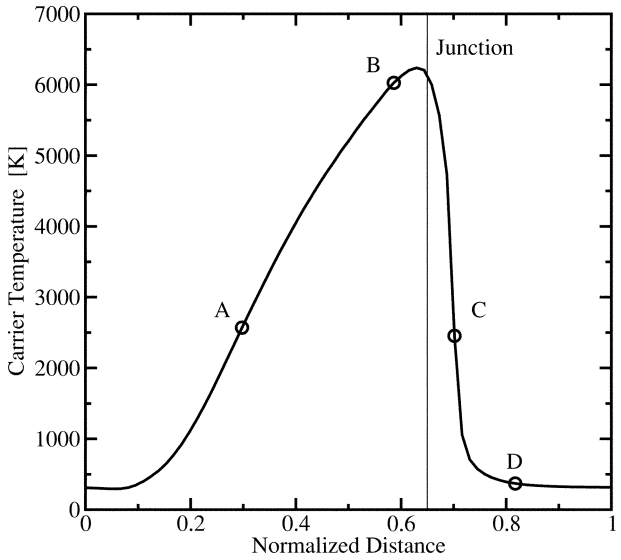
As device geometries are further reduced without according reduction of the supply voltages, the electric field occurring inside the devices increases rapidly. Furthermore, strong gradients in the electric field are observed. These highly inhomogeneous field distributions give rise to distribution functions which deviate significantly from the frequently assumed Maxwellian distribution. Furthermore, as has been pointed out in [15], [74], the distribution function is not uniquely described using just the average carrier energy. This is depicted in Fig. 5, which shows some electron distribution functions inside an n^+- n - n^+ structure obtained by MC simulation. Points A and B are in the channel, while points C and D are taken from the drain region. In the drain region, the overpopulation of the high-energy tail is obvious, whereas in the channel it is underpopulated, showing a significant thermal tail [75].

Several moment based models have been proposed so far, which aim at obtaining information about the distribution function in addition to the average energy. One approach is to split the energy range at some characteristic energy and handle both energy ranges with a two-population and two-temperature model [76], [77]. As these models were aimed at modeling impact ionization, the band gap energy was taken as the characteristic energy. This approach leads to various additional macroscopic parameters which model the transitions between the two energy regions. Determination of these parameters relies on carefully set up MC simulations. Due to this specialization to impact ionization, this model would have to be reformulated if another energy range is of interest, as is the case for the calculation of gate currents. Thus, this approach is difficult to generalize if both effects need to be captured at the same time, which is demanded for state-of-the-art devices. A special formulation using two electron populations has been proposed in [78] for those regions where the high-energy tail is heavily populated. In [79], a simplified version of the two-energy model [76] is given which used assumptions similar to those made by Cook and Frey [80] for ET models.

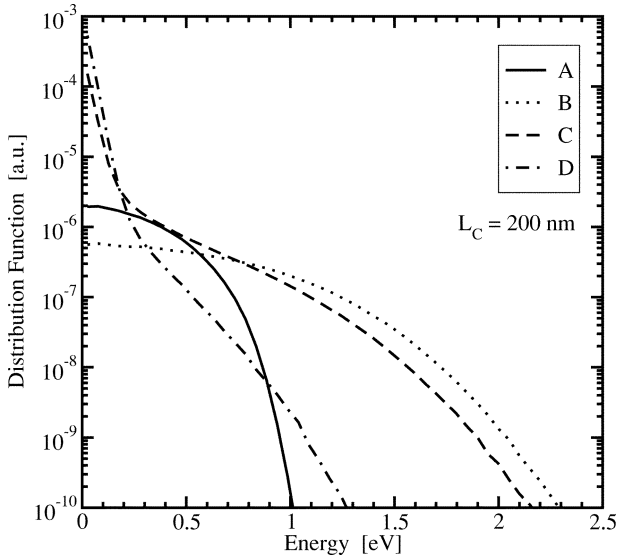
In [30], [81], it has been shown that important additional information about the distribution function can be obtained from a six-moments transport model. This model can be derived by the method of moments by including the two next higher order moments compared to a four-moments ET model [25], [82].

XI. Electrothermal Extensions

One of the problems resulting from the reduction in device geometries is that the generated heat has to be kept small



(a)



(b)

Fig. 5 Electron temperature and distribution functions at four characteristic points inside an n^+- n - n^+ structure with $Lc = 200$ nm. The average energies at the points A and C are the same, whereas the distribution function looks completely different. Note the high-energy tail at point D where the carrier temperature is 370 ° K.

because self-heating effects significantly influence the device characteristics. To capture these self-heating effects, the lattice heat flow equation has to be added to the transport models. Furthermore, the moment equations have to be extended to account for a nonconstant lattice temperature. A detailed treatment of this subject was given by Wachutka [5] for the DD equations. Chen *et al.* gave an extension for ET models in [83]. Benvenuti *et al.* introduced a thermal-fully hydrodynamic in [84].

XII. Critical Issues

The previously given models employ various approximations of different severity. As these approximations have been

discussed extensively in the literature, they will be summarized in this section.

A. Closure

The method of moments transforms the BTE into an equivalent, infinite set of equations. A severe approximation 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. Furthermore, the method of moments delivers more unknowns than equations which have to be eliminated by separate closure relations.

One approach to derive a suitable closure relation is to assume a distribution function and calculate the fourth-order moment. Geurts [85] expanded the distribution function around a drifted and heated Maxwellian distribution using Hermite polynomials. This gives a closure relation which generalizes the standard Maxwellian closure. However, these closures proved to be numerically unstable for strong electric fields. Liotta and Struchtrup [32] investigated a closure using an equilibrium Maxwellian, which proved to be numerically very efficient but with unacceptable errors for strong electric fields. For a discussion on heated Fermi–Dirac distributions see [19], [57]. Different closure relations available in the literature are compared in [38].

By introducing a non-Maxwellian and nonparabolicity correction factor

$$\beta_n = \frac{3}{5}\frac{\langle \mathcal{E}^2 \rangle}{\langle \mathcal{E} \rangle^2} \tag{92}$$

in the closure for the highest order moment $\hat{R}$

$$\hat{R} = \frac{5}{2} k_B^2 T_n^2 \beta_n \hat{I} \tag{93}$$

we obtain the following expression for the energy flux:

$$\mathbf{S} = \mathbf{S}^M + \mathbf{S}^{nM} \tag{94}$$

$$\begin{aligned} n\mathbf{S}^{nM} = &-\frac{5}{2}\frac{k_B^2}{q}\frac{\mu_S}{\mu}\mu \\ &\times \left[2nT_n(\beta_n - 1)\nabla T_n + nT_n^2 \nabla \beta_n \right. \\ &\left. + T_n^2(\beta_n - 1)\nabla n\right]. \end{aligned} \tag{95}$$

The lower order moment equations are not affected by the shape of the distribution function. $\mathbf{S}^M$ is given by (64) and represents the energy flux, if the fourth moment is evaluated from a Maxwellian distribution function. The correction factor β_n is the kurtosis of the distribution function and gives the deviation from the Maxwellian shape. As shown in Fig. 6, the kurtosis behaves fundamentally differently than in bulk [30] where a unique relationship $\beta_{\mathrm{Bulk}}(T_n)$ exists. Especially at the drain side of the structures we observe a strong deviation from the Maxwellian shape. This deviation corresponds to the high-energy tail in Fig. 5. Note that (95) contains both a gradient of n and a gradient of β_n. Furthermore, $\beta_n - 1$ has a different sign in the channel and the drain region.

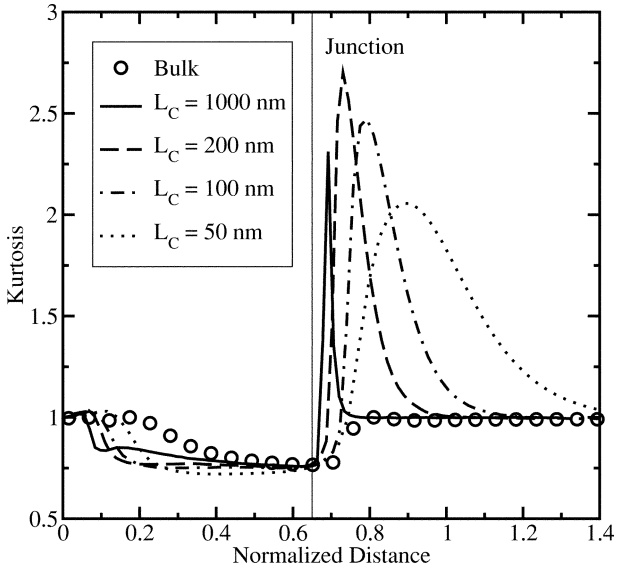


Fig. 6 The kurtosis for different n^+- n - n^+ structures. Note the different range on the x-axis which is required for small channel lengths.

B. Tensor Quantities

An issue which has only been vaguely dealt with is the approximation of the tensors by scalar quantities, such as the carrier mass and the carrier temperature. One-dimensional (1-D) simulations have been carried out in [15], which showed that the longitudinal temperature component T_{xx} is larger than the transverse temperature component T_{yy}. This indicates that the distribution function is elongated along the field direction and thus that the normally assumed equipartition of the energy is invalid (see Fig. 7). A rigorous approach has been taken by Pejčinović *et al.* [86], who model four components of the temperature tensor. They observed no significant difference between the scalar temperature and $tr\hat{T}_n/3$ for ballistic diodes and bipolar transistors but a 15% difference for submicrometer MOSFETs in the linear region of the transfer characteristics.

In addition, it has been observed that the energy tensor $\hat{U}$ is not a single-valued function of the average energy [43], and models using available moments have been given [see (86) and (88)].

C. Drift Energy Versus Thermal Energy

Another common approximation is the neglection of the drift energy in the average carrier energy [80]

$$w = \frac{1}{2}m^*\mathbf{v}^2 + \frac{3}{2}k_B T_n \approx \frac{3}{2}k_B T_n. \tag{96}$$

As has been pointed out by Baccarani and Wordeman [87], the convective energy can reach values comparable to the thermal energy. A plot of the ratio

$$\frac{\langle \frac{1}{2}m^*\mathbf{u}^2 \rangle}{\langle \mathcal{E} \rangle} \tag{97}$$

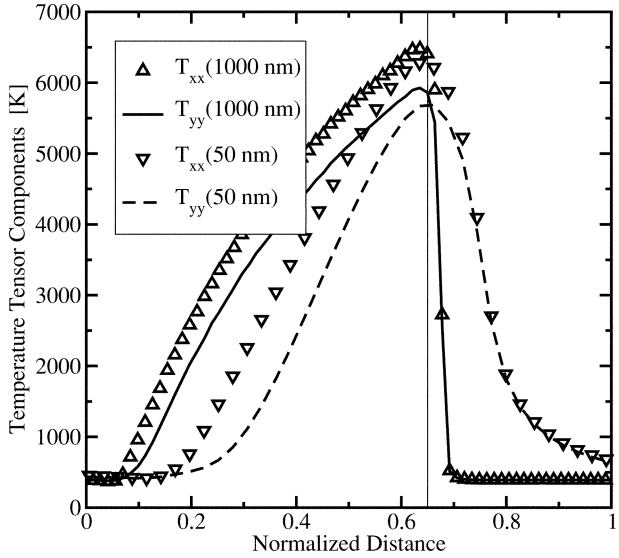


Fig. 7 Main components of the temperature tensor $\hat{T}$ for two different n^+- n - n^+ structures.

inside n^+- n - n^+ structures is given in Fig. 8. As can be seen, the error introduced by this approximation can reach 30% at the beginning of the channel where the carrier temperature is still low, and a velocity overshoot is observed. This effect has been studied in [87], [88].

