Technology CAD: Device simulation and characterization

T. Grasser and S. Selberherr
Institute for Microelectronics, Technische Universität Wien, Gußhausstraße 27–29, A-1040 Vienna, Austria

(Received 28 August 2001; accepted 26 November 2001)

State-of-the-art devices are characterized by the occurence of large gradients in the electric field. Due to the complex doping profiles which utilize ultrashallow junctions to obtain an optimum tradeoff between short-channel effects and parasitic resistances, two-dimensional numerical simulation of these structures is mandatory. From the modeling point of view, nonlocal and quantum effects gain more and more importance which need to be properly accounted for. We review the requirements for successful numerical simulation of these semiconductor devices. In addition, we give an overview of recent activities concerning device calibration and inverse modeling since inverse modeling of the doping profile in conjunction with calibration of the model parameters has proven to be an effective method of two-dimensional doping profile extraction. © *2002 American Vacuum Society.* [DOI: 10.1116/1.1445162]

I. INTRODUCTION

Continuous advances over the past years in integrated circuit technology impose new challenges on modeling of fabrication processes and electrical behavior of semiconductor devices. The routine utilization of process and device simulation has become indispensable for the development and redesign of ultralarge scale integration (ULSI) devices as well as for power devices.

Device simulation based on the self-consistent solution of the basic semiconductor equations dates back to the famous work of Gummel in 1964.[1] Since then numerical device modeling has been applied to nearly all important devices. The models used for the description of carrier transport through a device have frequently been the subject of discussions, in particular in view of their applicability to submicron devices. It has been argued that the classical drift-diffusion (DD) approach[2] looses its validity for devices smaller than 1 μm because nonlocal effects gain more importance. Nevertheless, most of the routinely performed simulations still employ the DD model and successful simulation of devices as small as 40 nm has been reported.[3] The reasons why the DD model still gives fairly accurate results, even though some fundamental underlying assumptions are definitely violated, has been a permanent issue of discussion and a recent study can be found in Ref. 4. One reason for the lasting popularity of the DD model is its simplicity and the stability of the available implementations. In addition, device and circuit designers are mainly interested in integral quantities, for instance the contact currents, which can be reproduced with reasonable accuracy when some parameters of the mobility model are adjusted accordingly. We review some hydrodynamic and energy-transport models which have been proposed to overcome the limitations of the DD model. Furthermore, we point out the most important effects which have to be considered when deep-submicron devices are simulated.

The calibration of model parameters is one of the routine tasks to be performed by device engineers. For the calibration of device models a proper description of the device in terms of doping profiles and geometries is required. Calibration of process simulators can be achieved by using measured doping profiles or by employing a followup device simulation. The latter has become known as inverse modeling of doping profiles and several successful algorithms have been published so far. Care must be taken that empirical models used in the device simulator do not affect the extracted doping profiles. We review some of these algorithms and point out their basic difficulties.

II. DEVICE SIMULATION

A device of a modern ULSI circuit is characterized by large electric fields in conjunction with steep gradients of the electric field and of the carrier concentrations. The DD model cannot cover nonlocal effects as the electron gas is assumed to be in thermal equilibrium with the lattice temperature. For rapidly increasing electric fields the energy lags behind the electric field because it takes the carriers some time to pick up energy from the field. 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. Thus, DD simulations predict the same velocity profile as for slowly varying fields which can dramatically underestimate the carrier velocities. Similar to the mobility, many other physical processes are more accurately described by a local energy model rather than a local electric-field model. Therefore, the assumption of a fixed energy-field relation can cause nonphysical results when used to predict, for example, impact ionization. Sophisticated device models, such as the hydrodynamic and energy-transport models,[5,6] the spherical harmonics expansion method[7,8] and the Monte-Carlo technique,[9–11] aim at overcoming the limitations of the established DD model. However, the increased physical rigor of a model comes at the expense of increased demand on computation time. This fact prevented widespread application of the models in the past, and probably in the near future. In addition, the modeling task becomes more complex due to the increased number of parameters required by these models.

