The stationary Monte Carlo method for device simulation. II. Event biasing and variance estimation

M. Nedjalkov, H. Kosina,[a] and S. Selberherr
Institute for Microelectronics, TU Vienna, Gußhausstraße 27-29, A-1040 Vienna, Austria

(Received 28 October 2002; accepted 13 December 2002)

A theoretical analysis of the Monte Carlo method for the solution of the stationary boundary value problem defined by the Boltzmann equation has been presented in Part I. Based on this analysis, the independent, identically distributed random variables of the simulated process are identified. Estimates of the stochastic error of the single-particle Monte Carlo method are derived. An event-biasing technique for carrier transport across an energy barrier is developed and its suitability for variance reduction is demonstrated. © *2003 American Institute of Physics.*
[DOI: 10.1063/1.1544655]

I. INTRODUCTION

In the Monte Carlo (MC) simulation of semiconductor devices, measures have to be taken to enhance the statistics in sparsely populated regions of interest. The purpose of statistical enhancement is to reduce the variance of the MC estimates in those regions. Such a reduction comes at the expense of increased variance in other, less interesting regions.[1] There are two general classes of statistical enhancement techniques, namely, population control techniques and event biasing techniques. To date, virtually all MC device simulation codes utilize population control techniques. On the other hand, the event-biasing technique was introduced in the field of semiconductor transport end of the 1980s, when the *weighted ensemble* MC method was proposed.[2,3] However, until now, this method has not been used in MC device simulation. The present work describes the application of the event-biasing technique to devices with a realistically large carrier concentration range. It is demonstrated that this method is suitable for variance reduction. On the other hand, the method turned out to have an inherent problem with the particle weights, which show a tendency to decrease exponentially in time. A solution to this problem for the steady state is shown.

II. RANDOM VARIABLES

The single-particle MC algorithm has been derived in a formal way in Part I, where, as one significant result, the independent, identically distributed random variables of the simulated process were identified. Knowledge of these random variables forms the basis for estimating the variance of the MC results and thus for estimating the stochastic error.

In the single-particle MC algorithm, a realization of some random variable, say X, is given by a complete numerical trajectory starting and terminating at the domain boundary. The ith realization of x consists of all generated random numbers for the ith trajectory, such as the initial state at the domain boundary, $\mathbf{k}_0$, $\mathbf{r}_0$, the free-flight times t_j, and the after-scattering states, $\mathbf{k}_j^a$

$$x_i=\{\mathbf{k}_0, \mathbf{r}_0, t_0, \mathbf{k}_1^a, t_1, \ldots \mathbf{k}_j^a, t_j, \ldots\}, \quad j\leqslant N_i, \tag{1}$$

where N_i+1 is the number of free-flight segments for the considered trajectory. The random variable $y(X)$ is defined by its realizations

$$y_i=\{\mathbf{k}_1^b, \mathbf{r}_1, \ldots \mathbf{k}_j^b, \mathbf{r}_j, \ldots\}, \quad j\leqslant N_i, \tag{2}$$

where $\mathbf{k}_b^j$ denote the before-scattering states and r_j the particle locations at the times of scattering.

Given a quantity of interest, $A(\mathbf{k}, \mathbf{r})$, a random variable $\psi_A(X)$ is defined, the realization of which is given by the time-averaging estimator

$$\psi_A(x_i)=\sum_{j=0}^{N_i} \int_0^{t_j} A[\mathbf{K}(\tau; 0, \mathbf{k}_j^a, \mathbf{r}_j), \mathbf{R}(\tau; 0, \mathbf{k}_j^a, \mathbf{r}_j)]\, \mathrm{d}\tau. \tag{3}$$

Another random variable $\psi_A^b(Y)$ is defined through the before-scattering estimator

$$\psi_A^b(y_i)=\sum_{j=1}^{N_i} \frac{A(\mathbf{k}_j^b, \mathbf{r}_j)}{\lambda(\mathbf{k}_j^b, \mathbf{r}_j)}, \tag{4}$$

with λ being the total scattering rate. Summation is over one complete trajectory.