However, when looking at the transport equations, we see that the gradient of the convective energy is of even more importance than its absolute value. A plot of the ratio

$$\frac{\nabla\langle\frac{1}{2}m^*\mathbf{u}^2\rangle}{\nabla\langle\mathcal{E}\rangle} \tag{98}$$

is given in Fig. 8, which indicates that this term is becoming important for nanoscale devices. In particular, pronounced spikes are observed at both junctions, the importance of which on fundamental quantities like the particle current is yet to be determined.

D. Relaxation Times

The relaxation times have traditionally been derived from homogeneous field measurements or MC simulations. For a homogeneous field, there is a unique relationship between the electric field and the carrier temperature via (5), which can be used as a definition for $\tau_{\mathcal{E}}$. However, due to the scattering operator in Boltzmann's equation, the relaxation times depend on the distribution function. Since the distribution function is not uniquely described by the average energy, models based on the average energy only are bound to fail. Furthermore, the band structure plays a dominant role. Nevertheless, all models should be able to correctly reproduce the homogeneous limit. In the following, some models for silicon are reviewed.

1) Mobility: Two models for the energy dependence of the mobility are frequently used, the model after Baccarani *et al.* [41], [87]

$$\frac{\mu(T_n)}{\mu_0} = \frac{T_L}{T_n} \tag{99}$$

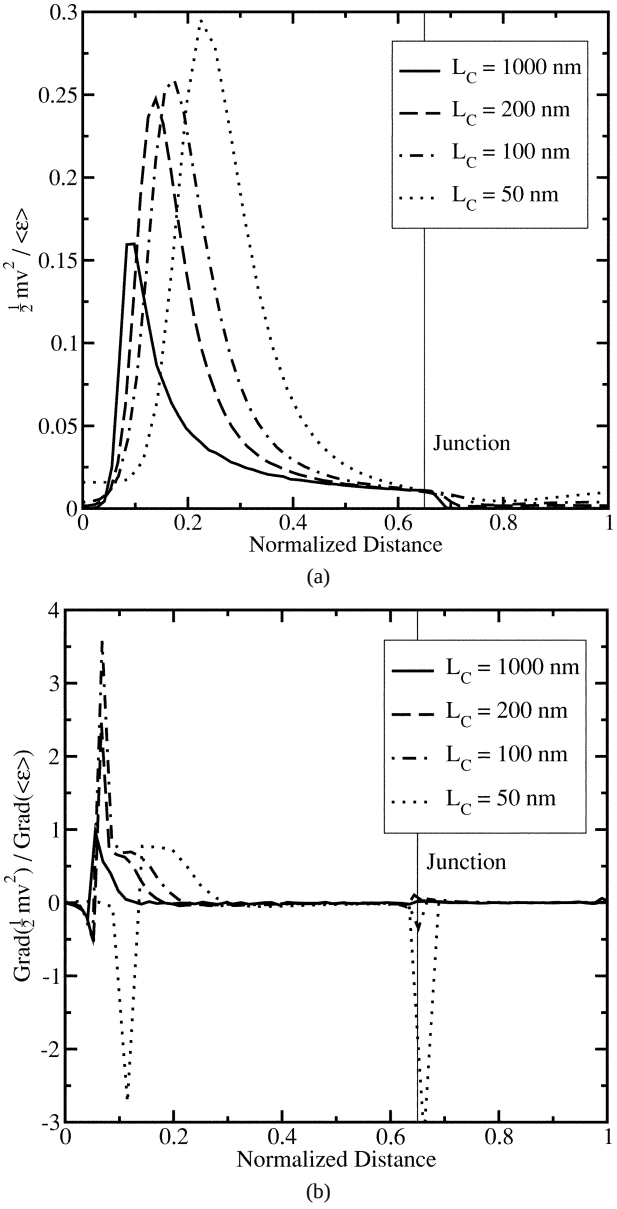


Fig. 8 The ratio of $\langle m\mathbf{u}^2/2\rangle$ and $\langle\mathcal{E}\rangle$ and the ratio of the gradients for different n^+- n - n^+ structures.

and the model after Hänsch [42], [89]

$$\frac{\mu(T_n)}{\mu_0} = \left(1 - \frac{3}{2}\frac{\mu_0}{\tau_{\mathcal{E}}v_s^2}\left(\frac{k_BT_L}{q} + \frac{2}{5}\frac{nS}{J}\right)\right)^{-1}. \tag{100}$$

Under homogeneous conditions the energy flux is proportional to the particle current

$$\frac{S}{J} = \frac{-5k_BT_n}{2q} \tag{101}$$

which can be used to simplify (100)

$$\frac{\mu(T_n)}{\mu_0} = \left(1 + \frac{3}{2}\frac{\mu_0 k_B}{q\tau_{\mathcal{E}}v_s^2}\left(T_n - T_L\right)\right)^{-1}. \tag{102}$$

As has been shown in [15], [90], (102) reproduces the mobility quite well in the regions with increasing $\mathbf{E}$. However, for decreasing $\mathbf{E}$, (100) should be used [15], [36]. A comparison of the three models with MC data is given in Fig. 9 and 10. All models have been evaluated using the data

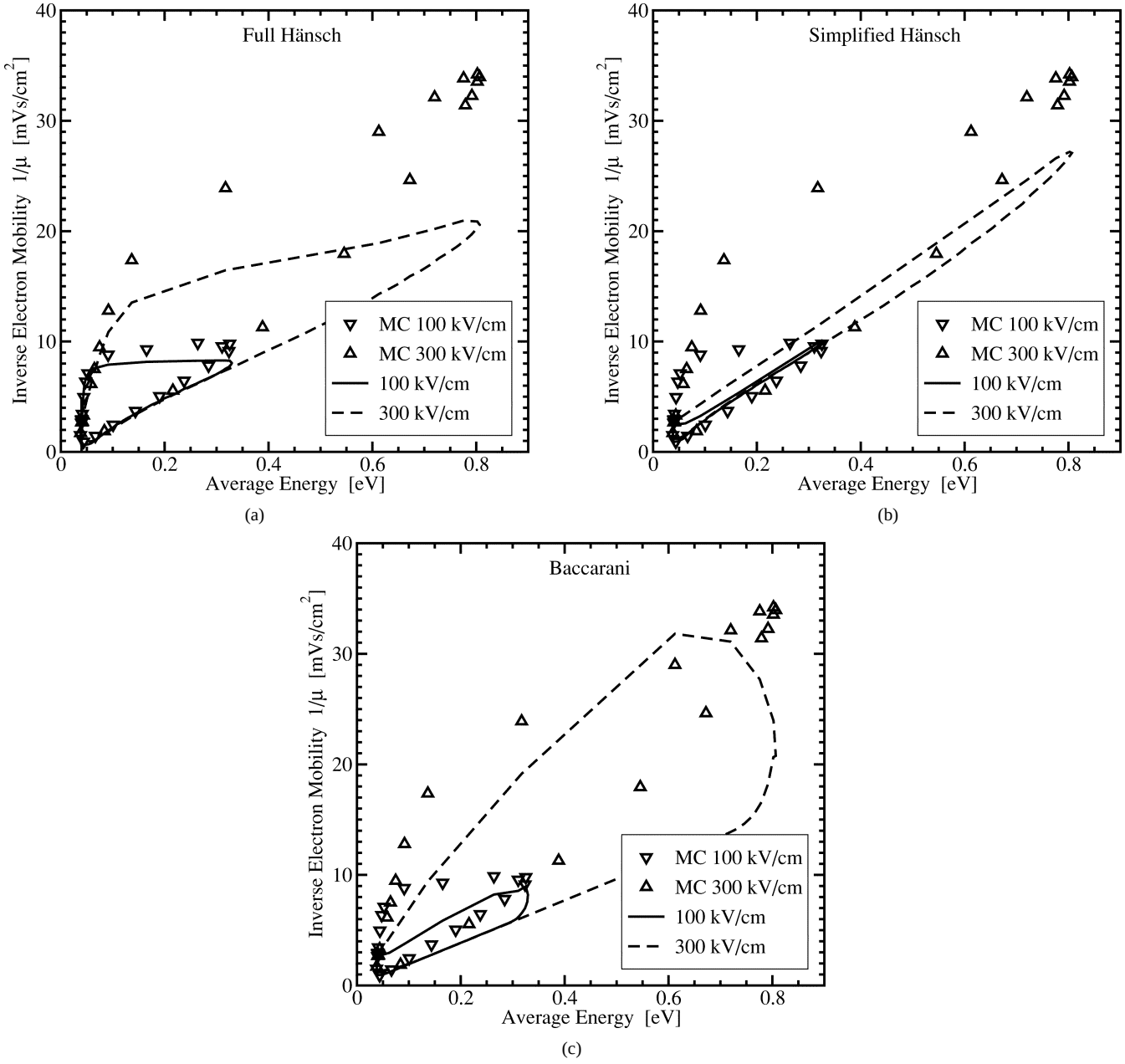


Fig. 9 Comparison of different mobility models with MC data.

from the MC simulation for an n^+- n - n^+ structure with $L_c = 200$ nm, once with $E_{\max} = 100$ kV/cm and once with $E_{\max} = 300$ kV/cm. For an electric field smaller than 100 kV/cm, (100) gives reasonable errors, but breaks down for larger electric fields. However, (100) is the only model that captures the hysteresis properly and thus the mobility at the beginning of the drain region. The hysteresis in the other models stems from the doping dependence of μ_0.

A different method for modeling the mobility has been proposed in [43], which is based on a separation of the homogeneous and inhomogeneous parts. They suggest modeling the collision term $\mathbf{C}_p$ as

$$n\mathbf{C}_p = \frac{\mathbf{J}}{\mu} = \frac{\mathbf{J}}{\mu^*} + \lambda_p n \nabla \cdot \hat{U} \tag{103}$$

with μ^* being the homogeneous mobility. The second term of (103) can then be moved to the left side of (31) to give a Stratton-like energy gradient expression $(1 - \lambda_p) n \nabla \hat{U}$. Accurate results for the quantity λ_p were obtained with (70), which is based on theoretically considerations [44].

