In-Advance CPU Time Analysis for Monte Carlo Device Simulations

C. Jungemann and B. Meinerzhagen

Institut für Theoretische Elektrotechnik und Mikroelektronik
Universität Bremen, Otto-Hahn-Allee, Postfach 33 04 40, D-28334 Bremen, Germany
E-mail: junge@item.uni-bremen.de

Abstract – **In this work it is shown for the first time how to calculate in advance by momentum-based noise simulation for stationary Monte Carlo (MC) device simulations the CPU time, which is necessary to achieve a predefined error level. In addition, analytical expressions for the simulation-time factor of terminal current estimation are given. Without further improvements of the MC algorithm MC simulations of small terminal currents are found to be often prohibitively CPU intensive.**

I. INTRODUCTION

Due to minimization transport in microelectronic devices is governed more and more by nonequilibrium effects which can be simulated by the Monte Carlo (MC) method [1, 2]. In addition to the discretization error, the MC results contain a stochastic error which is inversely proportional to the square root of the CPU time [3, 4, 5]. Although the stochastic error can be controlled automatically, the resultant CPU time is not known beforehand leaving the planning of MC simulations to trial and error, which makes the application of the rather CPU-intensive MC method in a TCAD framework difficult. Furthermore, in some cases it is hard to estimate the stochastic error at all because of strong temporal correlations [5]. In this case the presented methods might be the only way to calculate the stochastic error together with the necessary CPU time.

In this work it is demonstrated for the first time how to estimate the necessary CPU time for stationary MC calculations by momentum-based noise simulations. In addition, analytical approximations are discussed for terminal current estimation and the CPU time is investigated for different devices including an NMOSFET and a SiGe HBT. The proposed methods are validated by comparison with MC results.

II. THEORY

An average $\overline{X}$ calculated by stationary MC simulation for estimating the expected value $\langle X\rangle$ of a stochastic variable X contains a certain stochastic error which can be characterized by the standard deviation $\sigma_{\overline{X}}$. Due to the averaging the probability density of the average can be approximated in many cases by a Gaussian distribution [6], and for a confidence interval of the width 4σ the confidence level is 95.45% [4]. The relative error is defined as the relative half width of the confidence interval

$$r = \frac{2\sigma_{\overline{X}}}{\langle X\rangle} . \tag{1}$$

For an average calculated by integration over time for an ergodic system (T_{sim}: simulated time)

$$\overline{X} = \frac{1}{T_{\text{sim}}} \int_0^{T_{\text{sim}}} X(t)dt \tag{2}$$

the variance σ^2 is given by

$$\begin{aligned} \sigma^2_{\overline{X}} &= \frac{2}{\pi T_{\text{sim}}} \int_0^{\infty} S_{XX}\left(\frac{\nu}{T_{\text{sim}}}\right) \frac{1-\cos(\nu)}{\nu^2} d\nu \\ &\approx \frac{S_{XX}(0)}{T_{\text{sim}}} , \end{aligned} \tag{3}$$

where $S_{XX}(\omega)$ is the spectral intensity, which is the Fourier transform of the autocorrelation function of the fluctuation $X - \langle X\rangle$ and depends on the angular frequency ω [5]. The above approximation holds for sufficiently large T_{sim} and as a consequence the variance is proportional to the spectral intensity at zero frequency.

In the following only MC algorithms will be investigated which are based on uniformly weighted particles (no statistical enhancement) and where many-particle effects can be neglected. In this case the stochastic properties of the MC algorithm are described by the Langevin-Boltzmann equation (LBE) [5, 7]. Based on the LBE the spectral intensities in (3) can be calculated. A disadvantage of this approach is that it is rather time consuming. Therefore, Langevin-type momentum-based models, the drift-diffusion (LDDM) and hydrodynamic model (LHDM) are used here, which are CPU-time efficient approximations of the LBE, where consistency with the LBE is ensured by calculating all transport and noise parameters by bulk MC simulations [8, 9]. Of course, these models can only be used in those cases where stochastic variables are investigated which are covered by these models. These are quantities like the terminal currents or internal distributions like the particle density or velocity.