III. THE EVENT-BIASING METHOD

Expressing the solution of the conjugate equation by the Neumann series leads to a series expansion of the statistical average of $A(\mathbf{k}, \mathbf{r})$:

$$\langle\langle \mathrm{A} \rangle\rangle=\sum_{n=0}^{\infty} \langle\langle \mathrm{A} \rangle\rangle_n. \tag{5}$$

The explicit expression of a term of that series as given by Eq. (44) in Part I. There exists a variable transformation such that the multiple integral may be expressed as

[a]Electronic mail: kosina@iue.tuwien.ac.at

$$\langle\langle A\rangle\rangle_n = \Gamma_D \int d\xi_0 \ldots d\xi_n\, h(\xi_0) K(\xi_0, \xi_1) \times K(\xi_1, \xi_2) \ldots K(\xi_{n-1}, \xi_n) \frac{A(\xi_n)}{\lambda(\xi_n)}, \quad (6)$$

where the before-scattering states $\xi_i = (\mathbf{k}^b_{i+1}, \mathbf{r}_{i+1})$ are chosen as integration variables. The integrand of Eq. (6) contains an initial distribution h and the transition probability $K(xi_i, \xi_{i+1})$, given by the kernel of the underlying integral equation.

The integral (6) can be written as

$$\langle\langle A\rangle\rangle_n = \Gamma_D \int p(y)\psi(y)dy, \quad (7)$$

$$y = (\xi_0, \xi_1, \ldots \xi_n), \quad (8)$$

$$p = h_0(\xi_0) K(\xi_0, \xi_1) \ldots K(\xi_{n-1}, \xi_n), \quad (9)$$

$$\psi = \frac{A(\xi_n)}{\lambda(\xi_n)}. \quad (10)$$

Since $p(y)$ satisfies the properties of a probability density function, the integral can interpreted as the expected value of a random variable $\psi(y)$. The MC method can now be used to approximate the expected value $E\{\psi\}$ by a sample mean $\overline{\psi} = N^{-1}\Sigma\psi_i$.

In Eq. (6), the initial density h and the transition probability K reflect the physical properties of the system. Therefore, these physically interpreted probability distributions are the natural choice for the construction of the particle trajectory. However, it is possible to choose other than the natural probabilities for the MC integration obeying certain rules. In that case, one constructs numerical trajectories that in particular aim at statistical enhancement, for example, by guiding particles towards a sparsely populated region of interest.

One can choose an arbitrary initial distribution p_0 and arbitrary transition probabilities P for numerical trajectory construction. Since the product $p\psi$ has to remain unchanged, the random variable ψ has to compensate for the changes in the density p.

$$p(y) = p_0(\xi_0) P(\xi_0, \xi_1) \ldots P(\xi_{n-1}, \xi_n), \quad (11)$$

$$\psi = \frac{h_0(\xi_0) K(\xi_0, \xi_1) \ldots K(\xi_{n-1}, \xi_n)}{p_0(\xi_0) P(\xi_0, \xi_1) \ldots P(\xi_{n-1}, \xi_n)} \frac{A(\xi_n)}{\lambda(\xi_n)}. \quad (12)$$

The numerical initial distribution p_0 and the numerical transition probability P have to be nonzero where the physical counterparts are nonzero; that is, $p_0(\xi_0) \neq 0$ if $h_0(\xi_0) \neq 0$, and $P(\xi_i, \xi_j) \neq 0$ if $K(\xi_i, \xi_j) \neq 0$.

Furthermore, only normalized densities are considered: $\int p_0(\xi_0)\, d\xi_0 = 1$ and $\int P(\xi_i, \xi_j)\, d\xi_j = 1$ for all ξ_i.

The ratio of the physically interpreted density over the numerically interpreted density determines the weight of a particle:

$$w_n = \frac{h_0(\xi_0) K(\xi_0, \xi_1) \ldots K(\xi_{n-1}, \xi_n)}{p_0(\xi_0) P(\xi_0, \xi_1) \ldots P(\xi_{n-1}, \xi_n)}. \quad (13)$$

This formula states the rule that, whenever in the process of numerical trajectory construction a random variable, for example, a free-flight time or an after-scattering state, is selected from a numerically constructed density rather than from the physically interpreted density, the weight of the trajectory changes by the ratio of the two densities.

