Limitations of Hydrodynamic and Energy-Transport Models

T. Grasser and S. Selberherr

Institute for Microelectronics, TU Vienna, Gusshausstr. 27–29, A-1040 Vienna, Austria
Phone +43-1-58801-36023, FAX +43-1-58801-36099, e-mail: Tibor.Grasser@iue.tuwien.ac.at

Accurate description of nonlocal hot-carrier phenomena in modern semiconductor devices is becoming very important considering the rapid reduction of the feature size. Frequently used carrier transport models are the traditional drift-diffusion model and extended models which also consider the average carrier energy as an independent solution variable. While the drift-diffusion model is not capable of handling nonlocal effects, the latter model group provides a more accurate description. Several variants of hydrodynamic and energy-transport models have been published so far. Recent results show, however, that the average energy is in many cases not sufficient for accurate modeling. Both the transport models themselves and the models for the physical parameters seem to be affected.

Introduction

Numerical simulation of carrier transport in semiconductor devices dates back to the famous work of Scharfetter and Gummel [1]. Since then the transport models have been continously refined and extended to more accurately capture transport phenomena occurring in modern semiconductor 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 fields inside the devices have increased. Large electric fields which rapidly change over small length scales give 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.

Traditionally, the drift-diffusion model [2] has been used to describe carrier transport in semiconductor devices. However, the drift-diffusion model assumes an equilibrium between carrier energy and electric field, which is no longer valid in modern devices. Extended models have been proposed which consider the carrier energy an independent solution variable [3, 4]. These models are capable of describing nonlocal and hot-carrier effects to a first order. Recent results, however, suggest that the average energy is in many cases not sufficient for accurate modeling. Both the transport models themselves and the models for the physical parameters are affected. In this article we review the most commonly used transport models and point out their most important limitations.

Boltzmann's Transport Equation

Transport equations used in semiconductor device simulation are normally derived from Boltzmann's transport equation which provides a semiclassical description of carrier transport. For a general inhomogeneous material with arbitrary band structure it reads [5]

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla_{\mathbf{r}} f + \frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f = C[f] \qquad (1)$$

Here, $\mathbf{u}$ is the group velocity, $\mathbf{F}$ the force exerted on the particles, and C the collision operator. For inclusion of quantum effects equations based on the

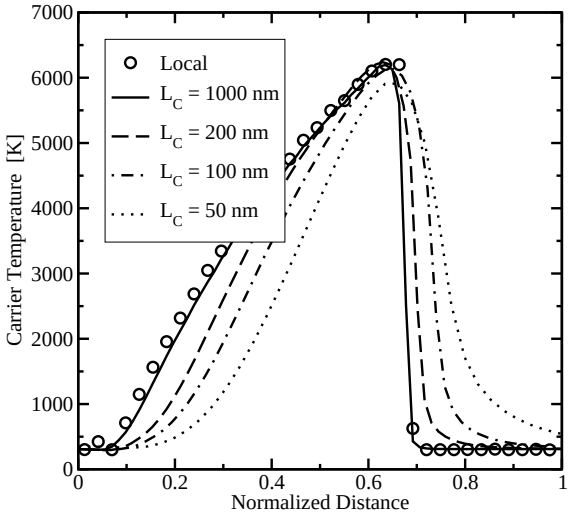

Fig. 1: The carrier temperature of comparable n^+-n-n^+ test structures with varying channel lengths where the spatial coordinates have been normalized to get an overlapping electric field.

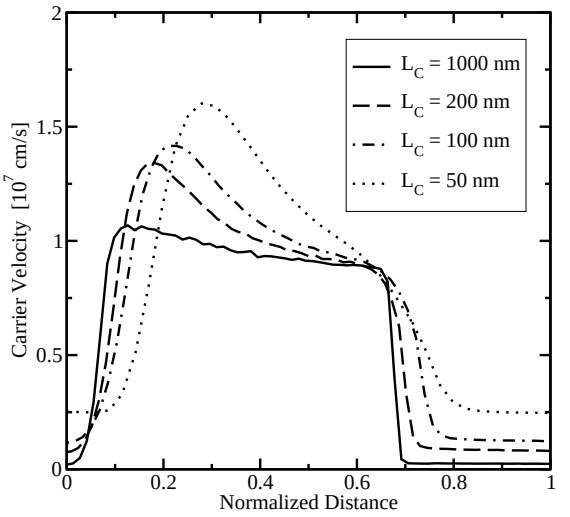

Fig. 2: A comparison of the average carrier velocities of comparable n^+-n-n^+ test structures. The velocity overshoot is caused by the nonlocality of the carrier temperature.