A. Hydrodynamic and energy-transport models

Monte-Carlo (MC) simulations have been proven to give accurate results but are often prohibitively time consuming. Therefore, a common simplification is to investigate only some moments of the distribution function, such as the carrier concentration and the carrier temperature. Extensions to the DD model have been proposed which basically add an additional balance equation for the average carrier energy.[5,6] Furthermore, a term is added to the current relation which is proportional to the gradient of the carrier temperature. Several different formulations have been proposed which vary considerably in complexity. Extensions to handle nonhomogeneous materials and nonparabolicity effects have also been given.[12–14]

Transport equations used in semiconductor device simulation are normally derived from Boltzmann's transport equation (BTE) which reads[15]

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

for a general inhomogeneous material with arbitrary band structure.[16] For inclusion of quantum effects equations based on the Wigner-Boltzmann equation have been considered.[17] The group velocity $\mathbf{u}$:

$$\mathbf{u}(\mathbf{k},\mathbf{r})=\frac{1}{\hbar}\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k},\mathbf{r}) \tag{2}$$

defines the inverse effective mass tensor

$$\hat{m}^{-1}(\mathbf{k},\mathbf{r})=\frac{1}{\hbar}\nabla_{\mathbf{k}}\otimes\mathbf{u}(\mathbf{k},\mathbf{r})=\frac{1}{\hbar^2}\nabla_{\mathbf{k}}\otimes\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k},\mathbf{r}), \tag{3}$$

where $\otimes$ denotes the tensor product.[16] In the following we will only consider position-independent masses but permit energy-dependent masses. Generalizations to position-dependent band structures will be given in the appropriate context. The force $\mathbf{F}$ exerted on the particles is generally given as

$$\mathbf{F}(\mathbf{k},\mathbf{r})=-\nabla_{\mathbf{r}}E_{c,0}(\mathbf{r})-q(\mathbf{E}(\mathbf{r})+\mathbf{u}\times\mathbf{B})-\nabla_{\mathbf{r}}\mathcal{E}(\mathbf{k},\mathbf{r}) \tag{4}$$

and depends both on $\mathbf{k}$ and $\mathbf{r}$. Omitting the influence of $\mathbf{u}\times\mathbf{B}$ (see Ref. 18 for a treatment of this term) and assuming homogeneous materials, $\mathbf{F}$ simplifies to

$$\mathbf{F}(\mathbf{r})=-q\mathbf{E}(\mathbf{r}). \tag{5}$$

The BTE is an equation in the seven-dimensional phase space which is too demanding to solve for engineering applications. Therefore, a common simplification is to investigate only some moments of the distribution function which are defined as $\langle\Phi\rangle=\int\Phi f d^3\mathbf{k}/(4\pi^3)$ with a suitable weight function $\Phi=\Phi(\mathbf{k})$. The simplest approximation for the complex band structure is a parabolic relationship between the energy and the crystal momentum

$$\mathcal{E}=\frac{\hbar^2k^2}{2m^*} \tag{6}$$

which is valid for energies close to the band minimum. A first-order nonparabolic relationship was given by Kane[19] as

$$\mathcal{E}(1+\alpha\mathcal{E})=\frac{\hbar^2k^2}{2m^*} \tag{7}$$

with α being the nonparabolicity correction factor. This correction is frequently used for the derivation of nonparabolic energy-transport models.[13,14] Bløtekjær[5] derived conservation equations by taking the moments of the BTE using the weight functions 1, $\hbar\mathbf{k}$, 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. In his original work Bløtekjær used Fourier's law to close the equation system, which has been shown to be problematic, see for instance Ref. 20. To overcome this shortcoming, the fourth moment of the BTE has been taken into account (see for example Ref. 21) which gives the following moment equations:[22]

$$\frac{\partial n}{\partial t}+\nabla\cdot(n\mathbf{v})=nC_n, \tag{8}$$

$$\frac{\partial(n\mathbf{p})}{\partial t}+\nabla\cdot(n\hat{\mathrm{U}})-n\mathbf{F}=n\mathbf{C}_p, \tag{9}$$

$$\frac{\partial(nw)}{\partial t}+\nabla\cdot(n\mathbf{S})-n\mathbf{v}\cdot\mathbf{F}=nC_{\mathcal{E}}, \tag{10}$$

$$\frac{\partial\langle\hbar\mathbf{k}\mathcal{E}\rangle}{\partial t}+\nabla\cdot(n\hat{\mathrm{R}})-n(w\hat{\mathrm{I}}+\hat{\mathrm{U}})\cdot\mathbf{F}=n\mathbf{C}_{p\mathcal{E}}, \tag{11}$$

with

$$n=\langle 1\rangle,\quad \mathbf{p}=\frac{1}{n}\langle\hbar\mathbf{k}\rangle,\quad \mathbf{v}=\frac{1}{n}\langle\mathbf{u}\rangle=-\frac{\mathbf{J}}{qn},\quad w=\frac{1}{n}\langle\mathcal{E}\rangle, \tag{12}$$

