Advanced Transport Models for Sub-Micrometer Devices

T. Grasser*, C. Jungemann†, H. Kosina°, B. Meinerzhagen†, and S. Selberherr°

* Christian Doppler Laboratory for TCAD in Microelectronics
at the Institute for Microelectronics

† NST, TU Braunschweig, 38023 Braunschweig, Germany

° Institute for Microelectronics, TU Vienna
Gußhausstraße 27–29, A-1040 Wien, Austria
Email: Grasser@iue.tuwien.ac.at

Abstract

To overcome limitations of the classic drift-diffusion model, several higher-order transport models such as Stratton's energy-balance model and Bløtekjær's hydrodynamic model have been proposed. While for the drift-diffusion model the only transport parameter to be modeled correctly is the carrier mobility, there are many more significant parameters in higher-order models which cannot be directly taken from measurements. These parameters have to be chosen consistently, because otherwise the accuracy of the transport model cannot be properly assessed. Although a considerable number of parameter suggestions exists, it is not clear how well these models work when applied to submicron devices. Here we attempt a practical comparison as consistent as possible.

1 Introduction

We restrict our discussion to macroscopic transport models derived from Boltzmann's transport equation [1]. A simplified solution of the seven-dimensional Boltzmann equation is obtained by investigating only low order moments of the distribution function, such as the carrier concentration and the carrier temperature. Moments are obtained by multiplying the distribution function with suitable weight functions $\phi = \phi(\mathbf{k})$ and integrating the product over $\mathbf{k}$-space as $n\langle\phi\rangle = \int \phi f \ \mathrm{d}^3\mathbf{k}$, with n being the carrier concentration. The equations which determine a given set of moments form a macroscopic transport model. Conventionally, the balance equations are obtained from powers of the energy $\mathcal{E}$, whereas the weight functions $\mathbf{p}\mathcal{E}^i$ with the momentum $\mathbf{p} = \hbar\mathbf{k}$ give the flux relations as

$$\begin{array}{lcllll}
\phi = 1 & \Rightarrow & \partial_t n + \nabla\cdot n\mathbf{V}_0 & & = 0 & (1.0)\\
\phi = \mathbf{p} & \Rightarrow & \partial_t n\mathbf{P}_0 + \nabla\cdot n\hat{\mathrm{U}}_1 - & n\mathbf{F} & = n\mathbf{Q}_0 & (2.0)\\
\phi = \mathcal{E} & \Rightarrow & \partial_t n w_1 + \nabla\cdot n\mathbf{V}_1 - 1\, n\mathbf{F}\cdot\mathbf{V}_0 & & = nq_1 & (1.1)\\
\phi = \mathbf{p}\mathcal{E} & \Rightarrow & \partial_t n\mathbf{P}_1 + \nabla\cdot n\hat{\mathrm{U}}_2 - & n\mathbf{F}\cdot(w_1\,\hat{\mathrm{I}} + \hat{\mathrm{U}}_1) & = n\mathbf{Q}_1 & (2.1)\\
\phi = \mathcal{E}^2 & \Rightarrow & \partial_t n w_2 + \nabla\cdot n\mathbf{V}_2 - 2\, n\mathbf{F}\cdot\mathbf{V}_1 & & = nq_2 & (1.2)\\
\phi = \mathbf{p}\mathcal{E}^2 & \Rightarrow & \partial_t n\mathbf{P}_2 + \nabla\cdot n\hat{\mathrm{U}}_3 - & n\mathbf{F}\cdot(w_2\,\hat{\mathrm{I}} + 2\,\hat{\mathrm{U}}_2) & = n\mathbf{Q}_2 & (2.2)\\
\vdots & & & & \vdots &
\end{array}$$