Wigner-Boltzmann equation have been considered [6]. Boltzmann's equation needs to be solved in the seven-dimensional phase space which is prohibitive for engineering applications. Monte Carlo simulations have been proven to give accurate results but are restrictively time consuming. Furthermore, if the distribution of high-energetic carriers is relevant, or if the carrier concentration is very low in specific regions of the device, Monte Carlo simulations tend to produce high variance in the results. Therefore, a common simplification is to investigate only some moments of the distribution function, such as the carrier concentration and the carrier temperature. These moments of the distribution function are typically defined as

$$\langle \Phi \rangle = \frac{1}{4\pi^3} \int \Phi f \ \mathrm{d}^3\mathbf{k} \tag{2}$$

with a suitable weight function $\Phi = \Phi(\mathbf{k})$.

The Drift-Diffusion Model

The drift-diffusion model is the simplest current transport model which can be derived from Boltzmann's transport equation by the method of moments [2] or from basic principles of irreversible thermodynamics [7]. It has been the working horse in industrial applications for over thirty years. Within the drift-diffusion model the well known continuity and current equations have to be solved which read in their static form

$$\nabla \cdot \mathbf{J} = \mathrm{q}R \tag{3}$$

$$\mathbf{J} = \mathrm{q}\mu\nu\mathbf{E} + \mu \mathrm{k_B} T_\mathrm{L} \nabla\nu \tag{4}$$

Here, μ denotes the electron mobility, T_L the lattice temperature, $\mathbf{E}$ the electric field, and R the recombination rate. The average energy w can be estimated via the local energy balance equation. This neglects the lag between the electric field and the average energy characterized by the energy relaxation time. One consequence of the lag is that the maximum energy can be much smaller than the one predicted by the local energy balance equation (Fig. 1). Furthermore, this lag gives rise to an overshoot in the carrier velocity because the mobility depends to first order on the energy and not on the electric field as shown in Fig. 2.

Therefore, modeling deep-submicron devices is becoming more and more problematic. Although successful reproduction of terminal characteristics of nano-scale MOS transistors has been reported [8], the values of the material parameters used significantly violate basic physical principles. In particular, the saturation velocity v_s had to be set to more than twice the value observed in bulk measurements.

The Full Hydrodynamic Model

The full hydrodynamic model was first derived by Bløtekjær [4]. In its original form the first three moments of Boltzmann's transport equation were considered. Closure was obtained by applying a heuristic model for the heat flux $\mathbf{Q}$ using Fourier's law. Furthermore, the band structure was assumed to be parabolic and the tensor quantities were approximated by scalars. The resulting equation set reads [4]

$$\nabla \cdot \mathbf{J} = \mathrm{q}R \tag{5}$$

$$\mathbf{J} - \frac{\tau_m}{\mathrm{q}} \nabla \cdot \left(\mathbf{J} \otimes \frac{\mathbf{J}}{n} \right) = \mathrm{q}n\mu\mathbf{E} + \mu \mathrm{k_B} \nabla (nT_n) \tag{6}$$

$$\nabla \cdot \mathbf{S} = \mathbf{E} \cdot \mathbf{J} - n \frac{w - w_0}{\tau_{\mathcal{E}}} \tag{7}$$

$$\mathbf{S} = -\frac{1}{\mathrm{q}}(w + \mathrm{k_B}T_n)\mathbf{J} - \kappa(T_n)\nabla T_n \tag{8}$$

where τ_m is the momentum relaxation time. The additional parameters are the energy relaxation time $\tau_{\mathcal{E}}$ and the thermal conductivity κ. For the latter an empirical relation in analogy to the Wiedemann-Franz law is used with a correction factor p.

$$\kappa(T_n) = (\tfrac{5}{2} - p) \left(\frac{\mathrm{k_B}}{\mathrm{q}} \right)^2 \mathrm{q}\mu n T_n \tag{9}$$

This equation system is similar to the Euler equations of gas dynamics with the addition of a heat conduction term and the collision terms. Thus, the electron gas has a sound speed and the electron flow may be either subsonic or supersonic [9]. In the case of supersonic flow, electron shock waves will in general develop inside the device. These shock waves occur at either short length scales or at low temperatures. Furthermore, the traditionally applied Scharfetter-Gummel [1] discretization scheme and its extensions cannot be used for this type of equation, which makes handling of the full hydrodynamic model quite difficult [9, 10].