$$\mathbf{S}=\frac{1}{n}\langle\mathbf{u}\mathcal{E}\rangle,\quad \hat{\mathrm{T}}=\frac{1}{k_Bn}\langle m(\mathbf{k})\mathbf{c}\otimes\mathbf{c}\rangle,\quad \hat{\mathrm{U}}=\frac{1}{n}\langle\hbar\mathbf{u}\otimes\mathbf{k}\rangle,$$

$$\hat{\mathrm{R}}=\frac{1}{n}\langle\hbar\mathbf{u}\otimes\mathbf{k}\mathcal{E}\rangle. \tag{13}$$

Note that these expressions are valid for arbitrary band structures, provided that the carrier mass is homogeneous. When $\mathbf{F}$ is allowed to be position dependent, additional force terms appear in Eqs. (8)–(10).[23] The collision terms are usually modeled with a macroscopic relaxation time approximation as

$$C_n=-\frac{1}{n}(R-G)=-\frac{1}{n}U,\qquad \mathbf{C}_p=-\frac{q\mathbf{v}}{\mu},$$

$$C_{\mathcal{E}}=-\frac{w-w_0}{\tau_{\mathcal{E}}},\qquad \mathbf{C}_{p\mathcal{E}}=-\frac{q\mathbf{S}}{\mu_S}, \tag{14}$$

which introduces the energy relaxation time $\tau_{\mathcal{E}}$ and the mobilities μ and μ_S. A discussion on this approximation is given in Ref. 24. This equation set is not closed as it contains more unknowns than equations. Closure relations have to be

found to express the equations in terms of the unknowns n, $\mathbf{v}$, and w. Due to the strong scattering the temperature tensor is normally assumed to be isotropic and is approximated by a scalar $\hat{\mathrm{T}} \approx T_n \hat{\mathrm{I}}$. Traditionally, parabolic bands were assumed, which gives the following closure relations for $\mathbf{p}$, $\hat{\mathrm{U}}$, and w:

$$\mathbf{p} = m^* \mathbf{v}, \qquad \hat{\mathrm{U}} = \frac{m^*}{n} \langle \mathbf{u} \otimes \mathbf{u} \rangle = k_B T_n \hat{\mathrm{I}} + m^* \mathbf{v} \otimes \mathbf{v},$$

$$w = \frac{3}{2} k_B T_n + \frac{m^* v^2}{2}. \tag{15}$$

With these equations a so-called hydrodynamic equation system is obtained which contains a convective term $\langle \mathbf{u} \otimes \mathbf{u} \rangle$ in the energy tensor $\hat{\mathrm{U}}$. The hydrodynamic (HD) 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, $v_c = \sqrt{\xi}\, 1.3 \times 10^7$cm/s while for $T_L = 77$ K, $v_c = \sqrt{\xi}\, 6.6 \times 10^6$ cm/s.[25] In the case of supersonic flow, electron shock waves will in general develop inside the device.[25] These shock waves occur at either short length scales or at low temperatures and require a special treatment of the equation system. To avoid the difficulties arising from this treatment, the convective term is normally neglected and an energy-transport model is obtained which is much simpler to deal with as it only covers the subsonic flows. Furthermore, the contribution of the velocity to the carrier energy is frequently neglected ($w \approx \frac{3}{2} k_B T_n$).

1. Critical issues

The method of moments transforms the BTE 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. For their generalized HD model, Thoma *et al.*[12] give a maximum error of 30%, which can be quite significant. One approach to derive a suitable closure relation is to assume a distribution function and calculate the fourth-order moment. Ramaswamy and Tang[26] gave a comparison of different closure relations available in the literature.

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 simulations have been carried out in Ref. 20 which indicate that the longitudinal temperature component T_l is larger than the transverse temperature component T_t, indicating 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.*[27] who model four components of the temperature tensor. They observed no significant difference between the scalar temperature and $\mathrm{Tr}(\hat{T}_n)/3$ for ballistic diodes and bipolar transistors but a 15% difference for aggressively scaled metal-oxide semiconductor field-effect transistor (MOSFETs) in the linear region of the transfer characteristics. Tang *et al.*[22] observed that the energy tensor is not a single valued function of the average energy and give models using available moments.