with $\mathbf{P}_i = \langle \mathbf{p}\mathcal{E}^i \rangle$, $\mathbf{V}_i = \langle \mathbf{u}\mathcal{E}^i \rangle$, $w_i = \langle \mathcal{E}^i \rangle$, $\hat{\mathrm{U}}_i = \langle \mathbf{u} \otimes \mathbf{p}\mathcal{E}^{i-1} \rangle$, the moments of the scattering integral $nq_i = \int \mathcal{E}^i \, Q[f(\mathbf{k})] \, \mathrm{d}^3\mathbf{k}$ and $n\mathbf{Q}_i = \int \mathbf{p}\mathcal{E}^i \, Q[f(\mathbf{k})] \, \mathrm{d}^3\mathbf{k}$, $\mathbf{F}$ the external force, in our case $\mathbf{F}(\mathbf{r}) = -\mathrm{q}\mathbf{E}(\mathbf{r})$ for homogeneous band structures and neglected magnetic fields. This equation system contains more unknowns than equations and each equation is coupled to the next higher equation. Since the current density is proportional to the average velocity $\mathbf{V}_0$ and many physical processes are modeled as a function of the average energy w_1, it is sensible to use the quantities $\mathbf{V}_i$ and w_i as solution variables. For such a choice the additional moments $\mathbf{P}_i$, $\hat{\mathrm{U}}_i$, q_i, and $\mathbf{Q}_i$ have to be expressed as functions of the solution variables, which is not exactly possible and known as closure problem.
Alternatively, the weight functions $\mathbf{u}\mathcal{E}^i$ can be used to obtain the flux relations [2]

$$\phi = \mathbf{u}\mathcal{E}^i \qquad \Rightarrow \qquad \partial_t n\mathbf{V}_0 + \nabla \cdot n\hat{\mathrm{U}}^*_{i+1} - n\mathbf{F} \cdot (\hat{\mathrm{W}}_i + i\hat{\mathrm{U}}^*_i) = n\mathbf{Q}^*_i \; , \tag{3.i}$$

with the tensors $\hat{\mathrm{U}}^*_i = \langle \mathbf{u} \otimes \mathbf{u}\,\mathcal{E}^{i-1} \rangle$ and $\hat{\mathrm{W}}_i = \langle \hat{\mathrm{m}}^{-1}\mathcal{E}^i \rangle$ and the moment of the scattering integral $n\mathbf{Q}^*_i = \int \mathbf{u}\mathcal{E}^i \, Q[f(\mathbf{k})] \, \mathrm{d}^3\mathbf{k}$.
Stratton's approach [3] is based on the microscopic relaxation time and gives another set of flux relations which, however, is not included in our comparison.
The balance equations (1) together with the flux relations (2) or (3) form a hierarchy of transport equations which has to be truncated at a certain order N to obtain a tractable equation set. Here we consider the orders $N = 4$ which results in hydrodynamic and energy-transport models, whereas $N = 6$ gives a six moments model.

2 Closure Relations

The closure relations used to model the additional moments in terms of the solution variables determine the differences between the various transport models. Some of the most commonly used approximations are discussed in the following.

2.1 Energy-Like Tensors

For parabolic energy bands we have a simple relation between the momentum $\mathbf{p}$ and the group velocity $\mathbf{u}$, $\mathbf{p} = m\mathbf{u}$, with m being the effective mass. If we further assume that the distribution function can be reasonably described by a displaced and heated Maxwellian distribution, simple relationships can be derived for the energy-like tensors, for instance $\hat{\mathrm{U}}_1 = \frac{2}{3}w_1\hat{\mathrm{I}} - m\mathbf{V}_0^2/3 + m\mathbf{V}_0 \otimes \mathbf{V}_0$. However, even the simplest model of this type, Bløtekjær's hydrodynamic model [4], is difficult to solve for multidimensional domains due to the existence of hyperbolic modes. Thus, claiming that the system is diffusion dominated [5], the convective terms and the time derivatives in the flux relations are commonly neglected, resulting in parabolic partial differential equation systems [6] which are considerably simpler to solve.
In the diffusion limit which we assume to hold in the following, the energy-like tensors for non-parabolic bands can be expressed by scalars, $\hat{\mathrm{U}}_i = U_i\,\hat{\mathrm{I}}$, which evaluate to $U_i = \frac{2}{3}w_i\,H_i$, where H_i considers the influence of a non-parabolic band structure on the streaming terms ($H_i = 1$ for parabolic bands). The non-parabolicity correction factors are defined through the trace of the energy-like tensors $\hat{\mathrm{U}}_i$ as $H_i = \mathrm{tr}\,\hat{\mathrm{U}}_i/(2w_1)$ and have been modeled as either energy-dependent using a simple analytical expression [7], by the incorporation of bulk Monte Carlo data [2], or via analytic models

for the distribution function [8]. Since the equation system is truncated after N equations, where we consider only an even number N, the highest-order solution variable is $w_{N/2-1}$. However, in the highest-order equation the moment $\hat{\mathrm{U}}_{N/2}$ appears which has to be expressed as a function of the available moments. Often, a heated Maxwellian distribution is used to derive such a relation [2–4].