The Energy-Transport Model

As the closure of the full hydrodynamic model has been shown to be problematic, the fourth moment of Boltzmann's equation is added to give a more accurate description for the energy flux $\mathbf{S}$. For the closure of the equation system a heated Maxwellian distribution is generally assumed [11]. Since the resulting equation system is difficult to handle, simplifications are generally considered. The four-moments energy-transport model is obtained by the simplification of the four-moments hydrodynamic model: The convective term

$$\frac{\tau_m}{\mathrm{q}} \nabla \cdot \left(\mathbf{J} \otimes \frac{\mathbf{J}}{n} \right) \tag{10}$$

in the current relation, the corresponding convective term in the energy flux relation, and the contribution of the kinetic energy to the total carrier energy are neglected.

$$w = \frac{m\mathbf{v}^2}{2} + \frac{3}{2}\mathrm{k_B}T_n \approx \frac{3}{2}\mathrm{k_B}T_n \tag{11}$$

This gives the four-moments energy-transport model which reads

$$\nabla \cdot \mathbf{J} = \mathrm{q}R \tag{12}$$

$$\mathbf{J} = \mathrm{q}n\mu\mathbf{E} + \mu\mathrm{k_B}\nabla(nT_n) \tag{13}$$

$$\nabla \cdot \mathbf{S} = \mathbf{E} \cdot \mathbf{J} - n\tfrac{3}{2}\mathrm{k_B}\frac{T_n - T_\mathrm{L}}{\tau_{\mathcal{E}}} + G_{\mathcal{E}_\nu} \tag{14}$$

$$\mathbf{S} = -\tfrac{5}{2}\frac{\mu_S}{\mu}\frac{\mathrm{k_B}T_n}{\mathrm{q}}\mathbf{J} - \tfrac{5}{2}\frac{\mu_S}{\mu}\left(\frac{\mathrm{k_B}}{\mathrm{q}}\right)^2 \mathrm{q}\mu n T_n \nabla T_n \tag{15}$$

Here the energy flux mobility μ_S appears instead of the thermal conductivity in (8). Considering the different definitions for the mobilities the energy-transport model is equivalent to the energy-balance model proposed by Stratton [3], see for example [12].

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

Fundamental Problems

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

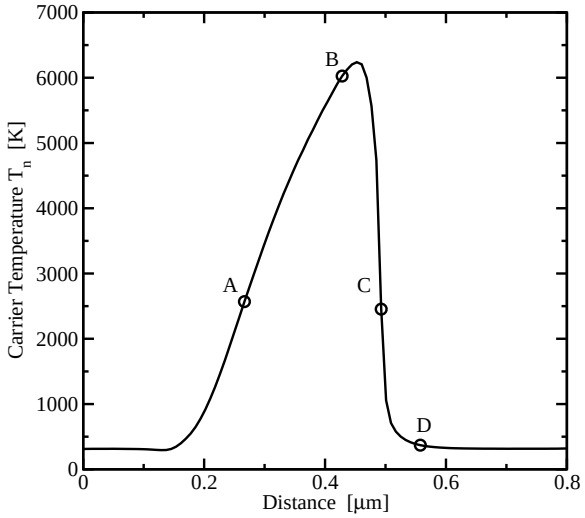

Fig. 3: Electron temperature inside an n^+-n-n^+ test structure with $L_c = 200\,\text{nm}$.

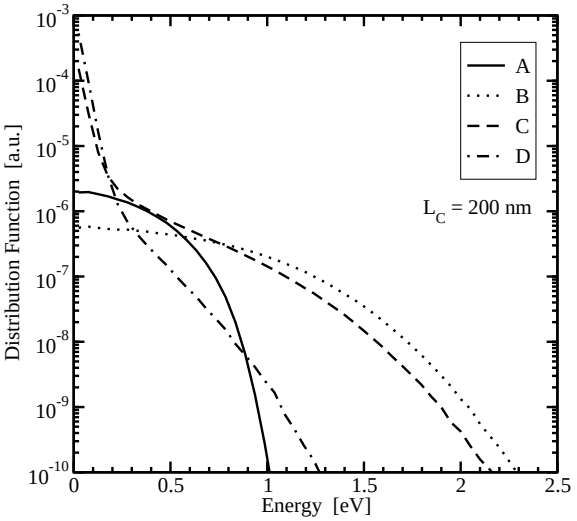

Fig. 4: The distribution function at the four characteristics points. Note that the average energies at the points A and C are the same whereas the distribution function looks completely different. Note the high-energy tail at point D where the carrier temperature is 370 K.