Another common approximation is to neglect the drift energy in the average carrier energy[28] $w = m\mathbf{v}^2/2 + \frac{3}{2} k_B T_n \approx \frac{3}{2} k_B T_n$. As has been pointed out by Baccarani and Wordeman,[29] the convective energy can reach values comparable to thermal energy. This effect has been studied in Ref. 30.

The relaxation times have traditionally been derived from homogeneous field measurements or MC simulations. For homogeneous fields, there is a unique relationship between the electric field and the carrier temperature which can be used as a definition for $\tau_\mathcal{E}$. However, due to the modeling of the collision terms, 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. Anile and Romano[31] and Muscato[32] derived expressions for the closure $\hat{U}$ and $\hat{R}$ using the maximum entropy principle. In addition, they were able to derive expressions for the collision terms. They found that their model fulfills Onsagers' reciprocity principle and gave a comparison with other hydrodynamic models.

B. Extended models

The general hydrodynamic equations (8)–(10) are valid for any band structure as $\mathbf{F}$ depends only on the spatial gradient of the dispersion relation. However, parabolicity assumptions are invoked to derive the closure relations (15). On the other hand, nonparabolicity effects enter the HD 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.[10]

Bløtekjær's[5] equations were originally devised for semiconductors with multiple bands. Woolard *et al.*[33,34] extended these expressions for multiple nonparabolic bands in GaAs. Other GaAs models can be found in Refs. 35 and 36. Wilson[37] gave an alternate form of the HD model which he claims to be more accurate than Ref. 5. Another multivalley nonparabolic energy-transport model was proposed in Ref. 38.

The highly nonhomogeneous field distributions found in modern devices give rise to distribution functions which de-

viate significantly from the frequently assumed Maxwellian distribution. Furthermore, as has been pointed out in Ref. 39, the distribution function is not uniquely described using just the average carrier energy.

Several moment based models have been proposed so far which aim at obtaining some additional information about the distribution function 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.[40,41] 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 Ref. 42 for those regions where the high-energy tail is heavily populated. In Ref. 43 Tang gave a simplified version of the two energy model.[40]

C. Contact models

For the ever shrinking devices the influence of the contacts should no longer be neglected. Traditionally, for Ohmic contacts, the potential is modeled via the builtin potential and the metal quasi-Fermi level, the carrier concentration is determined via the builtin potential and a suitable statistic, and the carrier temperature is set equal to the lattice temperature. The last assumption, especially, requires the carriers to loose their energy before reaching the contact while a more realistic model would allow for a heat flow over the contact and thus a contact temperature different from the lattice temperature. Some advanced models which capture the transition from Ohmic to Schottky contacts can be found in Refs. 44–46. A tunneling contact model, suitable for small raised source drain MOSFET's and Schottky source drain MOSFET's is given in Ref. 47.

D. Quantum effects

When device dimensions are comparable to the de Broglie wavelength, quantum effects have to be considered in the simulation. With shrinking feature size of complementary metal-oxide semiconductor (CMOS) devices especially, quantum effects become important for mainstream technology. Quantum effects at interfaces cause the continuous energy bands to split into discrete subbands where the carriers are confined in a two-dimensional electron gas. This has a direct impact on both the amount of charge which can be induced by the gate electrode through the gate oxide and the profile of the channel charge in the direction perpendicular to the surface which is shifted away from the surface by 2–3 nm. In addition, the mobility at the surface is influenced. Proper solutions of this problem can be obtained by solving the Schrödinger and the Poisson equation self-consistently. Multidimensional solutions of this problem are very difficult to obtain but fortunately not necessary in most cases.[48]

The first analytical correction for capturing quantum effects was proposed by Hänsch *et al.*[49] who modeled the repulsive boundary condition for channel carriers at the Si/SiO_2 interface. They propose to use a depth dependent density-of-states

$$N_c(z)=N_c\left[1-\exp\left(-\frac{(z+z_0)^2}{\lambda^2}\right)\right] \tag{16}$$

with z_0 being an offset to model nonzero carrier concentrations at the interface due to finite barrier height and λ a characteristic length which indicates how fast quantum effects diminish away from the interface. This approach introduces an additional driving force in the current relation due to the position dependence of N_c. Another model has been proposed by van Dort *et al.*,[50] who tries to capture the effect of energy quantization by introducing a quantum related band-gap increase as

$$\Delta E_g(z)=\beta F_S^{2/3} g(z) \tag{17}$$

with F_S being the pressing force, β a constant, and $g(z)$ the depth dependent function. Care must be taken to apply these models only to those regions where the carriers are confined to the interface.[50] As pointed out in Ref. 51, van Dort's model shows a discontinuity around the flat band voltage when simulating $C-V$ curves. A hybrid model, which combines the latter two models, has been proposed in Ref. 51.