2.2 Scattering Integral

The closure relations for the even moments of the scattering integral q_i are in general not considered to be too critical. Conventionally, relaxation times of the form $\tau_i = -(w_i - w_{i,\mathrm{eq}})/q_i$ are introduced, with $w_{i,\mathrm{eq}}$ being the equilibrium value of w_i. The relaxation times are then modeled as either constant or energy-dependent [6]. However, since a constant energy-relaxation time τ_1 in conjuction with an analytical mobility model often does not properly reproduce the homogeneous velocity-field characteristics, τ_1 has frequently been used as a fit parameter. Particularly, a smaller energy-relaxation time reduces the average energy and thus reduces the influence of hot-carrier effects. This makes energy-transport models appear to perform better for short-channel lengths where hot-carrier effects are often overestimated. However, the current in long-channel devices will be degraded as well. To avoid any uncertainties arising from this point we tabulate the bulk values for τ_i and model the relaxation times as functions of the average energy w_1 and doping.
The closure relations for the odd moments $\mathbf{Q}_i$, however, are more critical and subject to much debate [6, 8]. Rather simple expressions are obtained with the macroscopic relaxation time approximation where mobilities μ_i are introduced in analogy to the drift-diffusion model. $\mathbf{Q}_i$ is thus expressed as $\mathbf{Q}_i = -\mathrm{q}\mathbf{V}_i/\mu_i$, where μ_i is usually modeled as a function of the average energy w_1 only. A rigorous treatment reveals, however, that these odd moments of the scattering integral $\mathbf{Q}_i$ depend on the odd moments of the distribution function and thus on all fluxes [9, 10] as $\mathbf{Q}_i = \sum Z_{ij}\mathbf{V}_j$. Closure relations of this type cause an additional coupling between the flux equations and require a detailed description of the energy-like tensors to obtain an overall improvement of the transport model [11]. We therefore restrict our discussion to models based on the macroscopic relaxation time approximation.

2.3 Energy-Transport Models in the Heat-Flux Formulation

By expressing the term $n\mathbf{F}$ through $\mathbf{V}_0$ and inserting this definition into the energy-flux equation (2.1) we obtain by neglecting the terms ∇H_i, defining $w_1 = 3\mathrm{k_B}T_n/2$, calculating w_2 via a heated Maxwellian distribution, and assuming $H_2 \approx H_1(3 + 2H_1)/5$, the heat-flux formulation of (2.1)

$$n\mathbf{V}_1 = n\mathbf{V}_0 \tfrac{5}{2}\mathrm{k_B}T_n\delta_\mathrm{h} - \tfrac{5}{2}\delta_\mathrm{i} n \frac{\mathrm{k_B^2}T_n}{\mathrm{q}}\mu_0\nabla T_n \ , \quad \delta_\mathrm{h} = \frac{\mu_1}{\mu_0}\frac{3+2H_1}{5} \ , \quad \delta_\mathrm{i} = H_2\frac{\mu_1}{\mu_0} \ . \tag{4}$$

Various models have been suggested for the parameters δ_h and δ_i, simplifying the expressions given above (see Section 3). Equation (4) has often been subject to discussions. By comparison with Monte Carlo simulations it was found that for short-channel devices the heat-flux term (∇T_n) has to be reduced by multiplying δ_i with a constant factor f_hf in the range $0.1 - 0.2$ to obtain acceptable agreement [12, 13]. Reviewing

the derivation of (4) reveals one possible reason, namely that the ∇T_n term only enters the energy-flux relation through the heated Maxwellian assumption for the term $\nabla(U_{i+1}/U_1)$, which is known to be poor for short-channel devices. Another important contribution has been identified by Tang *et al.* [14] where an inhomogeneous term was added to the homogeneous mobility, which caused a significant reduction of the heat-flux term in the final relations.