Since the momentum-based models were developed for the simulation of real electronic noise, the resultant spectral intensities have to be scaled in order to account for the difference in the charge of a simulation particle and a real electron [5, 10]

$$S_{XX}(\omega) = \frac{Q_{\text{tot}}}{2qN_{\text{par}}} W_{XX}(\omega) , \tag{4}$$

where the total particle charge Q_{tot} is the sum of the absolute values of the uniform particle charges (electrons and holes), N_{par} the total particle number, q the physical electron charge, and the factor 2 is due to the different definitions of the power spectrum W_{XX} used in the calculation of electronic noise and

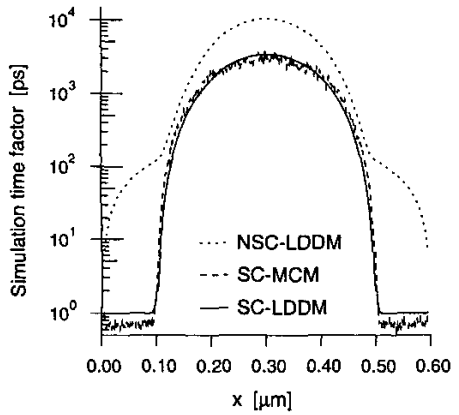

Fig. 1: Simulation-time factors of the electron density for an N^+NN^+ structure biased at zero volts as obtained by self-consistent MC and DD simulations and nonself-consistent DD results.

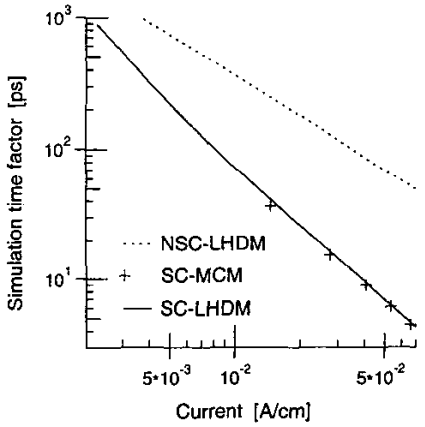

Fig. 2: Simulation-time factors of the terminal current for an N^+NN^+ structure as obtained by self-consistent MC and HD simulations and nonself-consistent HD results.

the spectral intensity. The variance of an average $\overline{X}$ is now given by

$$\sigma^2_{\overline{X}} = \frac{Q_{\text{tot}}}{2qN_{\text{par}}} \frac{W_{XX}(0)}{T_{\text{sim}}} \tag{5}$$

for sufficiently large T_{sim}.

In the case that the CPU time for solving the Poisson equation is negligible compared to the CPU time of the MC simulation of the particles, the total CPU time T_{CPU} is proportional to the particle number N_{par} and the time T_{sim} each particle is simulated

$$T_{\text{CPU}} = \alpha N_{\text{par}} T_{\text{sim}} , \tag{6}$$

where α is the cost factor, which depends on the device structure, bias point, and CPU speed and can be determined by running an MC simulation for a short period of time. Solving (1) with (5) and (6) for a given relative error yields the required CPU time

$$T_{\text{CPU}} = \frac{2\alpha}{r^2} \frac{Q_{\text{tot}}}{q} \frac{W_{XX}(0)}{\langle X \rangle^2} , \tag{7}$$

which depends on the relative error and the investigated quantity X.

The intrinsic simulation-time factor

$$P_X = \frac{r^2}{\alpha} T_{\text{CPU}} = 2 \frac{Q_{\text{tot}}}{q} \frac{W_{XX}(0)}{\langle X \rangle^2} , \tag{8}$$

which is fully determinable by the momentum-based models, is discussed below.