Closure

The method of moments transforms Boltzmann's equation into an equivalent, infinite set of equations. One of the severest 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. One approach to derive a suitable closure relation is to assume a distribution function and calculate the fourth order moment, where a heated Maxwellian shape is almost exclusively used. Ramaswami and Tang [13] gave a comparison of different closure relations available in literature.

In commercially available simulators, a heated Maxwellian shape is assumed for the distribution function to close the equation system. This assumption has been shown to be significantly violated in modern semiconductor devices. In general, the average carrier energy or carrier temperature is the only parameter that determines the shape of the heated Maxwellian distribution function. Monte Carlo simulation results of an n^+-n-n^+ test structure with $L_c = 200\,\text{nm}$ are shown in Fig. 3 and Fig. 4. Even though the average energy is the same at points A and C, the distribution function looks completely different in both cases. A heated Maxwellian distribution, which gives a straight line in a semi-logarithmic plot, is definitely a poor approximation throughout the whole device.

This error in the closure relation is one reason for the spurious overshoots observed in the velocity profiles obtained by energy-transport models [14]. As shown in Fig. 5, there is a spurious peak in the velocity profile which does not exist in the Monte Carlo simulation. For these simulations the relaxation times and mobilities have been taken from a coupled Monte Carlo 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 [15], the spurious peak is reduced. When in addition to the relaxation times also the closure is taken from the Monte Carlo simulation, the spurious peaks dissappear.

Tensor Quantities

An issue which has only been vaguely dealt with is the approximation of the tensors by scalar quantities, such as the trace of the tensors. For

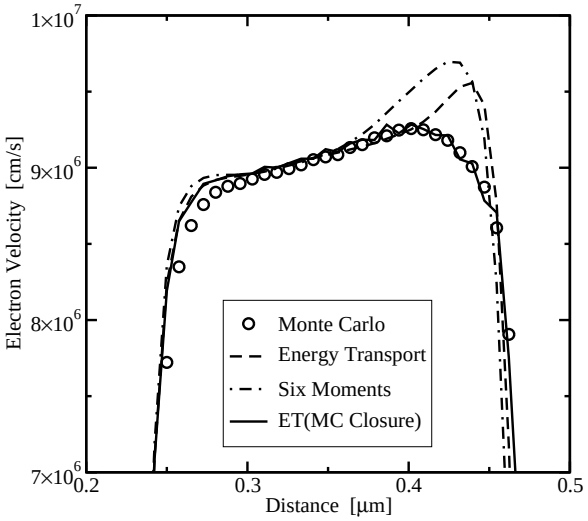

Fig. 5: Comparison of velocity profiles delivered by two transport models with Monte Carlo data. Both transport models use relaxation times and mobilities from the Monte Carlo simulation. In addition, when the energy-transport (ET) model is closed with the data from the Monte Carlo simulation, the spurious velocity overshoot disappears.

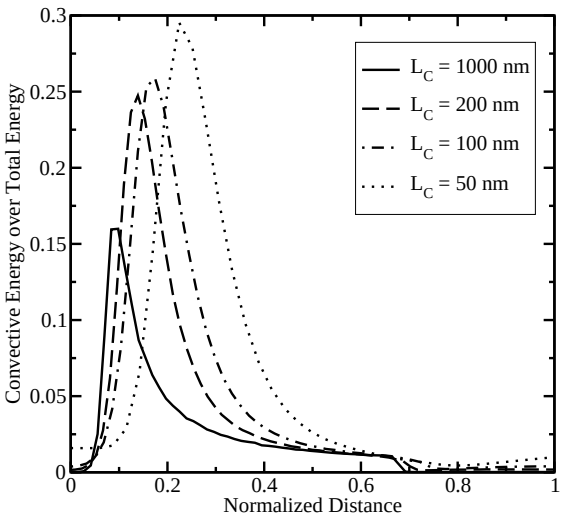

Fig. 6: Effect of approximation (11) on the carrier temperature and the deviation introduced by it obtained by a MC simulation of n^+-n-n^+ test structures.