While Eq. (12) is an estimator for one iteration term, in the single-particle Monte Carlo method, one uses estimators for the whole iteration series, such as Eqs. (3) and (4). In these estimators, each element of the sum has to be multiplied by the weight defined earlier. In the case of event biasing, the time-averaging and the before-scattering estimators, Eqs. (3) and (4), respectively, get extended to

$$\psi_A(x_i) = \sum_{j=0}^{N_i} w_j \int_0^{t_j} A[\mathbf{K}(\tau; 0, \mathbf{k}^a_j, \mathbf{r}_j), \mathbf{R}(\tau; 0, \mathbf{k}^a_j, \mathbf{r}_j)]\, d\tau, \quad (14)$$

$$\psi^b_A(y_i) = \sum_{j=1}^{N_i} w_j \frac{A(\mathbf{k}^b_j, \mathbf{r}_j)}{\lambda(\mathbf{k}^b_j, \mathbf{r}_j)}. \quad (15)$$

A. Modified probabilities

The purpose of the event-biasing method is to enhance the statistics in phase-space regions of interest. To guide the particle trajectory towards such regions, various probabilities used for trajectory construction can be modified,[4] including those for selecting the free-flight time, the scattering mechanism, the after scattering state, or the initial state at a contact.

1. Biasing the phonon-absorption probability

On the rising edge of an energy barrier, carrier diffusion can be increased by introducing artificial carrier heating. Controlled by a parameter $M_1 \geqslant 1$, the probability for phonon absorption is increased at the expense of phonon emission:

$$\lambda'_a = \lambda_a + \lambda_e\left(1 - \frac{1}{M_1}\right), \quad \lambda'_e = \frac{\lambda_e}{M_1}. \quad (16)$$

If, in the MC simulation, phonon absorption is selected, the particle weight must be multiplied by λ_a/λ'_a, otherwise by $\lambda_e/\lambda'_e = M_1$. The distribution of the flight time is not affected, because the sum of emission and absorption rate is not changed.

2. Biasing the scattering-angle distribution

Carrier diffusion can also be enhanced by modifying the distribution of the scattering angle. The event bias technique is applied only to isotropic processes. For these, the distribution of $\chi = \cos\theta$ is constant: $p(\chi) = 1/2$ for $\chi \in (-1, 1)$. Here, θ is defined as the angle between the after-scattering momentum and the field direction. The following modified density function increases the probability for forward scattering at the expense of backscattering:

$$p'(\chi) = \begin{cases} \dfrac{1}{2M_2} & -1 \leqslant \chi < \chi_0 \\ \dfrac{M_2}{2} & \chi_0 \leqslant \chi < 1, \end{cases} \quad (17)$$

where $M_2 \geqslant 1$ is a given parameter and χ_0 is determined from the normalization. The cumulative probability at this

point evaluates to $P'(\chi_0)=(1+M_2)^{-1}$. With a random number r, evenly distributed between 0 and 1, one obtains for $r<P'(\chi_0)$,

$$\chi_r=2M_2r-1, \quad \frac{p}{p'}=M_2,$$

and otherwise

$$\chi_r=1-\frac{2(r-1)}{M_2}, \quad \frac{p}{p'}=\frac{1}{M_2}.$$

This means that the particle weight is either reduced or increased by the factor M_2 whenever χ is generated from the density in Eq. (17).

3. Biasing the boundary distribution

In the emitter region, artificial carrier heating can be introduced by increasing the probability for phonon absorption as described earlier. Additionally, at the emitter contact, carriers can be injected from some heated distribution to support the formation of a hot carrier distribution at the emitter–base energy barrier.