3 Macroscopic Transport Models

In the following we shortly describe the non-parabolic transport models used for our comparison. Except otherwise noted, all transport models use the same mobilities μ_i and relaxation times τ_i extracted from homogeneous full-band simulations [15] as a function of w_1 and doping. To ensure consistency in the bulk case, the mobilities are extracted from the homogeneous flux relations as

$$\mu_i = \frac{V_i}{E\, w_i(1 + \frac{2}{3} i H_i)} \tag{5}$$

and the relaxation times from the homogenous balance relations as

$$\tau_i = -\frac{w_i - w_{i,\mathrm{eq}}}{i\, E\, V_{i-1}} \,. \tag{6}$$

Note that the definition of the non-parabolicity factors H_i only enters the definition of the higher-order mobilities whereas the carrier mobility μ_0 is not affected ($w_0 = 1$). We also consider so-called parabolic models where the influence of H_i is neglected, because it is still not clear how to best extract the tensors $\hat{\mathrm{U}}_i$ from a full-band simulation where the wave-vector $\mathbf{k}$ is not zero in a band-minimum as opposed to analytic-band models. However, the model still contains the full-band information through the mobilities and relaxation times.

3.1 Consistent Bulk Parameter Models

We define consistent models as those which exactly reproduce the even moments w_i and the fluxes $\mathbf{V}_i$ ($i \leq N/2 - 1$) from the Monte Carlo simulator under homogenous conditions. Furthermore, we require that all parameters of the model can be determined under homogeneous conditions. This results in models with 'no knobs to turn' [13]. Having too many adjustable parameters is a particular inconvenience inherent in many energy-transport models based on analytical models for the mobilities and relaxation times [6, 13].

- A non-parabolic six moments (NPSM) model is obtained by empirically setting $U_3 = (35/9) w_1^3 \beta^c H_3$. The quantity $\beta = (3/5) w_2/w_1^2$ is the kurtosis of the distribution function and indicates the deviation from a heated Maxwellian distribution for which $\beta = 1$ holds (for parabolic bands). The parameter $c = 2.7$ was obtained by a best fit of w_3 under homogeneous conditions. A consistent parabolic representation (PSM) is obtained by setting $H_i = 1$ and using different higher-order mobilities.

- A non-parabolic energy-transport (NPET) model is obtained by assuming a heated Maxwellian distribution to calculate a relationship between w_2 and w_1 which gives $U_2 = (10/9)w_1^2 H_2$. The NPET model can be written in heat-flux form (HFET) (cf. (4)) and the parabolic representation (PET) is obtained by setting $H_i = 1$.

- The generalized hydrodynamic model (GHDM) [2] is based on the flux equations (3.i) and various approximations for the tensors $\hat{\mathrm{U}}_i^*$ and $\hat{\mathrm{W}}_i$. In addition, the quantity $T^* = m^* \langle \mathbf{u}^2 \rangle / (3\mathrm{k_B})$ is used instead of w_1. This modifies the balance equations where a modified energy relaxation time τ_1^* is required to account for this variable transformation. Since all quantities are expressed as functions of T^* the GHDM is difficult to compare with the models based on w_1 and a practical evaluation is mandatory.

3.2 Simplified Energy-Transport Models

The energy-transport models given above require the modeling of two mobilities, one relaxation time, and one or two non-parabolicity factors. Since these models are not always available several simplified models have been proposed which are considered in the following. These models aim at eliminating the energy-flux mobility from the equation system by assuming $\mu_1 = r\mu_0$ with r close to unity. This assumption introduces an error in the energy-flux even under homogeneous conditions but does not influence the homogeneous current. More severe is the assumption of a constant energy-relaxation time which also influences the current.

- In [13] Tang *et al.* proposed to approximate δ_h as 0.8 and to empirically set δ_i in the range $0.1 - 0.2$. Thus the energy-flux mobility is removed from the equation system. Here, we model U_1 by a multiplicative factor H_1 in contrast to the additive factor proposed in [16]. We do not expect this to give significantly different results, however.