A different approach suitable for macroscopic device simulation has been proposed by Ancona[52] and is termed the density gradient method. This method has been derived from a quantum corrected BTE and is an extension to the classic drift-diffusion transport model. In the driving force an additional density gradient term appears which is proportional to

$$\nabla\left(\frac{\nabla^2\sqrt{n}}{\sqrt{n}}\right) \tag{18}$$

and turns the continuity equation into a fourth-order partial differential equation (PDE). A recent implementation delivered promising results.[53] For a detailed discussion and a comparison of quantum correction models see Refs. 48 and 54.

1. Tunneling currents

The calculation of tunneling currents is closely related to the quantum corrections mentioned before and should be treated together. Due to the finite barrier height, the probability density is not equal to zero at the interface and a current can flow between gate and substrate. Accurate predictions of the currents can be obtained by solving Schrödinger's equation which is, however, very time consuming. Thus simplified models have been proposed to capture the different effects which eventually lead to a gate current. These effects are direct tunneling, Fowler-Nordheim tunneling, and thermionic emission. Detailed studies of these tunneling processes are available in literature and many models have been

reported to accurately reproduce measured results.[53,55–57] The issue of hot electron injection, however, is still critical. Cold carriers in the channel of a silicon *n*MOS transistor face a barrier of 3.2 eV. As they travel along the channel they pick up energy, which decreases the height of the barrier to be surmounted and thus increases the tunneling probability. Unfortunately the distribution function of hot carriers is only poorly described by the traditionally assumed heated Maxwellian shape which introduces a large uncertainty into the result. An attempt to use a more realistic distribution function can be found in Ref. 58. However, it is to be expected that an extension to the energy-transport model is required to properly capture this effect.

E. Polysilicon depletion and lightly doped drain source/drain series resistances

The modeling of both the depletion of polysilicon gates and the series resistances found in the source or drain extension of LDD structures, fits well into the concept of semiclassical device simulation as both effects are captured by the semiconductor device equations and no additional modeling is necessary. For the extraction of series resistances see, for instance, Refs. 59 and 60. When modeling poly depletion, the straightforward solution of the semiconductor equations in the gate increases the computational burden. Under steady-state conditions, where there is negligible electron and hole current flow, simplified models have been proposed which capture the essential effects with sufficient accuracy.[61] For steady-state the carrier concentrations can be calculated as a function of the potential using Fermi–Dirac statistics. Using the gradual channel approximation for the potential near the interface the problem can be reduced to one dimension where the potential ψ_c at the gate is replaced by $\psi_c - \Delta(x)$ with $\Delta(x)$ modeling the position dependence of the gate potential.[61] See Refs. 62 and 63 for other detailed discussions.

F. Impact ionization

To properly account for substrate currents accurate impact ionization models are of utmost importance. In a classical paper Chynoweth[64] proposed a local electric-field model for the ionization coefficients. This model accurately predicts the ionization rates in homogeneous bulk material. For inhomogeneous fields, local temperature dependent models proved to be more accurate.[65] However, recent research revealed that for submicron devices local temperature dependent models break down and tend to overestimate the maximum of the ionization rate. Furthermore, a lag in space was observed, that is, the maximum of the ionization rate occurred after the temperature had already started to decrease. Although the parameters of local temperature models had been successfully adjusted to reproduce measured substrate currents, the distribution of the physical quantities inside the device are wrong. Even worse, bulk results cannot be recovered with these adjusted parameters and the parameters are only valid for a distinct technology. Several corrections have been proposed to cover these effects, for instance a ''dead length'' correction.[66] We believe that the necessary information is contained in the distribution function and a more accurate description of the distribution function can improve the quality of the model. The calculation of impact ionization using a six moment transport model can be found in Refs. 67 and 68. For results obtained by two carrier type models see Refs. 40–42.

III. CHARACTERIZATION: CALIBRATION AND INVERSE MODELING

Although calibration of model parameters is widely used by engineers around the world, a lot of experience and understanding of the involved physical models is required to obtain meaningful results. This is partly due to the fact that many models used in process and device simulation are empirical and due to their complex interplay when applied to a realistic simulation. Even physics based models normally employ some simplifying assumptions to yield analytical expressions. Furthermore, most of the models were derived under homogeneity assumptions or slowly varying electric fields, which is not applicable to state-of-the-art devices. So the parameters of these models need to be adjusted using an ''effective'' value to fit the process under consideration.