2) Energy Relaxation Time: The simplest approach for modeling the energy relaxation time $\tau_{\mathcal{E}}$ is the use of a constant value. Typically used values for silicon at room temperature are in the range $[0.3, 0.4$ ps$]$, although values in the range $[0.08, 0.68$ ps$]$ have been used [72]. A constant value is insofar justified as MC simulations show only a small hysteresis and a small energy dependence [43]. However, different energy dependencies have been published. The differences seem to originate from the different band structures employed in the various MC codes.

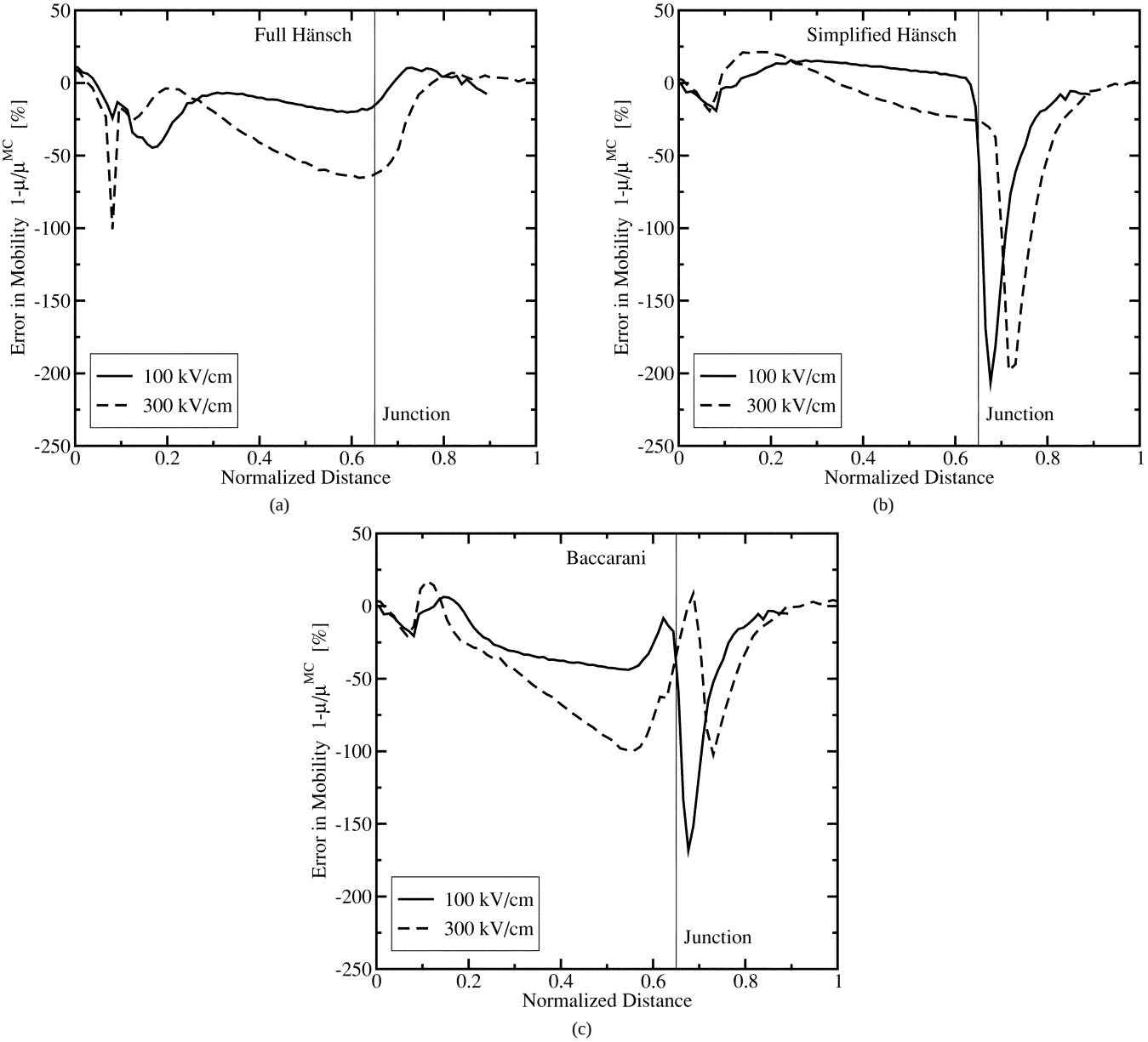


Fig. 10 Comparison of different mobility models with MC data.

Based on theoretical considerations, Baccarani *et al.* [41], [87] proposed the expression

$$\tau_{\mathcal{E}}(T_n) = \frac{3}{2}\frac{k_B\mu_0}{qv_s^2}\frac{T_nT_L}{T_n+T_L} + \frac{m^*\mu_0}{2q}\frac{T_n}{T_L}. \tag{104}$$

Equation (104) should be used together with (99) to correctly reproduce the homogeneous limit. Within Hänsch's approach, $\tau_{\mathcal{E}}$ is required only to be independent of the carrier temperature for (100) to correctly predict the homogeneous limit. When

$$\tau_{\mathcal{E}} = \frac{3k_B\mu_0T_L}{2qv_s^2} \tag{105}$$

is used in the Hänsch mobility models (100) and (102), the models are equivalent to Baccarani's model in the homogeneous case. A discussion of the inconsistencies resulting from mixing arbitrary energy-dependent mobility and energy relaxation time models can be found in [91].

Agostinelli *et al.* [92] proposed a model for the energy relaxation time for silicon which is fit to the data of Fischetti [93]

$$\frac{\tau_{\mathcal{E}}(W)}{1\text{ ps}} = \begin{cases} 0.172 + 2.656W - 3.448W^2, & \text{for } W \le 0.4 \\ 0.68, & \text{for } W > 0.4 \end{cases} \tag{106}$$

with $W = w/(1\text{ eV})$. Another more elaborate fit to the data of Fischetti is given in [94]. A maximum value of 0.68 ps seems to be too high, and yet another fit to newer data from Fischetti has been published by Hasnat *et al.* [95] as

$$\frac{\tau_{\mathcal{E}}(W)}{1\text{ ps}} = 0.27 + 0.62W - 0.63W^2 + 0.13W^3 + 0.01W^4 \tag{107}$$

with a maximum value of approximately 0.42 ps.

A detailed comparison of the effects of both relaxation times and transport models on the performance of silicon bipolar transistors is given in [49]. As the temperature profile

occurring inside the device is very sensitive to $\tau_{\mathcal{E}}$, this disagreement is rather unsatisfactory, although we believe for bulk silicon it should be in the range $[0.3, 0.4$ ps$]$.

3) Energy Flux Mobility: The ratio of the energy flux mobility and the mobility μ_S/μ is usually modeled as a constant with values in the range $[0.79, 1.0]$ (see, for instance, [15], [43]). In [43], it is proposed to model $\mathbf{C}_{p\mathcal{E}}$ as

$$\mathbf{C}_{p\mathcal{E}} = -\frac{q\mathbf{S}}{\mu_S^*} + \lambda_{p\mathcal{E}} \nabla \cdot \hat{R} \tag{108}$$

which is in analogy to (103). Here, μ_S^* is the homogeneous energy flux mobility. Expressions for μ_S^* and $\lambda_{p\mathcal{E}}$ can be found in [43].

E. Spurious Velocity Overshoot

Models based on Bløtekjær's approach have been frequently associated with spurious velocity overshoot (SVO), that is, nonplausible spikes in the velocity characteristics which do not occur in MC simulations. This effect can be seen in Fig. 11, where a spurious peak in the velocity profile which does not exist in the MC simulation is clearly visible.

Several theories have been put forward to explain this effect. Some authors argue that it is related to the hysteresis in the mobility [74], whereas others relate it to nonparabolicity effects [52]. Still others argue that it is related to the closure of the ET equation system [88]. The improvement obtained by the nonparabolic model [52] is probably due to the improved closure relation for $\hat{R}$. As already argued by [88], SVO is not likely to be caused only by the mobility, because the mobility is not properly modeled in the whole $\nabla \cdot \hat{U} < 0$ region and SVO is restricted to a very small area. In [15], SVO is investigated using different mobility models, and it is found that improvement is possible when proper mobility models are used. For example, with the Hänsch mobility model (100), these spikes are strongly diminished but not completely removed. Unfortunately, MC simulations show that (100) also overestimates the real velocity overshoot at the beginning of the channel. In [43], SVO is attributed to the improper modeling of λ_p and $\lambda_{p\mathcal{E}}$ in (103) and (108), respectively.

Chen *et al.* [96] proposed a model based on Stratton's approach. In their simplified analysis, they used Baccarani's mobility model, which gives $\nu_n = -1$ in (29) and thus removes any thermal diffusion current inside the whole device, which is questionable. As a result, SVO is overly suppressed, and the velocity overshoot at the beginning of the channel is overestimated.

The error in the closure relation for $\hat{R}$ is important for explaining SVO [97], as shown in Fig. 11. For these simulations, the relaxation times and mobilities have been taken from a coupled MC simulator in a self-consistent manner to rule out any errors introduced by these models. When the order of the transport model is increased to include six moments of Boltzmann's equation [25], the spurious peak is reduced. When in addition to the relaxation times, the closure is also taken from the MC simulation, the spurious peaks disappear.

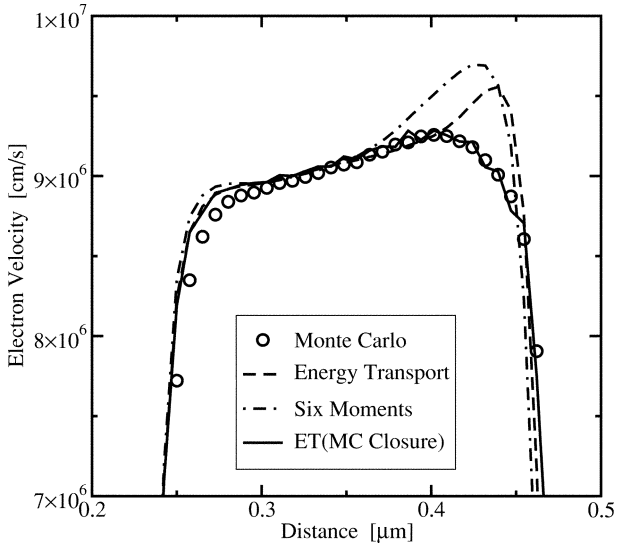


Fig. 11 Comparison of velocity profiles delivered by two transport models with MC data for an n^+- n - n^+ structure with $L_c = 200$ nm. Both transport models use relaxation times and mobilities from the MC simulation. In addition, when the ET model is closed with the data from the MC simulation, the SVO disappears.

XIII. Numerical Issues

The previously given transport models describe the spatial and temporal distributions of continuous quantities such as the carrier concentration and carrier temperature. As no closed-form solutions of these equation systems for practical devices can be given, they have to be solved numerically. This is done by discretizing the equation systems on suitable grids. The question of whether a grid is suitable for the problem under consideration is a complicated issue [98], [99]. From a practical point of view, however, it can be noted that improper grids are a common source of errors [100].