Consider a Maxwellian distribution at lattice temperature T_0 and a heated Maxwellian distribution at temperature $T'=M_3T_0$, with $M_3>1$:

$$f_b(\mathbf{k},\mathbf{r})=C(\mathbf{r})\exp\left[-\frac{\varepsilon(\mathbf{k})}{k_BT_0}\right], \tag{18}$$

$$f_b'(\mathbf{k},\mathbf{r})=C'(\mathbf{r})\exp\left[-\frac{\varepsilon(\mathbf{k})}{k_BT'}\right]. \tag{19}$$

The incident current density at some boundary point $\mathbf{r}$ with outward directed normal vector $\mathbf{n}(\mathbf{r})$ is given by

$$j_\perp(\mathbf{r})=-C(\mathbf{r})\int_{\mathbf{n}\cdot\mathbf{v}<0}\mathbf{n}(\mathbf{r})\cdot v(\mathbf{k})\exp\left[-\frac{\varepsilon(\mathbf{k})}{k_BT_0}\right]\mathrm{d}^3k. \tag{20}$$

Substituting the group velocity $v=(1/\hbar)\nabla\varepsilon(\mathbf{k})$ and assuming without loss of generality the x-axis to be parallel to $\mathbf{n}$, one obtains

$$j_\perp(\mathbf{r})=\frac{k_B}{\hbar}C(\mathbf{r})T_0\int_{\mathbf{n}\cdot\mathbf{v}<0}\mathbf{n}\cdot\nabla\exp\left[-\frac{\varepsilon(\mathbf{k})}{k_BT_0}\right]\mathrm{d}^3k, \tag{21}$$

$$=\frac{k_B}{\hbar}C(\mathbf{r})T_0\int\int\exp\left[-\frac{\varepsilon(0,k_y,k_z)}{k_BT_0}\right]\mathrm{d}k_y\mathrm{d}k_z. \tag{22}$$

In the last equation, the integral theorem of Gauss has been applied. From the integral over the closed surface, only the contribution from the (k_y,k_z) plane is nonzero.

Note that Eq. (18) is the physical boundary distribution, and hence the parameter $C(\mathbf{r})$ is known. The normalization factor $C'(\mathbf{r})$ in Eq. (19) is obtained from the condition $j_\perp'=j_\perp$. This requires evaluation of the double integral (22). For simple nonparabolic bands

$$\varepsilon(1+\alpha\varepsilon)=\frac{\hbar^2\mathbf{k}^2}{2m^*}, \tag{23}$$

C' must satisfy

$$\frac{C(\mathbf{r})}{C'(\mathbf{r})}=\frac{T'^2(1+2\alpha T')}{T_0^2(1+2\alpha T_0)}. \tag{24}$$

If the initial momentum is generated from a heated Maxwellian distribution, the initial weight of the particle has to be set to the ratio of the physical probability density over the numerical probability density.

$$w_0=\frac{v_\perp f_b(\mathbf{k},\mathbf{r})}{v_\perp f_b'(\mathbf{k},\mathbf{r})}$$

$$=M_3^2\frac{1+2\alpha M_3T_0}{1+2\alpha T_0}\exp\left[-\frac{\varepsilon(\mathbf{k})}{k_BT_0}\left(1-\frac{1}{M_3}\right)\right]. \tag{25}$$

The velocities $v_\perp$ in the velocity-weighted boundary distributions do not depend on the parameters of the distribution, and therefore cancel.

4. Biasing the injection probabilities

The initial weight (25) can become considerably larger than 1 for a large value of M_3 and not too high energies with respect to k_BT_0. However, particles with high weight are not desirable, especially at high energies, as their occurrence represents rare events which increase the variance. Therefore, it is better to inject the particles at a smaller weight and increase the rate of injection instead.

We consider the iteration term of order n, given by Eq. (44) in Part I, and write only the surface integral explictly, while for the sake of brevity the remaining integral is denoted by some function R_n.

$$\langle\langle \mathrm{A}\rangle\rangle_n=\Gamma_D\oint_{\partial D}\left\{\frac{j_\perp(\mathbf{r})}{\Gamma_D}\right\}R_n(\mathbf{r})\mathrm{d}\sigma(\mathbf{r}). \tag{26}$$