For successful device simulation a high-quality description of the device in terms of doping profiles and geometry, especially the gate thickness and gate length, is required. These device descriptions can be obtained by process simulators which need to be carefully calibrated to the process under consideration. This calibration is difficult due to the fact that two-dimensional doping profiles are complicated to extract. Another approach is to develop an analytical description of the device which is then optimized using measured electrical characteristics. As suitable fitting targets $C-V$ measurements,[69–71] subthreshold $I-V$ characteristics,[72] and the threshold voltage[73] have been identified. As pointed out in Ref. 72, $C-V$ measurements are problematic due to the extremely small dimensions and capacitance of modern submicron devices and special test structures are needed, which limits the applicability of this method. Furthermore, as quantum effects and polydepletion can no longer be neglected for oxides thinner than 4 nm, uncertainties in the modeling of these effects introduce an error in the extracted gate length. A recent comparison of the most sophisticated technology computer aided design tools (Schrödinger solvers) revealed a 20% difference in the extracted gate lengths.[74]

For inverse modeling, the crucial features of the doping profile need to be captured in the analytical description. Several analytical expressions have been used so far, including two-dimensional Gaussian functions for representing source or drain extensions,[72] one-dimensional splines for representing the depth dependence of the channel doping,[69] and Pearson Type IV functions.[2,75]

Inverse modeling and calibration tasks are described by a set of parameters which need to be optimized to meet a required target. For this optimization, several different algorithms have been investigated. Traditionally, gradient-based algorithms such as the Levenberg–Marquardt algorithm or

the simpler response surface method[76] have been used. Gradient-based methods work fast and efficiently when a proper initial guess is supplied but fail otherwise. Furthermore, they are only guaranteed to find a local minimum which does not necessarily coincide with the desired global minimum. A suitability study with respect to inverse modeling of global optimization algorithms is given in Ref. 75 where two global optimization methods, simulated annealing and genetic optimization, are compared with a Levenberg-Marquardt local gradient-based optimization technique. It was found that among the global optimization strategies that were evaluated, simulated annealing seems to be very well suited for the case of inverse modeling applications.

As a rigorous model calibration can be a cumbersome task, models are often poorly calibrated based on a very limited number of measurements. Thus, the models deliver satisfying results only in the vicinity of the calibration conditions. Due to such a restricted calibration within small subspaces of the model parameter set, miscalibrations are likely to occur meaning that differences between measurements and simulation are fitted with the wrong parameters. Although the model accuracy can be improved for the considered measurements, the accuracy might have even worsened for different operating regions not considered in the calibration. Such calibrations can hardly be carried out by hand and a flexible calibration framework is required.[77]

A typical calibration example is the adjustment of the saturation velocity employed in drift-diffusion models. As for submicron devices the velocity overshoot in the channel can no longer be neglected, it is common practice to increase the saturation velocity to reflect this fact. In a recent publication Bude[3] used $\upsilon_s = 2.2\times10^7$cm/s and $\beta=1$ in the Caughey–Thomas mobility model to fit the velocity profile in the channel of both a 150 nm and a 40 nm nMOS transistor. Furthermore, the zero-field mobility had to be adjusted. Several other examples indicate, that although the distributed quantities inside the devices may be incorrect, integral quantities such as contact currents can be well reproduced. However, these nonphysical values used for the model parameters make predictive simulations very questionable and due to the complex dependencies between the models it is difficult to adjust the appropriate parameters.[78]

IV. CONCLUSIONS

Simulation of semiconductor devices has been part of the design process for nearly forty years. With shrinking feature size, basic models loose their validity, which has led to successively refinement of the models. Furthermore, quantum-mechanical and nonlocal effects gain more importance and have to be properly accounted for. We have reviewed the requirements for successful simulation of modern semiconductor devices. As many of the models employed in process and device simulation are either empirical or contain some simplifying assumptions, an adjustment of the model parameters to fit the process under consideration is required. As has been pointed out, this is a complex task and care must be taken to obtain physically meaningful results.