With the simulation-time factor guidelines can be developed how to reduce the CPU time of an MC simulation beyond reducing the cost factor α by improving the program code or using faster computers. One option is the reduction of the total particle charge Q_{tot}, which should be reduced as much as possible by limiting the maximum doping in the nonactive regions of the device and the size of the highly doped contacts. Whether these modifications of the device structure have a negative impact on the simulation results can be checked by momentum-based simulations with and without the modifications. The other option is the reduction of the power spectrum $W_{XX}(0)$. For example, in [11] an unbiased estimator for the substrate current of MOSFETs is given which reduces the spectral intensity compared to the physical estimator. Furthermore, the efficiency of different simulation approaches (e.g. self-consistent (SC) or nonself-consistent (NSC) device simulations) can be investigated as long as the power spectrum W_{XX} can be evaluated with the momentum-based model.

III. RESULTS

In Fig. 1 the simulation-time factor of the electron density is shown for an N^+NN^+ structure at equilibrium and good agreement of the momentum-based simulator with the MC model is obtained for the SC results. Similar good agreement between the MC and momentum-based model is found for the simulation-time factor of the terminal current (Figs. 2 and 3) validating our new approach, where instead of the LDDM the LHDM was used, because it yields better results under nonequilibrium conditions. In addition, in Figs. 1 and 2 the simulation-time factor of NSC simulations is shown, where the electric field is kept constant during the MC simulation (frozen field). In the momentum-based simulators this corresponds to neglecting the derivatives with respect to the electrostatic potential in the small-signal analysis. The NSC simulation-time factors are found to be much larger than the SC ones. Thus, any reduction obtained in the cost factor by the NSC approach might be canceled by the increase in the simulation-time factor and it is not clear whether the NSC approach is more efficient than the SC or not. At least in the case of the N^+NN^+ structure the NSC approach is the less efficient one (see also [12]).

In Fig. 4 the simulation-time factors of the collector and base current of an npn SiGe HBT [9] are shown for a constant collector/emitter bias, where the base current calculation requires much more CPU time than the collector current due to a difference in the magnitude of the currents. In the limit of

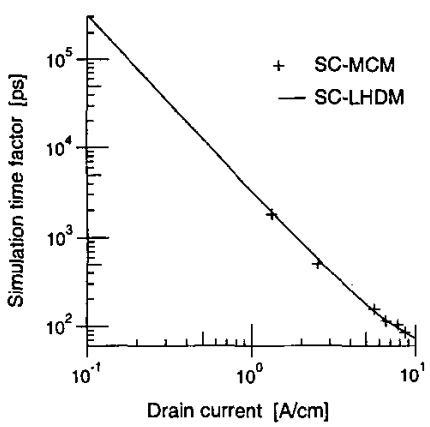

Fig. 3: Simulation time factors for the drain current of an 80nm NMOSFET biased at $V_G = 1.5$V as obtained by MC and HD simulations.

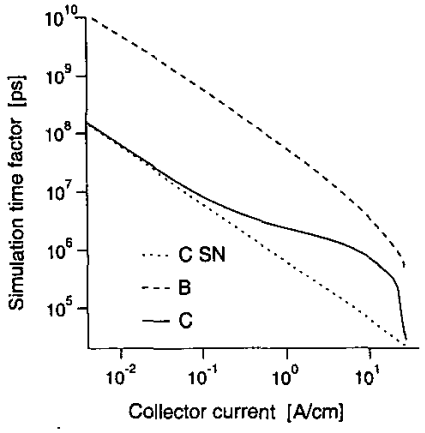

Fig. 4: Simulation time factors for collector (C) and base (B) currents of a 30nm base width HBT biased at $V_{CE} = 2.0$V and the shot noise limit for the collector current (C-SN).

small currents the collector current noise is shot-like

$$W_{I_C I_C}(0) \approx 2qI_C \tag{9}$$

and the simulation-time factor is inversely proportional to the current

$$P_{I_C} \approx \frac{4Q_{\text{tot}}}{I_C} . \tag{10}$$

Thus, the simulation of a ten times smaller current requires ten times more CPU time.