The domain boundary ∂D consists of N_c contacts, at which particles are exchanged with the environment, and Neumann boundaries, at which particles are reflected and hence the normal current density $j_\perp$ vanishes. Thus, one can partition the integral (26),

$$\langle\langle \mathrm{A}\rangle\rangle_n=\Gamma_D\sum_{l=1}^{N_c}\left\{\frac{\Gamma_l}{\Gamma_D}\right\}\int_{A_l}\left\{\frac{j_\perp(\mathbf{r})}{\Gamma_l}\right\}R_n(\mathbf{r})\mathrm{d}\sigma(\mathbf{r}), \tag{27}$$

where A_l denotes the area of the lth contact. The probability of injecting a particle at contact l is given by

$$p_l=\frac{\Gamma_l}{\Gamma_D}, \quad \Gamma_l=\int_{A_l}j_\perp(\mathbf{r})\mathrm{d}\sigma. \tag{28}$$

In a real simulation the p_l are known, for example, from the area and the equilibrium concentration of a contact, whereas the Γ_D and the Γ_l are commonly not known in advance.

In the following, we increase the probability of injection for the mth contact by $M_4>1$ and reformulate Eq. (27) accordingly.

$$\langle\langle \mathrm{A} \rangle\rangle_n = \Gamma_D \left(\sum_{l \neq m} p_l \int_{A_l} + M_4 p_m \frac{1}{M_4} \int_{A_m} \right) \tag{29}$$

$$= \Gamma'_D \left(\sum_{l \neq m} p'_l \int_{A_l} + p'_m \frac{1}{M_4} \int_{A_m} \right), \tag{30}$$

with $\Gamma'_D = s\Gamma_D$ and modified probabilities defined as

$$p'_l = \begin{cases} p_l/s, & l \neq m \\ M_4 p_l/s, & l = m \end{cases}. \tag{31}$$

The normalization factor s is obtained from

$$s = \sum_{l \neq m} p_l + M_4 p_m = 1 + (M_4 - 1) p_m . \tag{32}$$

Using these probabilities, the rate of injection at contact m is increased by M_4 and the rates at all other contacts are reduced accordingly. The factor M_4^{-1} in Eq. (30) determines the initial weight of a particle injected at contact m. The modified normalization coefficient Γ'_D is obtained from computing the average $\langle\langle 1 \rangle\rangle$ (see Sec. IV D in Part I).

One can combine a biased boundary distribution and biased injection probabilities. For example, one is free to choose M_4 equal to the energy-independent part of w_0 given by Eq. (25). Since now the initial weight also contains the multiplier M_4^{-1} from Eq. (30), the initial weight will be reduced:

$$M_4 = M_3^2 \frac{1 + 2\alpha M_3 T_0}{1 + 2\alpha T_0}, \tag{33}$$

$$w_0 = \exp\left[- \frac{\varepsilon(\mathbf{k})}{k_B T_0} \left(1 - \frac{1}{M_3} \right) \right]. \tag{34}$$

The algorithm assuming a modified injection probability and a heated Maxwellian distribution at contact m can be summarized as follows. When a particle has left the simulation domain, a new one must be injected at a contact selected by the probabilities (31). If $l \neq m$, the initial momentum must be generated from a cold Maxwellian distribution with an the initial weight set to one; otherwise the initial momentum must be generated from a heated Maxwellian, and the initial weight set to Eq. (34).

B. Evolution of the weights

The particle weight (13) evolves randomly along a trajectory. MC simulations show that at a given time the weights on different trajectories show a large spreading. Most of the weights evolve to fairly small values, such that new terms in the weighted sums (14) or (15) sooner or later become negligible.

This behavior can be investigated analytically for the simple case of the density function (17), which assumes only two discrete values, say, $0.5M$ and $0.5M^{-1}$. With a probability of $p_0 = (1+M)^{-1}$, the value M is selected as the multiplier of the weight, and with probability $p_1 = M(1+M)^{-1}$ the value M^{-1} is selected, respectively. The expected value of the selected multipliers equals $p_0 M + p_1 M^{-1} = 1$.

Consider a trajectory containing B biased events. On average, the multiplier M will appear $p_0 B$ times, and M^{-1} will appear $p_1 B$ times. The particle weight can be estimated as

$$w_B = M^{p_0 B} \cdot M^{-p_1 B} = \exp(-\alpha B), \tag{35}$$

$$\alpha = \frac{M-1}{M+1} \log M. \tag{36}$$

Since the function $\alpha(M)$ is positive for all positive $M \neq 1$ the weight w_B tends to zero exponentially for $B \to \infty$. An interpretation is that the significance of the trajectory diminishes with increasing number of biased events, as the contributions to the estimator (15) continuously decreases.