- A standard parabolic energy-transport (SPET) model as implemented in device simulators such as DESSIS [17] or MINIMOS-NT [18] can be obtained by setting $H_1 = 1$, $\delta_\mathrm{h} = 1$, and $\delta_\mathrm{i} = 1$. Again, only the carrier mobility enters this equation. Furthermore, a constant energy-relaxation time $\tau_1 = 0.35\,\mathrm{ps}$ will be assumed.

4 Comparison of Macroscopic Transport Models

To compare the performance of these transport models we simulated a series of double-gate MOSFETs similar to the ones used in [19] and compared the results to self-consistent full-band Monte Carlo results [15]. The gate length was varied from 250 down to $25\,\mathrm{nm}$ while the silicon layer thickness was set to $t_\mathrm{Si} = L_\mathrm{g}/4$. To reduce the Monte Carlo calculation time for the longer devices t_Si was limited to $t_\mathrm{Si} \leq 12.5\,\mathrm{nm}$. To avoid too high electron temperatures in the contact regions, the doping concentration was increased to $2 \times 10^{20}\,\mathrm{cm}^{-3}$ in the source/drain regions.
For the accurate description of carrier transport in sub-micron devices the quantization and its influence on the transport parameters inside the channel has to be considered. However, it is important to clarify whether the models can capture the non-local transport inside short-channel devices first before adding additional complexity to the

description. In the following quantum-mechanical effects are neglected which should give a pessimistic estimate of the validity of the respective transport models since the quantization of the energy-levels reduces the mean free path.
Another critical issue is the treatment of surface scattering. Since there is still no consistent way of transferring microscopic surface scattering models used in the Monte Carlo code to macroscopic models, the surface scattering parameters have to be calibrated. Because each model gives a slightly different carrier distribution inside the channel this calibration is not unique. Considerable research on this topic is still required.
Even without surface scattering, the full-band channel mobilities were found to be different from the bulk case because of the mere presence of the interface not accounted for in the tabulated models. We introduced this effect by scaling all channel mobilities with a constant factor (≈ 0.92) which was determined independently for each model from the simulation of a $250\,\mathrm{nm}$ MOSFET biased at $V_\mathrm{G} = 1\,\mathrm{V}$ and $V_\mathrm{D} = 100\,\mathrm{mV}$. This effect is regularly calibrated together with the surface mobility model.
The simulated output characteristics are shown in Fig. 1. All models give good results for the $250\,\mathrm{nm}$ device whereas for the short-channel device the results differ significantly. From the bulk parameter models, the six-moments models give the best results, the energy-transport models overestimate the currents by about 14%, while the drift-diffusion model underestimates it by about the same amount. When the heat-flux is reduced by $85 - 90\%$, the error in the currents stays below 10% for all variants of the energy-transport model. This improvement is linked to the more accurate velocity profile shown in Fig. 2: The six-moments models are in good agreement with the Monte Carlo results while the energy-transport models deliver the familiar overestimation which can only be controlled by reducing the heat-flux. The errors in the drain current and the quasi-static transit frequency [15] relative to the Monte Carlo results are shown in Fig. 3 and Fig. 4.

5 Conclusions

From this small set of simulations we might give some rough guidelines: first, the accuracy of the drift-diffusion model decreases rapidly for gate-lengths shorter than $100\,\mathrm{nm}$, particularly when transit frequencies are calculated. Secondly, energy-transport models which only use an estimation for the energy-flux mobility do not lose much accuracy. It is important, though, that the heated Maxwellian assumption is alleviated by empirically reducing the heat-flux by about 90%. With this measure energy-transport models can be used for gate-lengths down to $50\,\mathrm{nm}$ at the price of having one empirical fit-parameter and thus some uncertainty. Finally, the parabolic six-moments model gives good results for gate-lengths as small as $25\,\mathrm{nm}$. In addition the six moments models give the kurtosis which can be used to describe the distribution function beyond the heated Maxwellian approximation and thus to model hot-carrier effects more accurately.

Acknowledgment

The authors are grateful to Prof. T.-w. Tang for critically reviewing this manuscript. This work has been partly supported by *Infineon Technologies*, Villach, and *austriamicrosystems*, Graz, Austria.