example, the carrier mass and the carrier temperature are approximations introduced that way. One-dimensional simulations show [16] that the longitudinal temperature component is larger than the transverse temperature component. This indicates that the distribution function is elongated along the field direction and thus that the normally assumed equipartition of the energy is invalid. A rigorous approach has been taken by Pejčinović *et al.* [17] who model four components of the temperature tensor. They observed no significant difference between the scalar temperature and the trace of the temperature tensor for ballistic diodes and bipolar transistors but a 15 % difference for aggressively scaled MOSFETs in the linear region of the transfer characteristics.

Drift Energy versus Thermal Energy

Another common approximation is that the contribution of the drift energy to the total carrier energy is neglected [18]. As has been pointed out by Baccarani and Wordeman [19], the convective energy can reach values comparable to thermal energy. The error introduced by this approximation can be significant in the beginning of the channel where the carrier temperature is still low and a velocity overshoot is observed (Fig. 6). This effect has been studied in detail in [20].

Modeling of the Physical Parameters

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

Mobility

Two models for the energy dependence of the mobility are frequently used, the model after Baccarani and Wordeman [19]

$$\frac{\mu(T_n)}{\mu_0} = \frac{T_L}{T_n} \tag{16}$$

and the model after Hänsch [21, 22]

$$\frac{\mu(T_n)}{\mu_0} = \Big(1 - \frac{3}{2}\frac{\mu_0}{\tau_{\mathcal{E}} v_s^2}\Big(\frac{\mathrm{k_B} T_\mathrm{L}}{\mathrm{q}} + \frac{2}{5}\frac{S}{J}\Big)\Big)^{-1} \tag{17}$$

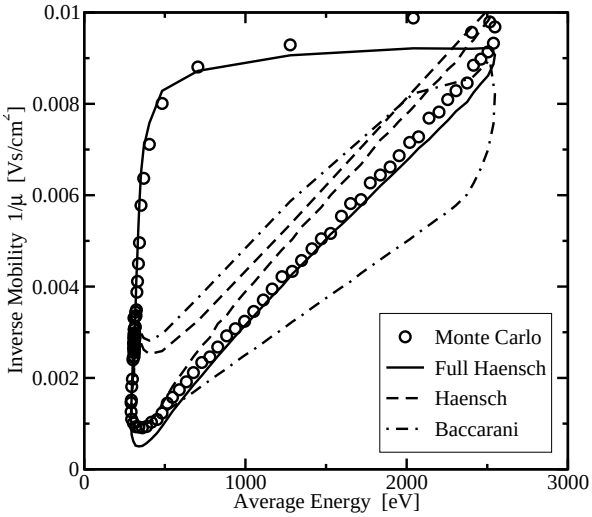

Fig. 7: Comparison of the mobility models data for an n^+-n-n^+ test structure with $L_c = 200\,\mathrm{nm}$.

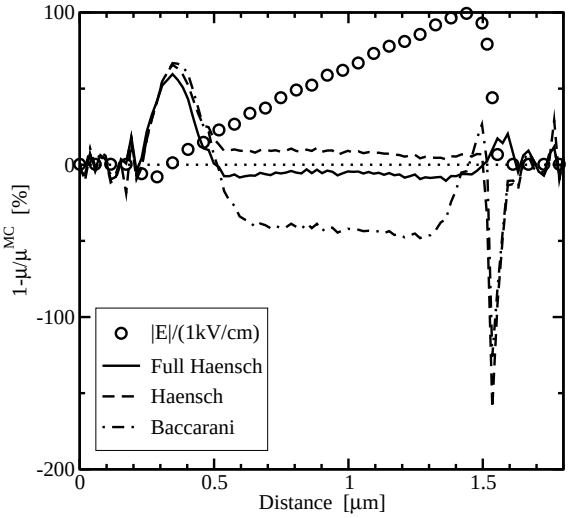

Fig. 9: Error in the analytical mobility models for an n^+-n-n^+ test structure with $L_c = 1000\,\mathrm{nm}$.