As discussed in Part I, setting $A = 1$, the before-scattering estimator (15) gives an estimate for the real time of the trajectory. Assuming a simple physical system with a constant scattering rate Γ, the estimated real time of the trajectory equals

$$T = \sum_n \frac{w_n}{\Gamma} = \frac{1}{\Gamma} \sum_{n=0}^{\infty} \exp(-\alpha n) = \frac{1}{\Gamma[1 - \exp(-\alpha)]}. \tag{37}$$

As the series converges, even a trajectory with infinitely many scattering events covers only a finite physical time interval. Only for $M = 1$, that is, when the physical probability density is used, α vanishes and the particle weight stays constant.

IV. VARIANCE ESTIMATION

Most generally, the result of a stationary MC device simulation can be expressed as a ratio of statistical averages:

$$C = \frac{\langle\langle \mathrm{A} \rangle\rangle}{\langle\langle \mathrm{B} \rangle\rangle}, \tag{38}$$

with the statistical average defined as $\langle\langle \cdot \rangle\rangle = \int_D d\mathbf{r} \int d\mathbf{k} \cdot f(\mathbf{k}, \mathbf{r})$. D denotes the simulation domain.

The random variable to be considered now is given by $\Psi_C = \Psi_A / \Psi_B$. In the MC simulation, one has to generate the samples ψ_{Ai} and ψ_{Bi} using estimators (14) or (15). The so-called classical estimator of C is given by the ratio of the sample means:

$$C \simeq \overline{C} = \frac{\overline{\psi}_A}{\overline{\psi}_B}, \quad \overline{\psi}_\alpha = \frac{1}{N} \sum_{i=1}^{N} \psi_{\alpha i}, \quad \alpha = A, B. \tag{39}$$

Additionally, the sample variances, s_A^2 and s_B^2, and the sample covariance s_{AB}^2 have to be evaluated using the definitions

$$s_\alpha^2 = \frac{1}{N-1} \left[\sum_{i=1}^{N} \psi_{\alpha i}^2 - \frac{1}{N} \left(\sum_{i=1}^{N} \psi_{\alpha i} \right)^2 \right], \quad \alpha = A, B, \tag{40}$$

$$s_{AB}^2 = \frac{1}{N-1} \left[\sum_{i=1}^{N} \psi_{Ai} \psi_{Bi} - \frac{1}{N} \sum_{i=1}^{N} \psi_{Ai} \sum_{i=1}^{N} \psi_{Bi} \right], \tag{41}$$

where N is the number of trajectories constructed in the simulation. From these inputs the variance of the random variable Ψ_C can be estimated.[5]

$$s_C^2 = s_A^2 - 2\overline{C} s_{AB}^2 + \overline{C}^2 s_B^2 . \tag{42}$$

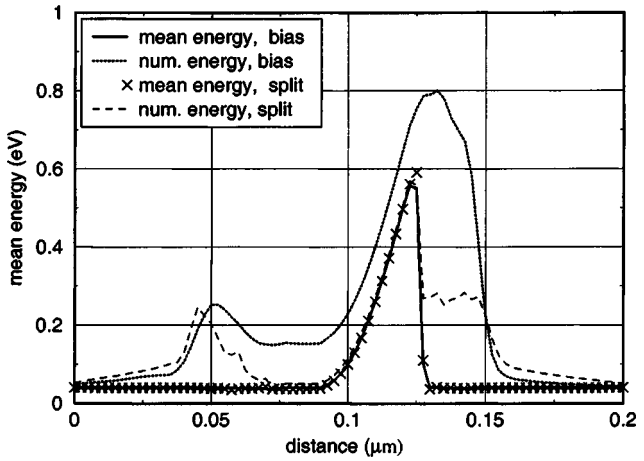

FIG. 1. Mean energy of the physical system (mean) and of the simulated carriers (num.) in the $n-p-n-$structure with 0.8-eV barrier height. Comparison of the event-bias method (W-MC) and a particle-split method is shown.