The deviation from the shot-noise formula of the collector-current simulation-time factor for large collector currents (Fig. 4) is due to hole scattering [9]. This increase in CPU time by the hole scattering can be avoided by solving a non-linear Poisson equation together with a constant quasi-Fermi potential for the holes instead of an MC simulation of the holes [13, 14]. In this case (10) also holds for large collector currents up to high injection and the simulation-time factor is considerably reduced.

In Fig. 5 results are shown for a constant base/emitter voltage. With increasing collector/emitter bias impact ionization becomes important and at very high voltages the collector simulation-time factor exceeds the one for the base, although the base current is still about ten times smaller than the collector current.

In Fig. 6 the simulation-time factor is shown for the drain current of an NMOSFET for a constant drain bias as a function of the gate voltage. For example, with $\alpha = 2.5 \cdot 10^{10}$ (1GHz PC) and $V_{\text{gate}} = V_{\text{drain}} = 1.5$V a relative error of $r = 5\%$ is achieved within 860 CPU seconds. In the sub-threshold regime the noise is shot-like and the CPU time is inversely proportional to the drain current similar to (10). For larger gate voltages the noise becomes more like thermal noise

$$W_{I_D I_D}(0) = 4\gamma k_B T G_{DD} , \tag{11}$$

where γ is 2/3 for long channel MOSFETs in saturation [15] and slightly larger for short channel devices, $k_B T$ is the thermal energy, and G_{DD} the drain self-conductance at zero drain/source bias. In this case the CPU time is inversely proportional to the square of the drain current

$$P_{I_D} = \frac{8\gamma k_B T G_{DD} Q_{\text{tot}}}{qI_D^2} . \tag{12}$$

and a ten times smaller current requires a one hundred times larger CPU time to achieve the same relative error. The change from shot to thermal noise behavior is clearly discernible in Fig. 7, where the simulation-time factor is plotted versus the drain current for constant gate voltages. The simulation-time factor is first inversely proportional to the square of the drain current and then inversely proportional to the drain current.

CONCLUSIONS

In conclusion, we have presented a CPU time analysis for stationary MC simulations including analytical expressions for the asymptotical behavior of the simulation-time factor for terminal currents. Based on CPU-efficient momentum-based device simulation we have shown for the first time how to estimate the CPU time of MC simulations in advance. Without further improvements of the MC algorithm MC simulations of small terminal currents are found to be often prohibitively CPU intensive. In the case of the NSC approach it has been shown that although the cost factor is reduced, the CPU time is not necessarily reduced, because the NSC approach increases the simulation-time factor for the examples investigated in this work.

REFERENCES

[1] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, Springer, Wien, 1989.

[2] K. Hess, Ed., *Monte Carlo Device Simulation: Full Band and Beyond*, Kluwer, Boston, 1991.

[3] J. M. Hammarsley and D. C. Handscomb, *Monte Carlo Methods*, Methuen/Chapman and Hall, London, 1964.

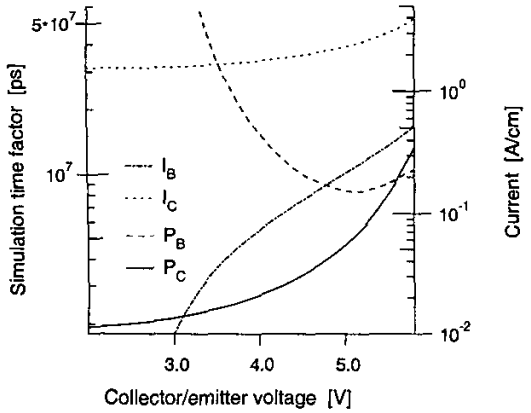

Fig. 5: Simulation time factors (P) and currents (I) for collector (C) and base (B) of a 30nm base width HBT biased at $V_{BE} = 0.85$V.