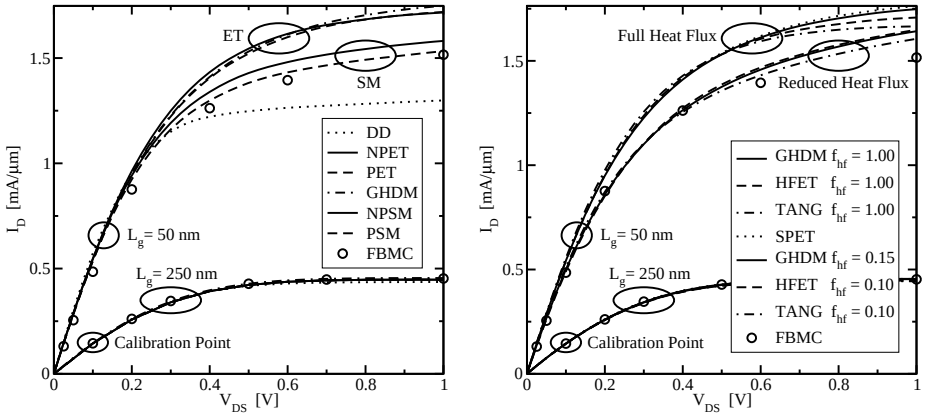

Figure 1: Comparison of the simulated output characteristics for the $50\,\mathrm{nm}$ device at $V_\mathrm{G} = 1\,\mathrm{V}$. The bulk parameter models are shown on the left while the other models are shown on the right.

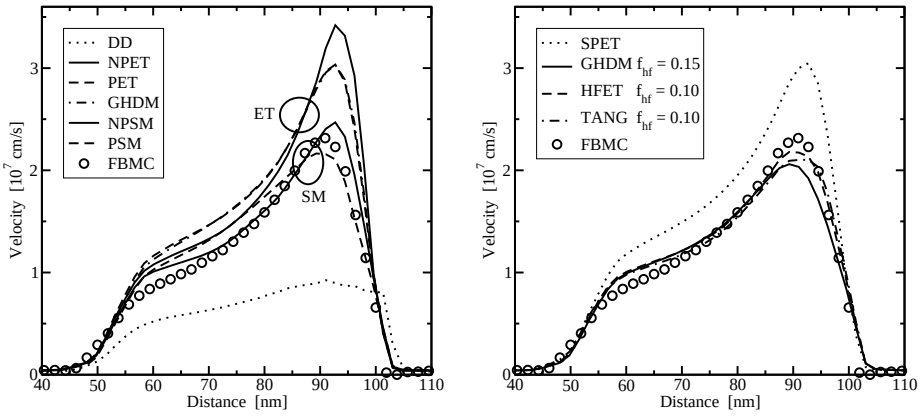

Figure 2: Comparison of the simulated velocity profiles for the $50\,\mathrm{nm}$ device at $V_\mathrm{G} = V_\mathrm{D} = 1\,\mathrm{V}$. The bulk parameter models are shown on the left while the other models are shown on the right.

[1] M. Lundstrom, *Fundamentals of Carrier Transport* (Cambridge University Press, 2000).

[2] R. Thoma *et al.*, IEEE Trans.Electron Devices **38**, 1343 (1991).

[3] R. Stratton, Physical Review **126**, 2002 (1962).

[4] K. Bløtekjær, IEEE Trans.Electron Devices **17**, 38 (1970).

[5] C. Ringhofer, C. Schmeiser, and A. Zwirchmayer, SIAM J.Numer.Anal. **39**, 1078 (2001).

[6] T. Grasser, T. Tang, H. Kosina, and S. Selberherr, Proc.IEEE **91**, 251 (2003).

[7] T. Bordelon, X.-L. Wang, C. Maziar, and A. Tasch, Solid-State Electron. **35**, 131 (1992).

[8] T. Grasser, H. Kosina, and S. Selberherr, in *Proc. Simulation of Semiconductor Processes and Devices* (Boston, USA, 2003), pp. 63–66.

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

[10] A. M. Anile and O. Muscato, Physical Review B **51**, 16728 (1995).