The revolutionary idea of Scharfetter and Gummel [1] enabled the first successful numerical device simulations and is still the basis of the most commonly used discretization schemes. Other schemes based on a different theoretical background have been proposed, but they still display similar mathematical properties to the scheme originally proposed by Scharfetter and Gummel.

As mentioned before, the FHD is hyperbolic in the supersonic regions, and special hyperbolic methods have to be used [27], [101]–[103]. Therefore, the Scharfetter–Gummel scheme cannot be applied to 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 [104]. However, this approach will degrade the convergence in cases where the variation in space or time is important [84]. Thus, to derive a spatial discretization, fluid dynamics methods known as upwinding are used [27], [54], [84], [105]. Other schemes such as essentially nonoscillatory, and Galerkin methods [106]–[108] also prove to be effective in solving nonlinear hyperbolic systems. However, these schemes require a time-dependent term; therefore, they are

less efficient when applied to steady-state problems. Furthermore, when the convective terms are involved, the handling of boundary conditions becomes more difficult [103], [109], [110].

When the convective terms are neglected, Scharfetter–Gummel-type schemes can be applied to higher order transport models. However, as this extension of the Scharfetter–Gummel scheme is not straightforward, several variants exist [111]–[116]. In general it can be noted that the convergence properties degrade significantly when compared to the classic DD model. Some improvement has been obtained by refining the discretization scheme [114], but the problem still persists. Other improvements in terms of convergence are based on iteration schemes where the equations are solve in a decoupled manner [8], [117], similar to an idea of Gummel for the DD model [118].

XIV. Validity of Moment-Based Models

When the critical dimensions of devices shrink below a certain value (around 200 nm for silicon at room temperature) MC simulations reveal strong off-equilibrium transport effects such as velocity overshoot and nonlocality of important model parameters. Therefore, the range of validity for moment-based models has been extensively examined. Furthermore, with shrinking device geometries, quantum effects gain more importance and limit the validity for the BTE itself [119]. Banoo and Lundstrom [120] compared the results obtained by (56)–(59) with a DD model and a solution of the BTE obtained by using the scattering matrix approach. They found that this ET model dramatically overestimates both the drain current and the velocity inside the device. Tomizawa *et al.* [121] found through a comparison with MC simulations that relaxation time based models tend to overestimate nonstationary carrier dynamics, especially the energy distribution. Nekovee *et al.* [31] compared moment hierarchy based models with a solution of the BTE and found that their model fails in the prediction of ballistic diodes because the equation hierarchy converges too slowly. However, their "moment" model was based on an Ansatz for the distribution function using a Maxwellian shape multiplied by a sum of Hermite polynomials. Indeed, such an expansion of the distribution function converges too slowly [30], but as there the parameters of the distribution function and not the moments of the BTE are considered, we do not feel their conclusions are equally valid for moment-based models. A conclusion similar to Nekovee's was drawn by Liotta and Struchtrup [32], who found that a hierarchy containing 12 moment equations was needed to reproduce results similar to those obtained by spherical harmonics expansions. In their model, the closure relations were calculated via an Ansatz for the distribution function; therefore, they cannot easily be generalized to moment-based models with carefully derived closure relations.

The nonequilibrium transport in nanoscale devices is characterized by the ratio $l_{\rm in}/L_c$, where $l_{\rm in}$ is the inelastic mean free path and L_c is the device characteristic length. Assuming $l_{\rm in} = v_{\rm th}\tau_{\mathcal{E}}$ where $v_{\rm th} = \sqrt{3k_BT/m^*}$ and taking the energy relaxation time $\tau_{\mathcal{E}}$ to be 0.3 ps for silicon, we have $l_{\rm in} = 31.7\sqrt{T_n/T_L}$. This gives $l_{\rm in} = 30$–80 nm depending on the electron temperature. It takes a distance of at least $3l_{\rm in}$ for electrons to attain a "local equilibrium" average energy. Thus, for any device with L_c less than 200 nm, the carrier transport in the channel is intrinsically nonstationary. For $L_c = 100$–50 nm, this ratio becomes smaller than one. This implies that in this quasiballistic regime, even if only 50% of the transport is ballistic, it is rather difficult to construct a "universal" hydrodynamic model because the model now depends on this ratio. In fact, our own simulations show that in the modeling of $\hat{U}$, the plot of $\hat{U} - \mathbf{v}\otimes\mathbf{p}$ versus w is almost a single-valued function of w, e.g., $r(w)$. For $L_c = 200$ nm, $r(w)$ is almost linear with a slope of 1.6/3. For $L_c = 100$ nm, $r(w)$ is a straight line with a slope of 3/5. For $L_c = 50$ nm, $r(W)$ is also a straight line but with a slope of 2/3. This implies that if we want to model $\hat{U}$ in this regime, we have approximately $\hat{U} = \mathbf{v}\otimes\mathbf{p} + zw$, where z is a function of $l_{\rm in}/L_c$. Therefore, when considering devices in the nanometer regime, it might be wise to compare the solution of moment-based models to a solution of Boltzmann's equation obtained by either MC or other methods.

XV. Simplified Models

Despite the limitations and approximations contained in the moment equations given previously, the handling of these equations is still far more complicated than that of the robust and well-studied DD equations. Thus, several researchers have tried to find suitable approximations to simplify the problem. These approximations were frequently used in postprocessors to account for an average energy distribution different from the local approximation. Slotboom *et al.* [122] used this technique to calculate energy-dependent impact ionization rates via a postprocessing model. Cook and Frey [80], [123] proposed a simplified model by using the approximations $v_x \gg v_y$ and $E_x \gg E_y$ in a two-dimensional (2-D) silicon MESFET to yield

$$\frac{\partial w}{\partial x} = \frac{21}{20}qE_x - \frac{9}{20}\left(\frac{40}{9}\frac{m^*(w-w_0)}{v_x\tau_{\mathcal{E}}\tau_p} + q^2E_x^2\right)^{1/2}. \tag{109}$$

Thus, the energy balance equation and the continuity equation become decoupled, and the complexity of the problem is considerably reduced. Approximations for GaAs were also given. Although these approximations might have delivered promising results, progress in the size reduction of state-of-the-art devices makes the assumptions $v_x \gg v_y$ and $E_x \gg E_y$ questionable. In particular, for deep-submicrometer MOSFETs, velocity overshoot influences the electric field distribution for a given bias condition and effectively defines a higher drain saturation voltage, which in turn defines a higher current [124].

To bring the ET equations into a self-adjoint form, Lin *et al.* [125] approximated the carrier temperature in the diffusion coefficient by the lattice temperature as

$$\mathbf{J} = \mu k_B \underbrace{T_n}_{\approx T_L}\nabla n + \mu k_B n\nabla T_n + qn\mu\mathbf{E} \tag{110}$$

which will underestimate the diffusion current by a factor of ≈ 10–20 for today's devices; therefore, this technique cannot be recommended.

Another simplified model is Thornber's generalized current equation [126], later called the augmented DD model [127], [128]. In these models, the field dependence of the transport coefficients is extended to include the gradient of the electric field. The motivation for this model is very similar to (109). However, the augmented DD model incorporates nonlocality effects using the electric field rather than the average energy. It can successfully catch nonlocal effects occurring in the device when the electric field is rising, but fails to do so when the electric field is abruptly decreasing, such as near the drain region in MOSFETs. Because of this, its usefulness in nanoscale device simulation is very limited.

XVI. Application to MOS Transistors

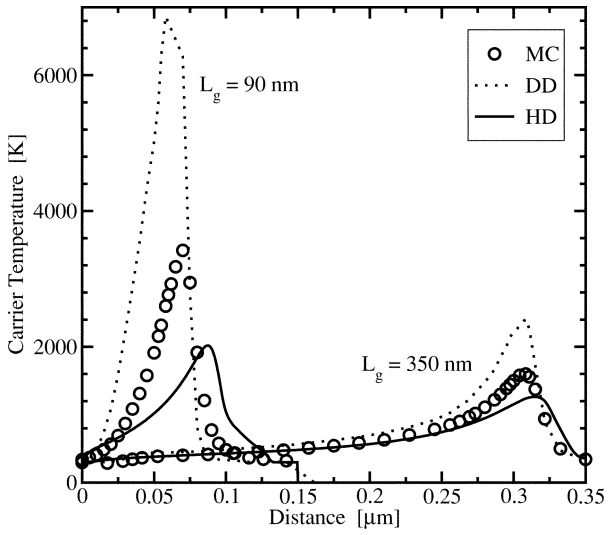
So far, the transport models have been checked with n^+-n-n^+ structures. Due to their simplicity, these are the most commonly used structures when a comparison with MC data is required. This is mainly because they are 1-D and require only one carrier type. Therefore, the influence of the various parameters on basic quantities like the velocity or the carrier temperature can be more easily separated and interpreted. As can be seen from the previous examples, even for these simple structures, such an interpretation is far from trivial.

Although it has been frequently claimed that n^+-n-n^+ structures emulate the behavior of MOS transistors, the most important devices in silicon technology, this is only partly true. MOS transistors are inherently 2-D devices, a fact that makes a comparison of moment-based models with MC more involved. In the following, the most important differences to n^+-n-n^+ structures are compared. The MC simulations were performed using MINIMOS [129], whereas MINIMOS-NT [130] was used for the ET model, which is based on (56)–(58) and (64). Hänsch's mobility model (102), a ratio μ_S/μ of 0.8, and $\tau_{\mathcal{E}} = 0.3$ ps have been chosen because this is basically the model available in commercial simulators. No fitting was performed.

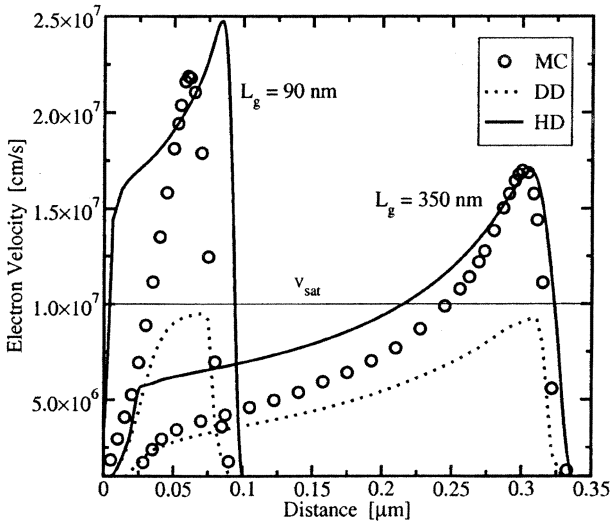
The basic quantities velocity and temperature show similar features for MOS transistors as shown in Fig. 12 and for n^+-n-n^+ structures (cf. Fig. 1). However, the velocity overshoot at the beginning of the channel and the SVO observed in n^+-n-n^+ structures coincide in MOS transistors to give a single overshoot in the pinch-off region.