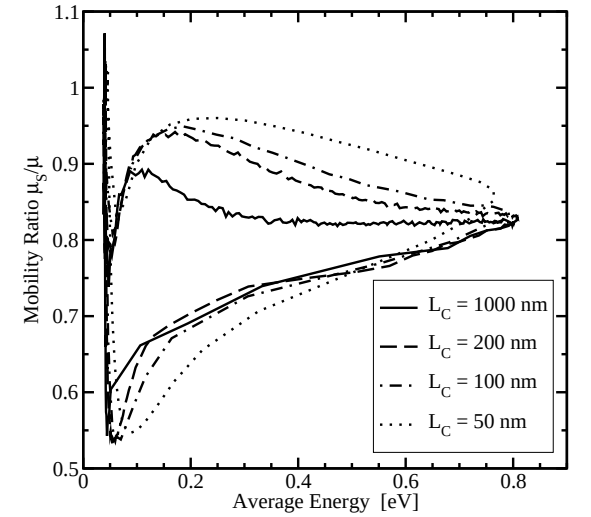

Fig. 8: Ratio of the mobilities obtained by Monte Carlo simulations for four n^+-n-n^+ test structures.

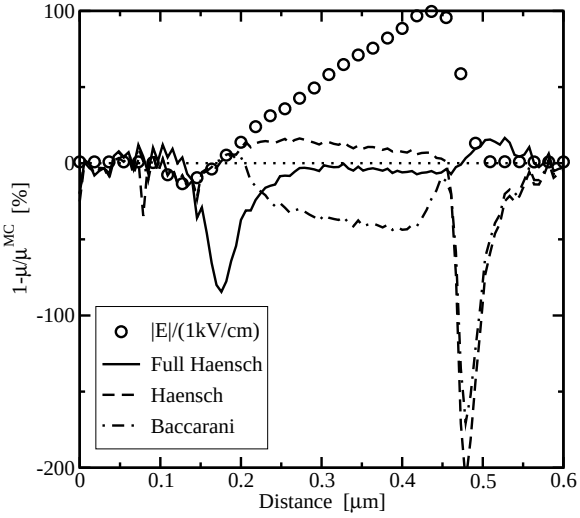

Fig. 10: Error in the analytical mobility models for an n^+-n-n^+ test structure with $L_c = 200\,\mathrm{nm}$.

For homogeneous materials $S/J = 5\mathrm{k_B}T_n/(2\mathrm{q})$ which can be used to simplify (17) as

$$\frac{\mu(T_n)}{\mu_0} = \left(1 + \frac{3}{2}\frac{\mu_0 \mathrm{k_B}}{\mathrm{q}\tau\mathcal{E}v_s^2}\Big(T_n - T_\mathrm{L}\Big)\right)^{-1} \qquad (18)$$

A comparison of these three expressions with Monte Carlo simulation results for an n^+-n-n^+ test structure with channel length $L_c = 200\,\mathrm{nm}$ is given in Fig. 7. The analytical expressions were evaluated using the data from the Monte Carlo simulation. Note that the temperature dependence of the inverse mobility is frequently plotted because of the expected linear dependence (16). The small hysteresis in the simplified Hänsch model and the Baccarani model is due to the doping dependence of the zero-field mobility μ_0.

The ratio of the mobilities μ_S/μ as a function of the carrier temperature is shown in Fig. 8 for four n^+-n-n^+ test structures. To obtain comparable

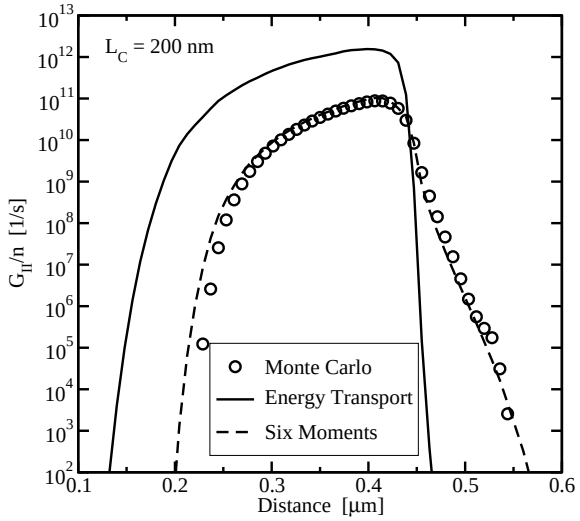

Fig. 11: Comparison of impact ionization rates predicted by an energy transport model and a six moments transport model with Monte Carlo data for an n^+-n-n^+ test structure with $L_c = 200\,\text{nm}$.

behavior the same doping profile has been used for all structures and the bias condition has been chosen to give a maximum electric field of 300 kV/cm. Note that in commercial device simulators the mobility ratio is normally assumed to be unity.