[1] H. K. Gummel, IEEE Trans. Electron Devices **11**, 455 (1964).
[2] S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer, New York, 1984).
[3] J. D. Bude, in *Simulation of Semiconductor Processes and Devices* (IEEE, New Jersey, 2000), pp. 23–26.
[4] K. Banoo and M. S. Lundstrom, Solid-State Electron. **44**, 1689 (2000).
[5] K. Bløtekjær, IEEE Trans. Electron Devices **17**, 38 (1970).
[6] R. Stratton, Phys. Rev. **126**, 2002 (1962).
[7] S. Liotta and H. Struchtrup, Solid-State Electron. **44**, 95 (2000).
[8] K. Rahmat, J. White, and D. A. Antoniadis, in *International Electron Devices Meeting* (IEEE, New Jersey, 1994), pp. 359–362.
[9] A. Abramo *et al.*, IEEE Trans. Electron Devices **41**, 1646 (1994).
[10] M. Fischetti and S. Laux, Phys. Rev. B **38**, 9721 (1988).
[11] M. Fischetti, IEEE Trans. Electron Devices **38**, 634 (1991).
[12] R. Thoma *et al.*, IEEE Trans. Electron Devices **38**, 1343 (1991).
[13] T. Bordelon, X.-L. Wang, C. Maziar, and A. Tasch, Solid-State Electron. **35**, 131 (1992).
[14] A. Smith and K. Brennan, Solid-State Electron. **39**, 1659 (1996).
[15] D. Ferry, *Semiconductors* (Macmillan, New York, 1991).
[16] E. Azoff, Solid-State Electron. **30**, 913 (1987).
[17] C. Gardner, in *Proceedings of the International Workshop on Computational Electronics* (University of Leeds, Leeds, 1993), pp. 25–36.
[18] K. Hess, *Theory of Semiconductor Devices* (IEEE, New Jersey, 2000).
[19] E. Kane, J. Phys. Chem. Solids **1**, 249 (1957).
[20] S.-C. Lee and T.-W. Tang, Solid-State Electron. **35**, 561 (1992).
[21] Y. Zhang and M. Nokali, Solid-State Electron. **36**, 1689 (1993).
[22] T. Tang, S. Ramaswamy, and J. Nam, IEEE Trans. Electron Devices **40**, 1469 (1993).
[23] E. Azoff, IEEE Trans. Electron Devices **36**, 609 (1989).
[24] M. Lundstrom, *Fundamentals of Carrier Transport, Vol. X of Modular Series on Solid State Devices* (Addison-Wesley, Reading, MA, 1990).
[25] C. Gardner, IEEE Trans. Electron Devices **38**, 392 (1991).
[26] S. Ramaswamy and T.-W. Tang, IEEE Trans. Electron Devices **41**, 76 (1994).
[27] B. Pejčinović *et al.*, IEEE Trans. Electron Devices **42**, 2147 (1995).
[28] R. Cook and J. Frey, Compel **1**, 65 (1982).
[29] G. Baccarani and M. Wordeman, Solid-State Electron. **28**, 407 (1985).
[30] M. Stettler, M. Alam, and M. Lundstrom, IEEE Trans. Electron Devices **40**, 733 (1993).
[31] A. Anile and V. Romano, "Hydrodynamical Modeling of Charge Carrier Transport in Semiconductors," Summer School on Industrial Mathematics, IST Lisboa, Portugal, http://www.dipmat.unict.it/~anile/preprint.html, 1999.
[32] O. Muscato, Physica A (to be published), preprint under http://www.dipmat.unict.it/~anile/preprint.html
[33] D. Woolard, R. Trew, and M. Littlejohn, Solid-State Electron. **31**, 571 (1988).
[34] D. L. Woolard *et al.*, Phys. Rev. B **44**, 11 119 (1991).
[35] R. Stewart and J. Churchill, Solid-State Electron. **33**, 819 (1990).
[36] R. Stewart, L. Ye, and J. Churchill, Solid-State Electron. **32**, 497 (1989).
[37] C. Wilson, IEEE Trans. Electron Devices **35**, 180 (1988).
[38] J. Cao and X. Lei, Solid State Electrons **41**, 1781 (1997).
[39] G. Wolokin and J. Frey, in Proceedings NASECODE VIII Vienna, 1992, pp. 107–108.
[40] P. Scrobohaci and T.-W. Tang, IEEE Trans. Electron Devices **41**, 1197 (1994).
[41] J.-G. Ahn *et al.*, IEEE Trans. Electron Devices **15**, 348 (1994).
[42] T. Bordelon *et al.*, Electron. Lett. **28**, 1173 (1992).
[43] T.-W. Tang and J. Nam, IEEE Electron Device Lett. **19**, 201 (1998).
[44] D. Schroeder, *Modelling of Interface Carrier Transport for Device Simulation* (Springer, New York, 1994).
[45] D. Schroeder and U. Witkowski, IEEE Trans. Electron Devices **44**, 679 (1997).
[46] K. Matsuzawa, K. Uchida, and A. Nishiyama, IEEE Trans. Electron Devices **47**, 103 (2000).
[47] M. Ieong *et al.*, in International Electron Devices Meeting (1998), pp. 733–736.
[48] Z. Yu, R. Dutton, and R. Kiehl, IEEE Trans. Electron Devices **47**, 1819 (2000).
[49] W. Hänsch, T. Vogelsang, R. Kircher, and M. Orlowski, Solid-State Electron. **32**, 839 (1989).
[50] M. van Dort, P. Woerlee, and A. Walker, Solid-State Electron. **37**, 411 (1994).