A striking difference can be observed in the now 2-D electron concentration, which spreads much deeper into the bulk than would be expected from MC simulations. A typical situation is depicted in Fig. 13, where the electron concentration of an MOS transistor with $L_g = 130$ nm resulting from an MC simulation is compared with that of an ET simulation. The overestimated spreading of the carriers in the ET simulation can be clearly seen.

The reason for this effect is that the thermal diffusion of carriers in the direction normal to the current flow is overestimated. Several measures have been taken to reduce this



(a)



(b)

Fig. 12 Comparison of the simulated velocities and temperatures for two MOS transistors.

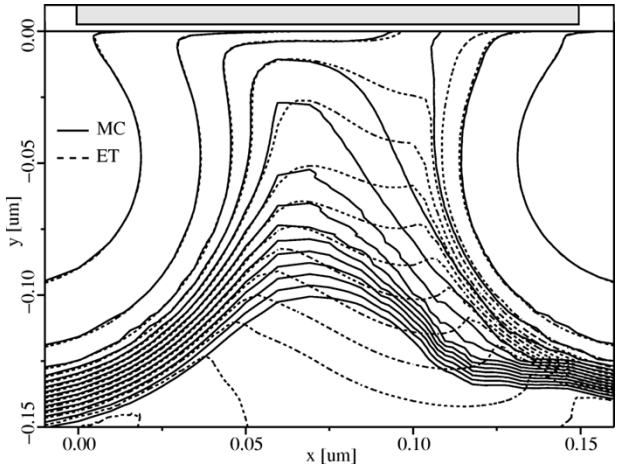


Fig. 13 Comparison of the simulated electron concentration of an $L_g = 130$ nm MOS transistor from an MC simulation and an ET model. Neighboring lines differ by a factor of ten. The overestimated carrier spreading into the bulk in the ET simulation can be clearly seen.

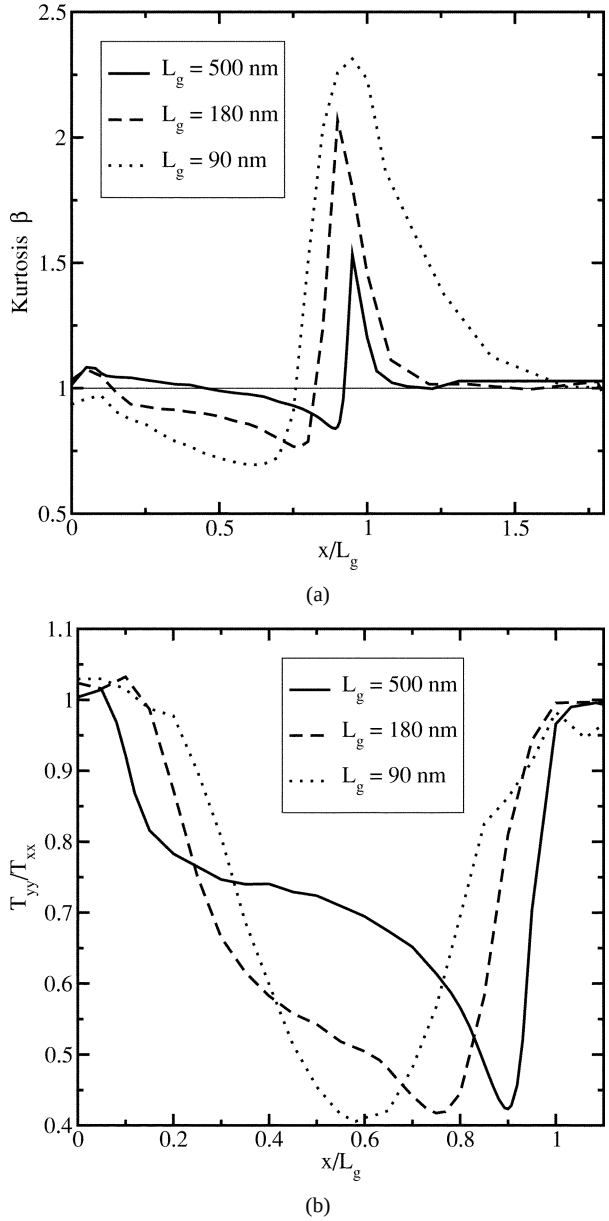


Fig. 14 The error in the closure condition given by the kurtosis β_n and the ratio of the temperature tensor components T_{yy} and T_{xx} in small MOS transistors.

effect, including the artificial reduction of the heat flux term by a factor of 0.2 [131].

In a recent study [132], this effect has been related to errors introduced by the closure of the equation system and to the error introduced by approximating the temperature tensor with a scalar value. These quantities are shown in Fig. 14 for three MOS transistors. Whereas the kurtosis β_n behaves qualitatively as in n^+-n-n^+ structures, the anisotropy observed in MOS transistors is much larger. As the energy is assumed to be equally partitioned over the components of the temperature tensor, an overestimation of the temperature component into the bulk is obtained.

This enhanced spreading of the carriers into the bulk leads to a complete breakdown of the ET model in the case of partially depleted silicon-on-insulator (SOI) transistors where the excess carriers recombine in the bulk and virtually turn the transistor off via the bulk effect. A modified ET model has been proposed in [132] where both the closure and the anisotropy was modeled based on empirical corrections. In MOS transistors, this effect is much less important and has as such been considered only as a cosmetic problem of ET models as the body potential is not influenced by this effect.

With respect to the terminal characteristics, it has been observed that ET models deliver higher currents than measurements and MC simulations, whereas the currents from DD simulations are too low [37], [133]. The sensitivities of the drain current with respect to a change in the parameter values

$$S(p) = \frac{p}{I_d}\left.\frac{\partial I_d}{\partial p}\right|_{V_d=V_g=1\text{ V}}$$

were found to be $S(\tau_\mathcal{E}) = 0.01$ and $S(\mu_S/\mu) = 0.045$ for $L_g = 350$ nm and $S(\tau_\mathcal{E}) = 0.17$ and $S(\mu_S/\mu) = 0.2$ for $L_g = 90$ nm. This gives for a 1% change in the energy relaxation time a 0.01% and a 0.17% change in I_d for $L_g = 350$ nm and $L_g = 90$ nm, respectively. As such, I_d is not too sensitive to inaccuracies in the model parameters. Other quantities, like the impact ionization rate, can be expected to show a higher sensitivity.

XVII. Conclusion

Many different hydrodynamic and ET models have been developed so far. They rely on either Stratton's or Bløtekjær's approach to find a suitable set of balance and flux equations. FHDs obtained by the method of moments are still too complicated to handle due to their hyperbolic nature. Within the diffusion approximation, hydrodynamic models simplify to ET models, which form the bulk of the models in use today. The crucial points are summarized in the following.

1) Band structure: Accurate description of the band structure is important for hot-carrier effects. However, most models in use rely on the single effective parabolic band model because even for the relatively simple nonparabolicity correction given by Kane, no closed-form solutions can be given. By the introduction of additional relaxation times, information about the full-band structure can be incorporated into the macroscopic models. So far, this has been achieved only for the homogeneous case, and the accuracy of this approach for deep-submicrometer devices is still to be determined.
2) Nonhomogeneous effects: The transport parameters like the mobilities are normally taken from bulk simulations or measurements and modeled as a function of the average carrier energy. For Bløtekjær's approach, this has been shown to give reasonable results inside the channel region, where the absolute value of the electric field increases. At the end of the channel region where the electric field decreases, however, these models give wrong results. In this region, a cold- and a hot-carrier population coexist, and the average carrier energy does not provide enough information to describe this circumstance. For long-channel devices,

this region is much shorter than the channel region, and the error introduced probably small. However, for devices smaller than approximately $L_g = 100$ nm, the length of this region is in the order of the channel length. This implies that the hot-carriers injected into the drain require a distance similar to the channel length to relax. Therefore, the influence of this region will become much more important for future technologies.

3) Closure: The method of moments transforms the BTE into an equivalent, infinite set of equations which has to be truncated to obtain a tractable equation set. A closure relation has to be formulated for the highest order moment, which is normally obtained by assuming a heated Maxwellian distribution function. This has been shown to be a rather crude approximation for the distribution function occurring in modern devices. As the lower order equations remain unchanged by this choice, it is worth noting that the *whole* information about the remaining higher order equations has to be packed into this closure.
4) Anisotropy: For the modeling of the temperature tensor equipartition of the energy is assumed. This approximation has been shown to be invalid both in n^+-n-n^+ structures and in MOS transistors. As the lateral component of the temperature tensor has no influence on the current in n^+-n-n^+ structures, and as it only indirectly influences the drain current in MOS transistors, this feature has been considered of minor importance. However, the carriers extend much deeper into the bulk of MOS transistors than predicted by MC simulations, which influences the modeling of energy-dependent parameters like the mobility and impact ionization. For partially depleted SOI transistors, on the other hand, this feature is crucial, and ET models cannot be used for the prediction of transfer characteristics. Rigorous handling of this issue was found to be difficult, as additional equations are required for each temperature tensor component, but empirical corrections achieved promising results.
5) Drift energy: As most ET models are based on the diffusion approximation, the drift energy is neglected. Investigations show that the drift energy can amount up to 30% of the total energy inside the channel region. However, to our best knowledge, a detailed investigation of this approximation on basic parameters like the drain current and carrier temperature is still missing.
6) Velocity overshoot: As a result of some of the above mentioned approximations, ET models tend to overestimate the velocity overshoot and show an SVO at the end of the channel region of n^+-n-n^+ structures. In MOS transistors, the spurious peak coincides with the velocity overshoot at the end of the channel and is as such not explicitly visible. A quantitative study of how the error in the velocity impairs the overall quality of ET models in modern MOS transistors is still missing.
7) Hot-carrier effects: As hydrodynamic and ET models provide only the first two moments of the energy distribution function, modeling of hot-carrier effects is difficult. It has been shown that the energy distribution function is not uniquely described by the concentration and the average energy only. Hot-carrier effects like impact ionization are particularly sensitive to the exact shape of the distribution function, and models based on the local average energy only are bound to fail. Improvements have been obtained by considering nonlocal models, which come with their own drawbacks and heuristics. Extension of the ET model to a six-moments model has proven to improve the accuracy considerably.

From a practical point of view, however, it is of note that only simple ET models are available in commercial simulators, which suffer from many of the demonstrated weaknesses. In such a case, it might be advantageous to compare the results with MC data or measurements. Of course, this applies also to devices which operate close to the ballistic limit. Application of these models in a "black-box" manner like in the "golden age" of the DD model ($L_g > 1\ \mu$m) is no longer possible. This is, however, not related to a principal weakness of higher order models compared to the DD model, but is rather a consequence of the complicated physics (e.g., quantum effects and semiballistic transport) which have to be captured in miniaturized devices. Despite these limitations, excellent results have been obtained in carefully set-up simulations. However, higher order moment based models still require a lot of fine-tuning and a detailed understanding of the underlying physical phenomena.