In Fig. 9 and Fig. 10 the error of the analytical models is shown for the two n^+-n-n^+ test structures. As has been pointed out in [16] expression (17) is the only expression which gives reasonable results in both increasing and decreasing field regions. However, at the beginning of the channel where the carrier temperature is still low, the mobility is considerably over- or underestimated. Furthermore, due to the quotient of the two vector quantities **S** and **J**, (17) is rather difficult to handle in a multidimensional device simulator.

Impact Ionization

Impact ionization is poorly described by models which use the local average energy as only parameter. In general, ionization rates obtained by local energy models start rising too early and fall off too sharply. Furthermore, local energy models considerably overestimate the ionization rates if not calibrated for the investigated device. In particular, local energy models cannot capture impact ionization caused by hot electrons in the drain because there the cold carriers dominate the average energy which is close to the equilibrium value. This is shown in Fig. 11 for an n^+-n-n^+ test structure with $L_c = 200\,\text{nm}$. In addition, simulation results obtained from a six moments transport model [15] are shown which considerably improve the accuracy of the model.

Several nonlocal models have been proposed [23] which are, however, both difficult to implement in a conventional device simulator and difficult to justify on a theoretical basis, especially for multidimensional problems.

Conclusions

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

References

[1] D. Scharfetter and H. Gummel, "Large-Signal Analysis of a Silicon Read Diode Oscillator," *IEEE Trans.Electron Devices*, vol. 16, no. 1, pp. 64–77, 1969.

[2] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Wien–New York: Springer, 1984.

[3] R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers," *Physical Review*, vol. 126, no. 6, pp. 2002–2014, 1962.

[4] K. Bløtekjær, "Transport Equations for Electrons in Two-Valley Semiconductors," *IEEE Trans.Electron Devices*, vol. 17, no. 1, pp. 38–47, 1970.

[5] D. Ferry, *Semiconductors*. New York: Macmillan, 1991.

[6] C. Gardner, "The Classical and Quantum Hydrodynamic Models," in *Proceedings of the International Workshop on Computational Electronics*, (University of Leeds), pp. 25–36, Aug. 1993.

[7] G. 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.

[8] J. Bude, "MOSFET Modeling into the Ballistic Regime," in *Simulation of Semiconductor Processes and Devices*, (Seattle, Washington, USA), pp. 23–26, Sept. 2000.

[9] C. 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.

[10] 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.

[11] T.-W. Tang, S. Ramaswamy, and J. Nam, "An Improved Hydrodynamic Transport Model for Silicon," *IEEE Trans.Electron Devices*, vol. 40, no. 8, pp. 1469–1476, 1993.

[12] 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, no. 9, pp. 1726–1732, 2000.

[13] S. Ramaswamy and T.-W. Tang, "Comparison of Semiconductor Transport Models Using a Monte Carlo Consistency Check," *IEEE Trans.Electron Devices*, vol. 41, no. 1, pp. 76–83, 1994.

[14] 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.

[15] 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.

[16] 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.

[17] B. Pejčinović, H. Tang, J. Egley, L. Logan, and G. Srinivasan, "Two-Dimensional Tensor Temperature Extension of the Hydrodynamic Model and Its Applications," *IEEE Trans.Electron Devices*, vol. 42, no. 12, pp. 2147–2155, 1995.

[18] R. 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.

[19] G. Baccarani and M. Wordeman, "An Investigation of Steady-State Velocity Overshoot in Silicon," *Solid-State Electron.*, vol. 28, no. 4, pp. 407–416, 1985.

[20] M. Stettler, M. Alam, and M. Lundstrom, "A Critical Examination of the Assumptions Underlying Macroscopic Transport Equations for Silicon Devices," *IEEE Trans.Electron Devices*, vol. 40, pp. 733–740, Apr. 1993.

[21] W. Hänsch, *The Drift Diffusion Equation and its Application in MOSFET Modeling*. Wien–New York: Springer, 1991.

[22] 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.

[23] C. Jungemann, R. Thoma, and L. Engl, "A Soft Threshold Lucky Electron Model for Efficient and Accurate Numerical Device Simulation," *Solid-State Electron.*, vol. 39, no. 7, pp. 1079–1086, 1996.