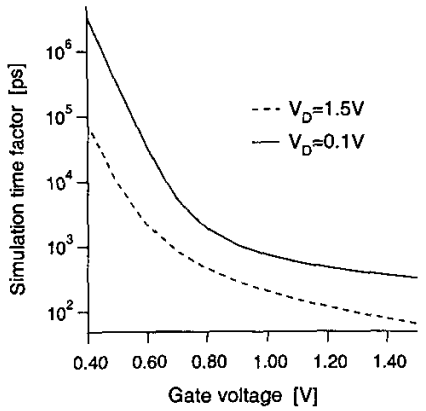

Fig. 6: Simulation time factors for the drain current of an 80nm NMOSFET biased at $V_D = 0.1$V and 1.5V.

[4] C. Jungemann, S. Yamaguchi, and H. Goto, "Convergence estimation for stationary ensemble Monte Carlo simulations," *IEEE J. Tech. Comp. Aided Design*, , no. 10, 1998.

[5] C. Jungemann and B. Meinerzhagen, "Analysis of the stochastic error of stationary Monte Carlo device simulations," *IEEE Trans. Electron Devices*, vol. 48, no. 5, pp. 985–992, 2001.

[6] A. Papoulis, *Probability,Random Variables and Stochastic Processes*, Mc Graw-Hill, 3rd edition, 1991.

[7] Sh. Kogan, *Electronic Noise and Fluctuations in Solids*, Cambridge University Press, Cambridge, New York, Melbourne, 1996.

[8] C. Jungemann, B. Neinhüs, and B. Meinerzhagen, "Hierarchical 2–D DD and HD noise simulations of Si and SiGe devices: Part I—Theory," *to be published in IEEE Trans. Electron Devices*, 2002.

[9] C. Jungemann, B. Neinhüs, S. Decker, and B. Meinerzhagen, "Hierarchical 2–D DD and HD noise simulations of Si and SiGe devices: Part II—Results," *to be published in IEEE Trans. Electron Devices*, 2002.

[10] C. Moglestue, "Monte-Carlo particle modelling of noise in semiconductors," in *International Conference on Noise in Physical Systems and 1/f Fluctuations*, 1983, pp. 23–25.

[11] Jack M. Higman, Karl Hess, Chang G. Hwang, and Robert W. Dutton, "Coupled Monte Carlo-drift diffusion analysis of hot-electron effects in MOSFET's," *IEEE Trans. Electron Devices*, vol. 36, pp. 930–937, 1989.

[12] C. Jungemann and B. Meinerzhagen, "On the applicability of nonself-consistent Monte Carlo device simulations," *to be published in IEEE Trans. Electron Devices*, 2002.

[13] 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, pp. 9721–9745, 1988.

[14] C. Jungemann, B. Neinhüs, and B. Meinerzhagen, "Comparative study of electron transit times evaluated by DD, HD, and MC device simulation for a SiGe HBT," *IEEE Trans. Electron Devices*, vol. 48, no. 10, pp. 2216–2220, 2001.

[15] F. M. Klaassen and J. Prins, "Thermal noise of MOS transistors," *Philips Res. Repts*, pp. 505–514, 1967.

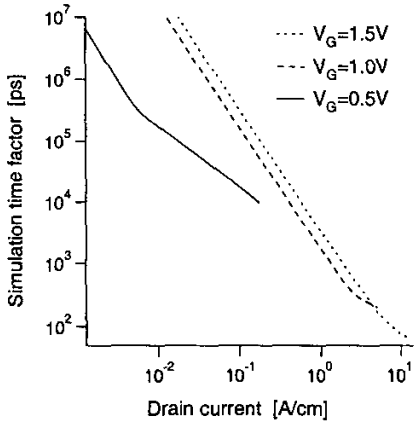

Fig. 7: Simulation time factors for the drain current of an 80nm NMOSFET biased at $V_G = 0.5$V, 1.0V, and 1.5V.