REFERENCES

[1] D. L. Scharfetter and H. K. Gummel, "Large-signal analysis of a silicon read diode oscillator," *IEEE Trans. Electron Devices*, vol. ED-16, pp. 64–77, Jan. 1969.
[2] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers," *Phys. Rev.*, vol. 126, no. 6, pp. 2002–2014, 1962.
[3] K. Bløtekjær, "Transport equations for electrons in two-valley semiconductors," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 38–47, Jan. 1970.
[4] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. New York: Springer-Verlag, 1984.
[5] G. K. Wachutka, "Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling," *IEEE Trans. Computer-Aided Design*, vol. 9, pp. 1141–1149, Nov. 1990.
[6] K. Hess, *Theory of Semiconductor Devices*. Piscataway, NJ: IEEE, 2000.
[7] J. D. Bude and M. Mastrapasqua, "Impact ionization and distribution functions in sub-micron nMOSFET technologies," *IEEE Electron Device Lett.*, vol. 16, pp. 439–441, Oct. 1995.
[8] B. Meinerzhagen and W. L. Engl, "The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon submicrometer MOS transistors," *IEEE Trans. Electron Devices*, vol. 35, pp. 689–697, May 1988.
[9] J. D. Bude, "MOSFET modeling into the ballistic regime," in *Proc. Simulation Semiconductor Processes and Devices*, 2000, pp. 23–26.
[10] D. K. Ferry, *Semiconductors*. New York: Macmillan, 1991.
[11] E. M. Azoff, "Generalized energy-momentum conservation equation in the relaxation time approximation," *Solid-State Electron.*, vol. 30, pp. 913–917, Sept. 1987.
[12] C. L. Gardner, "The classical and quantum hydrodynamic models," in *Proc. Int. Workshop Computational Electronics*, 1993, pp. 25–36.
[13] M. C. Vecchi and M. Rudan, "Modeling electron and hole transport with full-band structure effects by means of the spherical-harmonics expansion of the BTE," *IEEE Trans. Electron Devices*, vol. 45, pp. 230–238, Jan. 1998.

[14] C.-K. Lin, N. Goldsman, I. Mayergoyz, S. Aronowitz, and N. Belova, "Advances in spherical harmonic device modeling: Calibration and nanoscale electron dynamics," in *Proc. Simulation Semiconductor Processes and Devices*, 1999, pp. 167–170.
[15] S.-C. Lee and T.-W. Tang, "Transport coefficients for a silicon hydrodynamic model extracted from inhomogeneous monte-carlo calculations," *Solid-State Electron.*, vol. 35, no. 4, pp. 561–569, 1992.
[16] R. Thoma, A. Emunds, B. Meinerzhagen, H. J. Peifer, and W. L. Engl, "Hydrodynamic equations for semiconductors with nonparabolic band structure," *IEEE Trans. Electron Devices*, vol. 38, pp. 1343–1353, June 1991.
[17] E. O. Kane, "Band structure of indium antimonide," *J. Phys. Chem. Solids*, vol. 1, pp. 249–261, 1957.
[18] D. Cassi and B. Riccò, "An analytical model of the energy distribution of hot electrons," *IEEE Trans. Electron Devices*, vol. 37, pp. 1514–1521, June 1990.
[19] A. W. Smith and K. F. Brennan, "Non-parabolic hydrodynamic formulations for the simulation of inhomogeneous semiconductor devices," *Solid-State Electron.*, vol. 39, no. 11, pp. 1659–1668, 1996.
[20] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Phys. Rev. B*, vol. 38, no. 14, pp. 9721–9745, 1988.
[21] Y. Apanovich, E. Lyumkis, B. Polsky, A. Shur, and P. Blakey, "Steady-state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 702–711, June 1994.
[22] K. Hess, *Advanced Theory of Semiconductor Devices*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[23] E. M. Azoff, "Energy transport numerical simulation of graded AlGaAs/GaAs heterojunction bipolar transistors," *IEEE Trans. Electron Devices*, vol. 36, pp. 609–616, Apr. 1989.
[24] M. Lundstrom, *Fundamentals of Carrier Transport*. Reading, MA: Addison-Wesley, 1990, vol. 10, Modular Series on Solid State Devices.
[25] T. Grasser, H. Kosina, M. Gritsch, and S. Selberherr, "Using six moments of Boltzmann's transport equation for device simulation," *J. Appl. Phys.*, vol. 90, no. 5, pp. 2389–2396, 2001.
[26] K. Souissi, F. Odeh, H. H. K. Tang, and A. Gnudi, "Comparative studies of hydrodynamic and energy transport models," *COMPEL*, vol. 13, no. 2, pp. 439–453, 1994.
[27] C. L. Gardner, "Numerical simulation of a steady-state electron shock wave in a submicrometer semiconductor device," *IEEE Trans. Electron Devices*, vol. 38, pp. 392–398, Feb. 1991.
[28] C. Ringhofer, C. Schmeiser, and A. Zwirchmayer, "Moment methods for the semiconductor boltzmann equation in bounded position domains," *SIAM J. Numer. Anal.*, vol. 39, no. 3, pp. 1078–1095, 2001.
[29] H. Struchtrup, "Extended moment method for electrons in semiconductors," *Physica A*, vol. 275, pp. 229–255, 2000.
[30] T. Grasser, H. Kosina, C. Heitzinger, and S. Selberherr, "Characterization of the hot electron distribution function using six moments," *J. Appl. Phys.*, vol. 91, no. 6, pp. 3869–3879, 2002.
[31] M. Nekovee, B. J. Geurts, H. M. J. Boots, and M. F. H. Schuurmans, "Failure of extended moment equation approaches to describe ballistic transport in submicron structures," *Phys. Rev. B*, vol. 45, no. 10, pp. 6643–6651, 1992.
[32] S. F. Liotta and H. Struchtrup, "Moment equations for electrons in semiconductors: Comparison of spherical harmonics and full moments," *Solid-State Electron.*, vol. 44, pp. 95–103, 2000.
[33] R. Stratton, "Semiconductor current-flow equations (Diffusion and Degeneracy)," *IEEE Trans. Electron Devices*, vol. ED-19, pp. 1288–1292, Dec. 1972.
[34] P. T. Landsberg and S. A. Hope, "Two formulations of semiconductor transport equations," *Solid-State Electron.*, vol. 20, pp. 421–429, 1977.
[35] P. T. Landsberg, "$D\mathrm{grad}\nu$ or $\mathrm{grad}(D\nu)$?," *J. Appl. Phys.*, vol. 56, no. 4, pp. 1119–1122, 1984.
[36] T.-W. Tang and H. Gan, "Two formulations of semiconductor transport equations based on spherical harmonic expansion of the boltzmann transport equation," *IEEE Trans. Electron Devices*, vol. 47, pp. 1726–1732, Sept. 2000.
[37] M. Ieong and T.-W. Tang, "Influence of hydrodynamic models on the prediction of submicrometer device characteristics," *IEEE Trans. Electron Devices*, vol. 44, pp. 2242–2251, Dec. 1997.
[38] S. Ramaswamy and T.-W. Tang, "Comparison of semiconductor transport models using a Monte Carlo consistency check," *IEEE Trans. Electron Devices*, vol. 41, pp. 76–83, Jan. 1994.
[39] M. C. Vecchi and L. G. Reyna, "Generalized energy transport models for semiconductor device simulation," *Solid-State Electron.*, vol. 37, no. 10, pp. 1705–1716, 1994.
[40] H. Gan and T.-W. Tang, "A new method for extracting carrier mobility from Monte Carlo device simulation," *IEEE Trans. Electron Devices*, vol. 48, pp. 399–401, Feb. 2001.
[41] G. Baccarani, M. Rudan, R. Guerrieri, and P. Ciampolini, "Physical models for numerical device simulation," in *Process and Device Modeling*, W. Engl, Ed. Amsterdam, The Netherlands: North-Holland, 1986, vol. 1, Advances in CAD for VLSI, pp. 107–158.
[42] W. Hänsch, *The Drift Diffusion Equation and Its Application in MOSFET Modeling*. New York: Springer-Verlag, 1991.
[43] T.-W. Tang, S. Ramaswamy, and J. Nam, "An improved hydrodynamic transport model for silicon," *IEEE Trans. Electron Devices*, vol. 40, pp. 1469–1476, Aug. 1993.
[44] T.-W. Tang, X. Wang, H. Gan, and M. K. Leong, "An analytic expression of thermal diffusion coefficient for the hydrodynamic simulation of semiconductor devices," *VLSI Design*, vol. 13, no. 1-4, pp. 131–134, 2000.
[45] S. Selberherr, "MOS device modeling at 77 K," *IEEE Trans. Electron Devices*, vol. 36, pp. 1464–1474, Aug. 1989.
[46] A. Leone, A. Gnudi, and G. Baccarani, "Hydrodynamic simulation of semiconductor devices operating at low temperature," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 1400–1408, Nov. 1994.
[47] A. H. Marshak and C. M. VanVliet, "Electrical current and carrier density in degenerate materials with nonuniform band structure," *Proc. IEEE*, vol. 72, pp. 148–164, Feb. 1984.
[48] R. Thoma, A. Emunds, B. Meinerzhagen, H. Peifer, and W. L. Engl, "A generalized hydrodynamic model capable of incorporating Monte Carlo results," in *Proc. Int. Electron Devices Meeting*, 1989, pp. 139–142.
[49] A. D. Sadovnikov and D. J. Roulston, "A study of the influence of hydrodynamic model effects on characteristics of silicon bipolar transistors," *COMPEL*, vol. 12, no. 4, pp. 245–262, 1993.
[50] C. Jungeman, B. Neinhüs, and B. Meinerzhagen, "Comparative study for electron transit times evaluated by DD, HD, and MC device simulation for a SiGe HBT," *IEEE Trans. Electron Devices*, vol. 48, pp. 2216–2220, Oct. 2001.
[51] T. J. Bordelon, X.-L. Wang, C. M. Maziar, and A. F. Tasch, "Accounting for bandstructure effects in the hydrodynamic model: A first-order approach for silicon device simulation," *Solid-State Electron.*, vol. 35, no. 2, pp. 131–139, 1992.
[52] ——, "An efficient nonparabolic formulation of the hydrodynamic model for silicon device simulation," in *Proc. Int. Electron Devices Meeting*, 1990, pp. 353–356.
[53] D. Chen, E. C. Kan, U. Ravaioli, C.-W. Shu, and R. W. Dutton, "An improved energy transport model including nonparabolicity and nonmaxwellian distribution effects," *IEEE Electron Device Lett.*, vol. 13, pp. 26–28, Jan. 1992.
[54] D. Xu, T.-W. Tang, and S. S. Kucherenko, "Time-dependent solution of a full hydrodynamic model including convective terms and viscous effects," *VLSI Design*, vol. 6, no. 1-4, pp. 173–176, 1998.
[55] H. Gan, "Re-examination of the hot-carrier transport model using spherical harmonic expansion of the Boltzmann transport equation," M.S. thesis, Univ. Massachusetts Amherst, Amherst, 1999.
[56] A. W. Smith and A. Rohatgi, "Non-isothermal extension of the Scharfetter-Gummel technique for hot carrier transport in heterostructure simulations," *IEEE Trans. Computer-Aided Design*, vol. 12, pp. 1515–1523, Oct. 1993.
[57] A. W. Smith and K. F. Brennan, "Comparison of nonparabolic hydrodynamic simulations for semiconductor devices," *Solid-State Electron.*, vol. 39, no. 7, pp. 1055–1063, 1996.
[58] A. Anile and V. Romano. (1999) Hydrodynamical modeling of charge carrier transport in semiconductors. Summer School on Industrial Mathematics, IST Lisboa, Lisbon, Portugal. [Online]. Available: http://www.dipmat.unict.it/~anile/preprint.html
[59] O. Muscato, "The onsager reciprocity principle as a check of consistency for semiconductor carrier transport models," *Physica A*, vol. 289, pp. 422–458, 2001.
[60] M. Trovato and L. Reggiani, "Maximum entropy principle within a total energy scheme: Application to hot-carrier transport in semiconductors," *Phys. Rev. B*, vol. 61, no. 24, pp. 16 667–16 681, 2000.