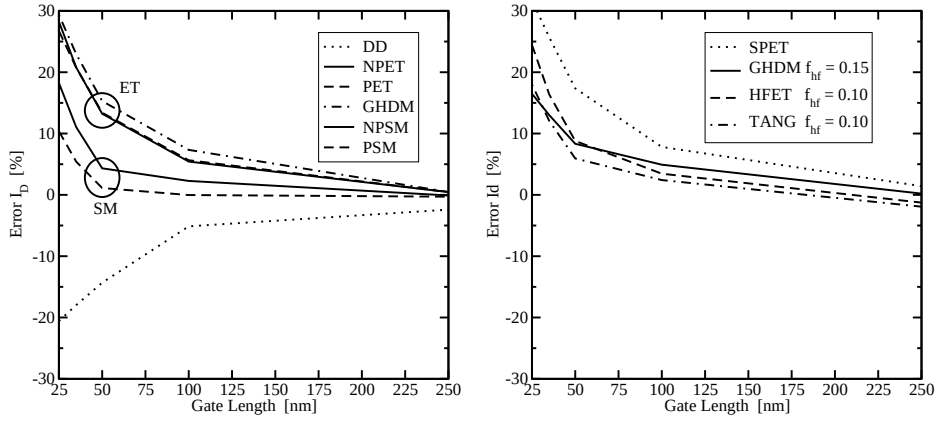

Figure 3: Error in the drain currents relative to the Monte Carlo results. The bulk parameter models are shown on the left while the other models are shown on the right.

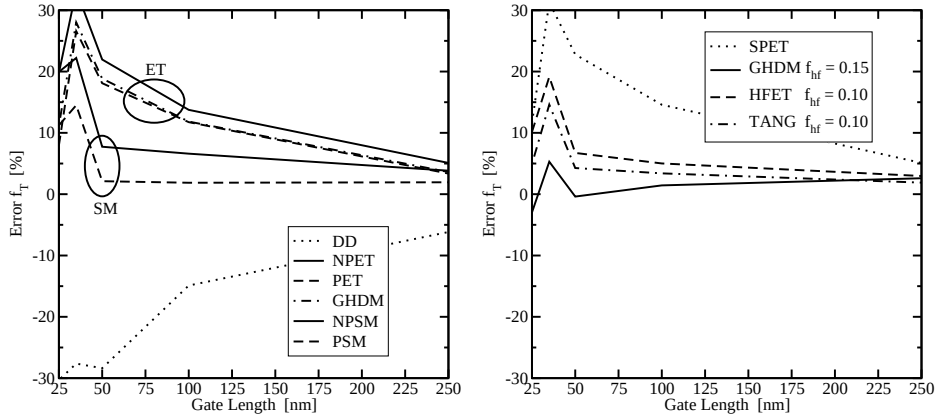

Figure 4: Error in the transit frequency relative to the Monte Carlo results. The bulk parameter models are shown on the left while the other models are shown on the right.

[11] T. Grasser, H. Kosina, and S. Selberherr, in *Proc. Simulation of Semiconductor Processes and Devices* (Munich, Germany, 2004).

[12] I. Bork, C. Jungemann, B. Meinerzhagen, and W. Engl, in *Intl. Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits NUPAD V* (Honolulu, 1994), pp. 63–66.

[13] T. Tang, in *Semiconductor TCAD Workshop & Exhibition* (Hsinchu, Taiwan, 1999), pp. 1–19.

[14] T. Tang, X. Wang, H. Gan, and M. Leong, VLSI Design **13**, 131 (2000).

[15] C. Jungemann and B. Meinerzhagen, *Hierarchical Device Simulation: The Monte-Carlo Perspective* (Springer, Wien–New York, 2003).

[16] T. Tang, S. Ramaswamy, and J. Nam, IEEE Trans.Electron Devices **40**, 1469 (1993).

[17] *DESSIS-ISE, ISE TCAD Release 9.0*, ISE Integrated Systems Engineering AG, Zürich, Switzerland, 2003.

[18] IμE, *MINIMOS-NT 2.0 User's Guide*, Institut für Mikroelektronik, Technische Universität Wien, Austria, 2002, http://www.iue.tuwien.ac.at/software/minimos-nt.

[19] F. Bufler, A. Schenk, and W. Fichtner, J. of Comp. Electronics **2**, 81 (2003).