[51]C.-H. Choi *et al.*, IEEE Trans. Electron Devices **47**, 1843 (2000).
[52]M. Ancona, IEEE Trans. Electron Devices **47**, 1449 (2000).
[53]M. Ancona *et al.*, IEEE Trans. Electron Devices **47**, 2310 (2000).
[54]H. Tsuchiya and T. Miyoshi, IEICE Trans. Electron. **E82-C**, 880 (1999).
[55]C. Fiegna *et al.*, IEEE Trans. Electron Devices **38**, 603 (1991).
[56]S. Mohammad *et al.*, Solid-State Electron. **38**, 807 (1995).
[57]S. Mudanai *et al.*, IEEE Trans. Electron Devices **47**, 1851 (2000).
[58]K. Hasnat *et al.*, IEEE Trans. Electron Devices **44**, 129 (1997).
[59]G. Samudra, B. Seah, and C. Ling, Solid-State Electron. **39**, 753 (1996).
[60]C.-H. Choi, Z. Yu, and R. Dutton, Superlattices Microstruct. **27**, 191 (2000).
[61]P. Habaš, Dissertation, Technische Universität Wien, 1993, http://www.iue.tuwien.ac.at.
[62]C.-L. Huang and N. Arora, IEEE Trans. Electron Devices **40**, 2330 (1993).
[63]N. Arora, IEICE Trans. Electron. **E82-C**, 967 (1999).
[64]A. Chynoweth, Phys. Rev. **109**, 1537 (1958).
[65]K. Souissi *et al.*, IEEE Trans. Electron Devices **40**, 1501 (1993).
[66]R. McIntyre, IEEE Trans. Electron Devices **46**, 1623 (1999).
[67]K. Sonoda, M. Yamaji, K. Taniguchi, C. Hamaguchi, and C. Hamaguchi, J. Appl. Phys. **80**, 5444 (1996).
[68]T. Grasser, H. Kosina, and S. Selberherr, in Proceedings of the International Conference on Modeling and Simulation of Microsystems, Hilton Head Island, South Carolina, 2001, pp. 474–477.
[69]N. Khalil, J. Faricelli, D. Bell, and S. Selberherr, IEEE Electron Device Lett. **16**, 17–19 (1995).
[70]A. Das *et al.*, in International Electron Devices Meeting (1997), pp. 687–690.
[71]C. Chiang, Y. Yeow, and R. Ghodsi, IEEE Trans. Electron Devices **47**, 1385 (2000).
[72]Z. Lee, M. McIlrath, and D. Antoniadis, IEEE Trans. Electron Devices **46**, 1640 (1999).
[73]H. Hayashi, H. Matsuhashi, K. Fukuda, and K. Nishi, IEICE Trans. Electron. **E82-C**, 862 (1999).
[74]C. Richter, A. Hefner, and E. Vogel, IEEE Electron Device Lett. **22**, 35 (2001).
[75]T. Binder, C. Heitzinger, and S. Selberherr, Proc. MSM, Hilton Head, March 2001, pp. 466–469.
[76]H. Sato, K. Aoyama, K. Tsuneno, and H. Masuda, Second International Workshop on Statistical Metrology, Kyoto, Japan, June 1997.
[77]R. Strasser and S. Selberherr, IEICE Trans. Electron. **E83-C**, 1303 (2000).
[78]M. Duane, IEICE Trans. Electron. **E82-C**, 976–982 (1999).