[61] R. Thoma and W. L. Engl, "A new approach for the derivation of macroscopic balance equations beyond the relaxation time approximation," in *Simulation of Semiconductor Devices and Processes*, W. Fichtner and D. Aemmer, Eds. Konstanz, Germany: Hartung-Gorre, 1991, vol. 4, pp. 185–194.
[62] T.-W. Tang and M. K. Ieong, "Why are there so many different hydrodynamic transport models in semiconductor device simulation?," in *Proc. 2nd NASA Device Modeling Workshop*, 1997, pp. 127–136.
[63] A. H. Marshak and K. M. van Vliet, "Electrical current in solids with position-dependent band structure," *Solid-State Electron.*, vol. 21, pp. 417–427, 1978.
[64] E. M. Azoff, "Closed-form method for solving the steady-state generalized energy-momentum conservation equations," *COMPEL*, vol. 6, no. 1, pp. 25–30, 1987.
[65] D. L. Woolard, R. J. Trew, and M. A. Littlejohn, "Hydrodynamic hot-electron transport model with Monte Carlo–generated transport parameters," *Solid-State Electron.*, vol. 31, no. 3/3, pp. 571–574, 1988.
[66] D. L. Woolard, H. Tian, R. J. Trew, M. A. Littlejohn, and K. W. Kim, "Hydrodynamic electron-transport model: Nonparabolic corrections to the streaming terms," *Phys. Rev. B*, vol. 44, no. 20, pp. 11 119–11 132, 1991.
[67] D. L. Woolard, H. Tian, M. A. Littlejohn, K. W. Kim, R. J. Trew, M. K. Ieong, and T.-W. Tang, "Construction of higher-moment terms in the hydrodynamic electron-transport model," *J. Appl. Phys.*, vol. 74, no. 10, pp. 6197–6207, 1993.
[68] R. A. Stewart and J. N. Churchill, "A fully nonparabolic hydrodynamic model for describing hot electron transport in GaAs," *Solid-State Electron.*, vol. 33, no. 7, pp. 819–829, 1990.
[69] R. A. Stewart, L. Ye, and J. N. Churchill, "Improved relaxation-time formulation of collision terms for two-band hydrodynamic models," *Solid-State Electron.*, vol. 32, no. 6, pp. 497–502, 1989.
[70] C. L. Wilson, "Hydrodynamic carrier transport in semiconductors with multiple band minima," *IEEE Trans. Electron Devices*, vol. 35, pp. 180–187, 1988.
[71] J. C. Cao and X. L. Lei, "Simulation of an $Al_xGa_{1-x}As$ ballistic diode using multivalley nonparabolic hydrodynamic balance equations," *Solid-State Electron.*, vol. 41, no. 11, pp. 1781–1785, 1997.
[72] M.-K. Ieong, "A multi-valley hydrodynamic transport model for GaAs extracted from self-consistent Monte Carlo data," M.S. thesis, Univ. Massachusetts Amherst, Amherst, 1993.
[73] M. Ieong and T.-W. Tang, "Transport coefficients for a GaAs hydrodynamic model extracted from inhomogenous Monte Carlo calculations," in *Proc. Int. Workshop Computational Electronics*, 1993, pp. 65–69.
[74] G. Wolokin and J. Frey, "Overshoot effects in the relaxation time approximation," in *Proc. 8th Int. Conf. Numerical Analysis Semiconductor Devices and Integrated Circuits*, 1992, pp. 107–108.
[75] A. Abramo and C. Fiegna, "Electron energy distributions in silicon at low applied voltages and high electric fields," *J. Appl. Phys.*, vol. 80, no. 2, pp. 889–893, 1996.
[76] P. G. Scrobohaci and T.-W. Tang, "Modeling of the hot electron subpopulation and its application to impact ionization in submicron silicon devices-part I: Transport equations," *IEEE Trans. Electron Devices*, vol. 41, pp. 1197–1205, July 1994.
[77] J.-G. Ahn, C.-S. Yao, Y.-J. Park, H.-S. Min, and R. W. Dutton, "Impact ionization modeling using simulation of high energy tail distributions," *IEEE Electron Device Lett.*, vol. 15, pp. 348–350, Sept. 1994.
[78] T. J. Bordelon, V. M. Agostinelli, X.-L. Wang, C. M. Maziar, and A. F. Tasch, "Relaxation time approximation and mixing of hot and cold electron populations," *Electron. Lett.*, vol. 28, no. 12, pp. 1173–1175, 1992.
[79] T.-W. Tang and J. Nam, "A simplified impact ionization model based on the average energy of hot-electron subpopulation," *IEEE Electron Device Lett.*, vol. 19, pp. 201–203, June 1998.
[80] R. K. Cook and J. Frey, "An efficient technique for two-dimensional simulation of velocity overshoot effects in Si and GaAs devices," *COMPEL*, vol. 1, no. 2, pp. 65–87, 1982.
[81] K. Sonoda, S. T. Dunham, M. Yamaji, K. Taniguchi, and C. Hamaguchi, "Impact ionization model using average energy and average square energy of distribution function," *Jap. J. Appl. Phys.*, vol. 35, no. 2B, pp. 818–825, 1996.
[82] K. Sonoda, M. Yamaji, K. Taniguchi, C. Hamaguchi, and S. T. Dunham, "Moment expansion approach to calculate impact ionization rate in submicron silicon devices," *J. Appl. Phys.*, vol. 80, no. 9, pp. 5444–5448, 1996.
[83] D. Chen, Z. Yu, K.-C. Wu, R. Goossens, and R. W. Dutton, "Dual energy transport model with coupled lattice and carrier temperatures," in *Simulation of Semiconductor Devices and Processes*, S. Selberherr, H. Stippel, and E. Strasser, Eds. Berlin, Germany: Springer-Verlag, 1993, vol. 5, pp. 157–160.
[84] A. Benvenuti, W. M. Coughran, and M. R. Pinto, "A thermal-fully hydrodynamic model for semiconductor devices and applications to III–V HBT simulation," *IEEE Trans. Electron Devices*, vol. 44, pp. 1349–1359, Sept. 1997.
[85] B. J. Geurts, "An extended Scharfetter-Gummel scheme for high order momentum equations," *COMPEL*, vol. 10, no. 3, pp. 179–194, 1991.
[86] B. Pejčinović, H. H. K. Tang, J. L. Egley, L. R. Logan, and G. R. Srinivasan, "Two-dimensional tensor temperature extension of the hydrodynamic model and its applications," *IEEE Trans. Electron Devices*, vol. 42, pp. 2147–2155, Dec. 1995.
[87] G. Baccarani and M. R. Wordeman, "An investigation of steady-state velocity overshoot in silicon," *Solid-State Electron.*, vol. 28, no. 4, pp. 407–416, 1985.
[88] M. A. Stettler, M. A. Alam, and M. S. Lundstrom, "A critical examination of the assumptions underlying macroscopic transport equations for silicon devices," *IEEE Trans. Electron Devices*, vol. 40, pp. 733–740, Mar. 1993.
[89] W. Hänsch and M. Miura-Mattausch, "The hot-electron problem in small semiconductor devices," *J. Appl. Phys.*, vol. 60, no. 2, pp. 650–656, 1986.
[90] S.-C. Lee, T.-W. Tang, and D. H. Navon, "Transport models for MBTE," in *NASECODE VI—Numerical Analysis of Semiconductor Devices and Integrated Circuits*, J. J. H. Miller, Ed. Dublin, Ireland: Boole, 1989, pp. 261–265.
[91] K. Rahmat, J. White, and D. A. Antoniadis, "Computation of drain and substrate current in ultra-short-channel nMOSFET's using the hydrodynamic model," *IEEE Trans. Computer-Aided Design*, vol. 12, pp. 817–824, June 1993.
[92] V. M. Agostinelli, T. J. Bordelon, X. L. Wang, C. F. Yeap, C. M. Maziar, and A. F. Tasch, "An energy-dependent two-dimensional substrate current model for the simulation of submicrometer MOSFET's," *IEEE Electron Device Lett.*, vol. 13, pp. 554–556, Nov. 1992.
[93] M. V. Fischetti, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures-part I: Homogeneous transport," *IEEE Trans. Electron Devices*, vol. 38, pp. 634–649, Mar. 1991.
[94] B. Gonzales, V. Palankovski, H. Kosina, A. Hernandez, and S. Selberherr, "An energy relaxation time model for device simulation," *Solid-State Electron.*, vol. 43, pp. 1791–1795, 1999.
[95] K. Hasnat, C.-F. Yeap, S. Jallepalli, S. A. Hareland, W.-K. Shih, V. M. Agostinelli, A. F. Tasch, and C. M. Maziar, "Thermionic emission model of electron gate current in submicron NMOSFET's," *IEEE Trans. Electron Devices*, vol. 44, pp. 129–138, Jan. 1997.
[96] D. Chen, E. Sangiorgi, M. R. Pinto, E. C. Kan, U. Ravaioli, and R. W. Dutton, "Analysis of spurious velocity overshoot in hydrodynamic simulations," in *Proc. Int. Numerical Modeling Processes and Devices Integrated Circuits (NUPAD IV)*, May 1992, pp. 109–114.
[97] T. Grasser, H. Kosina, and S. Selberherr, "Investigation of spurious velocity overshoot using Monte Carlo data," *Appl. Phys. Lett.*, vol. 79, no. 12, pp. 1900–1903, 2001.
[98] S. E. Laux and R. J. Lomax, "Effect of mesh spacing on static negative resistance in GaAs MESFET simulation," *IEEE Trans. Electron Devices*, vol. ED-28, pp. 120–122, Jan. 1981.
[99] V. Axelrad, "Grid quality and its influence on accuracy and convergence in device simulation," *IEEE Trans. Computer-Aided Design*, vol. 17, pp. 149–157, Feb. 1998.
[100] M. Duane, "TCAD needs and applications from a user's perspective," *IEICE Trans. Electron.*, vol. E82-C, no. 6, pp. 976–982, 1999.
[101] A. M. Anile, C. Maccora, and R. M. Pidatella, "Simulation of $n^+ - n - n^+$ devices by a hydrodynamic model: Subsonic and supersonic flows," *COMPEL*, vol. 14, no. 1, pp. 1–18, 1995.
[102] E. Fatemi, J. Jerome, and S. Osher, "Solution of the hydrodynamic device model using high-order nonoscillatory shock capturing algorithms," *IEEE Trans. Computer-Aided Design*, vol. 10, pp. 232–244, Feb. 1991.
[103] E. Thomann and F. Odeh, "On the well-posedness of the two-dimensional hydrodynamic model for semiconductor devices," *COMPEL*, vol. 9, no. 1, pp. 45–57, 1990.

[104] M. Rudan and F. Odeh, "Multi-dimensional discretization scheme for the hydrodynamic model of semiconductor devices," *COMPEL*, vol. 5, no. 3, pp. 149–183, 1986.
[105] E. Fatemi and F. Odeh, "Upwind finite difference solution of boltzmann equation applied to electron transport in semiconductor devices," *J. Comput. Phys.*, vol. 108, no. 2, pp. 209–217, 1993.
[106] G. Jiang and C.-W. Shu, "Efficient implementation of weighted ENO schemes," *J. Comput. Phys.*, vol. 126, no. 2, pp. 202–228, 1996.
[107] D. S. Balsara and C.-W. Shu, "Monotonicity preserving weighted essentially nonoscillatory schemes with increasingly high order of accuracy," *J. Comput. Phys.*, vol. 160, no. 2, pp. 405–452, 2000.
[108] B. Cockburn and C.-W. Shu, "The local discontinuous Galerkin method for time-dependent covection- diffusion systems," *SIAM J. Numer. Anal.*, vol. 35, no. 6, pp. 2440–2463, 1998.
[109] N. R. Aluru, K. H. Law, P. M. Pinsky, and R. W. Dutton, "An analysis of the hydrodynamic semiconductor device model—boundary conditions and simulations," *COMPEL*, vol. 14, no. 2/3, pp. 157–185, 1995.
[110] L.-M. Yeh, "Well-posedness of the hydrodynamic model for semiconductors," *Math. Meth. Appl. Sci.*, vol. 19, no. 18, pp. 1489–1507, 1996.
[111] T.-W. Tang, "Extension of the Scharfetter-Gummel algorithm to the energy balance equation," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1912–1914, Dec. 1984.
[112] C. C. McAndrew, K. Singhal, and E. L. Heasell, "A consistent nonisothermal extension of the Scharfetter-Gummel stable difference approximation," *IEEE Electron Device Lett.*, vol. EDL-6, pp. 446–447, Sept. 1985.
[113] A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan, and G. Baccarani, "A new discretization strategy of the semiconductor equations comprising momentum and energy balance," *IEEE Trans. Computer-Aided Design*, vol. 7, pp. 231–242, Feb. 1988.
[114] W.-S. Choi, J.-G. Ahn, Y.-J. Park, H.-S. Min, and C.-G. Hwang, "A time dependent hydrodynamic device simulator SNU-2D with new discretization scheme and algorithm," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 899–908, July 1994.
[115] A. Gnudi and F. Odeh, "An efficient discretization scheme for the energy-continuity equation in semiconductors," in *Simulation of Semiconductor Devices and Processes*, G. Baccarani and M. Rudan, Eds. Bologna, Italy: Technoprint, 1988, vol. 3, pp. 387–390.
[116] T.-W. Tang and M.-K. Ieong, "Discretization of flux densities in device simulations using optimum artificial diffusivity," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 1309–1315, Nov. 1995.
[117] B. Meinerzhagen, K. H. Bach, I. Bork, and W. L. Engl, "A new highly efficient nonlinear relaxation scheme for hydrodynamic MOS simulations," in *Proc. Int. Numerical Modeling Processes and Devices Integrated Circuits (NUPAD IV)*, 1992, pp. 91–96.
[118] H. K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculations," *IEEE Trans. Electron Devices*, vol. ED-11, pp. 455–465, Oct. 1964.
[119] Z. Yu, R. W. Dutton, and R. A. Kiehl, "Circuit/device modeling at the quantum level," *IEEE Trans. Electron Devices*, vol. 47, pp. 1819–1825, Oct. 2000.
[120] K. Banoo and M. S. Lundstrom, "Electron transport in a model Si transistor," *Solid-State Electron.*, vol. 44, pp. 1689–1695, 2000.
[121] M. Tomizawa, K. Yokoyama, and A. Yoshii, "Nonstationary carrier dynamics in quarter-micron Si MOSFET's," *IEEE Trans. Computer-Aided Design*, vol. 7, pp. 254–258, Feb. 1988.
[122] J. W. Slotboom, G. Streutker, M. J. v. Dort, P. H. Woerlee, A. Pruijmboom, and D. J. Gravesteijn, "Non-local impact ionization in silicon devices," in *Proc. Int. Electron Devices Meeting*, 1991, pp. 127–130.
[123] R. K. Cook and J. Frey, "Two-dimensional numerical simulation of energy transport effects in Si and GaAs MESFET's," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 970–977, June 1982.
[124] D. Chang and J. G. Fossum, "Simplified energy-balance model for pragmatic multi-dimensional device simulation," *Solid-State Electron.*, vol. 41, no. 11, pp. 1795–1802, 1997.
[125] Q. Lin, N. Goldsman, and G.-C. Tai, "Highly stable and routinely convergent 2-Dimensional hydrodynamic device simulation," *Solid-State Electron.*, vol. 37, no. 2, pp. 359–371, 1994.
[126] K. K. Thornber, "Current equations for velocity overshoot," *IEEE Electron Device Lett.*, vol. EDL-3, pp. 69–70, Mar. 1982.
[127] D. Chen, E. C. Kan, and U. Ravaioli, "An analytical formulation of the length coefficient for the augmented drift-diffusion model including velocity overshoot," *IEEE Trans. Electron Devices*, vol. 38, pp. 1484–1490, June 1991.
[128] M. G. Ancona, "Hydrodynamic models of semiconductor electron transport at high fields," *VLSI Design*, vol. 3, no. 2, pp. 101–114, 1995.
[129] H. Kosina and S. Selberherr, "A hybrid device simulator that combines Monte Carlo and drift-diffusion analysis," *IEEE Trans. Computer-Aided Design*, vol. 13, pp. 201–210, Feb. 1994.
[130] *MINIMOS-NT 2.0 User's Guide*, Institute for Microelectronics, Technical Univ. Vienna, Vienna, Austria, 2002.
[131] I. Bork, C. Jungemann, B. Meinerzhagen, and W. L. Engl, "Influence of heat flux on the accuracy of hydrodynamic models for ultra-short Si MOSFET's," in *Proc. Int. Workshop Numerical Modeling Processes and Devices Integrated Circuits (NUPAD V)*, 1994, pp. 63–66.
[132] M. Gritsch, H. Kosina, T. Grasser, and S. Selberherr, "Revision of the standard hydrodynamic transport model for SOI simulation," *IEEE Trans. Electron Devices*, vol. 49, pp. 1814–1820, Oct. 2002.
[133] C. Jungeman and B. Meinerzhagen, "On the applicability of nonself-consistent Monte Carlo device simulations," *IEEE Trans. Electron Devices*, vol. 49, pp. 1072–1074, June 2002.

Tibor Grasser was born in Vienna, Austria, in 1970. He received the Diplomingenieur degree in communications engineering, the Ph.D. degree in technical sciences, and the *venia docendi* degree in microelectronics from the Technical University Vienna, Vienna, Austria, in 1995, 1999, and 2002, respectively.

He is currently an Associate Professor at the Institute for Microelectronics, Technical University Vienna. Since 1997, he has headed the MINIMOS-NT development group, working on the successor of the highly successful MINIMOS program. In 1997, he was also a Visiting Research Engineer at Hitachi Ltd., Tokyo, Japan. In Summer 2001, he was also a Visiting Researcher in the Alpha Development Group, Compaq Computer Corp., Shrewsbury, MA. His current research interests include circuit and device simulation, device modeling, and physical and software aspects in general.

Ting-Wei Tang (Fellow, IEEE) received the B.S. degree in electrical engineering from the National Taiwan University, Taiwan, in 1957, and the M.Sc. and Ph.D. degrees in engineering from Brown University, Providence, RI, in 1961 and 1964, respectively.

From 1963 to 1964, he was an Instructor of Electrical Engineering and from 1964 to 1968 he was an Assistant Professor of Aerospace Engineering at the University of Connecticut, Storrs. From 1968 to 1974, he was an Associate Professor in the Electrical and Computer Engineering Department at the University of Massachusetts, Amherst; since 1974, he has been a Professor in that department. In Spring 1975, he was also a Visiting Professor at Oxford University, Oxford, U.K. In Fall 1990, he was also Visiting Samsong Professor of Electronics Engineering at Seoul National University, Seoul, South Korea. His past research includes electromagnetic theory and plasma physics. His current research interest is in semiconductor device physics and numerical modeling of semiconductor devices. He teaches a graduate-level course in which students gain hands-on experience in developing an elementary device simulation software.

Dr. Tang received an Outstanding Senior Faculty Award and an Outstanding Teaching Award from the College of Engineering, University of Massachusetts, in 1989 and 1991, respectively, and the Chancellor's Medal from the University of Massachusetts in 2000. He was a Keynote Speaker at the 1993 Symposium on Semiconductor Modeling and Simulation in Taipei, Taiwan; the 1995 Fourth International Seminar on Simulation of Devices and Technologies at Kruger National Park, South Africa; and the 1999 Seminar Technology CAD Workshop and Exhibition in Hsinchu, Taiwan.

Hans Kosina (Member, IEEE) received the Diplomingenieur degree in electrical engineering, the Ph.D. degree in technical sciences, and the *venia docendi* degree in microelectronics from the Technical University Vienna, Vienna, Austria, in 1987, 1992, and 1998, respectively.

He is currently an Associate Professor at the Institute for Microelectronics, Technical University Vienna. His current research interests include modeling of carrier transport and quantum effects in semiconductor devices, development of novel Monte Carlo algorithms, and computer aided engineering in ultralarge-scale integration technology.

Siegfried Selberherr (Fellow, IEEE) was born in Klosterneuburg, Austria, in 1955. He received the Diplomingenieur degree in electrical engineering, the Ph.D. degree in technical sciences, and the *venia docendi* degree in computer-aided design from the Technical University Vienna, Vienna, Austria, in 1978, 1981, and 1984, respectively.

He is currently the Head of the Institute for Microelectronics, Technical University Vienna, and, since 1999, has been Dean of the Faculty of Electronic and Information Technology, Technical University Vienna. His current research interests are modeling and simulation of problems for microelectronics engineering.