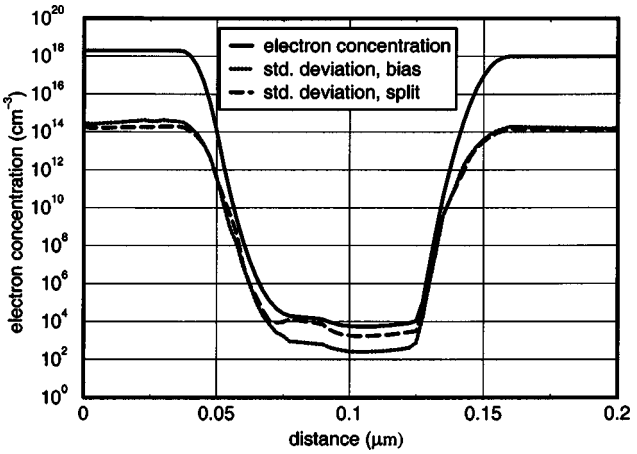

FIG. 3. The electron concentration varies by more than 14 orders. In the base region the event bias method gives significantly less variance than the split method.

The standard deviation of the result $\overline{C}$ is finally given by

$$\sigma\{C\}=\frac{s_C}{\overline{\psi}_B\sqrt{N}}. \tag{43}$$

For a confidence coefficient $(1-\delta)$, a confidence interval $\overline{I}$ can be constructed:[5]

$$\overline{I}=[\overline{C}\pm z_\delta\sigma\{C\}, \tag{44}$$

$$z_\delta=\Phi^{-1}\left(1-\frac{\delta}{2}\right), \tag{45}$$

where Φ denotes the standard normal distribution function. An often used value is $z_\delta=3$, which gives $1-\delta=0.997$.

V. RESULTS AND DISCUSSION

A one-dimensional $n-p-n$ structure has been analyzed. The three segments of the device are referred to as emitter, base, and collector. The semiconductor model of silicon assumes analytical bands and the phonon spectrum reported in Ref. 6.

A. The Event-biasing technique

The modified probabilities described in Sec. III have been used to simulate electron transport through the $n-p-n$ structure, assuming an emitter–base barrier of 0.8 eV and a collector–emitter voltage of 1 V.

To enhance statistics in the emitter–base barrier region, artificial carrier heating is introduced. In the barrier and the base region, the distribution of the scattering angle is biased in order to induce artificial carrier diffusion towards the collector. Optimal values for the parameters M_1 and M_2 controlling the bias are not known *a priori*. For instance, if M_1 is chosen too small, not enough particles will surmount the barrier, rendering statistical enhancement inefficient. Choosing M_1 and M_2 too large, plenty of numerical trajectories will pass through the low-concentration region. However, due to the aggressive bias, the individual particle weights

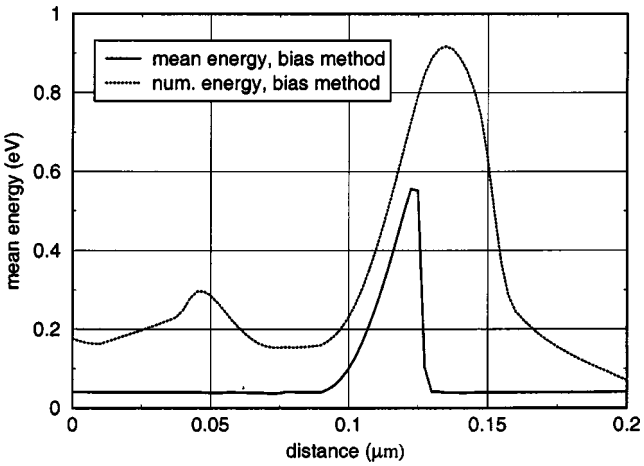

FIG. 2. Mean energy in the 0.8-eV structure. In addition to Fig. 1, electrons are injected at the left contact at 1500 K, whereas at the right contact a Maxwellian at 300 K is assumed.

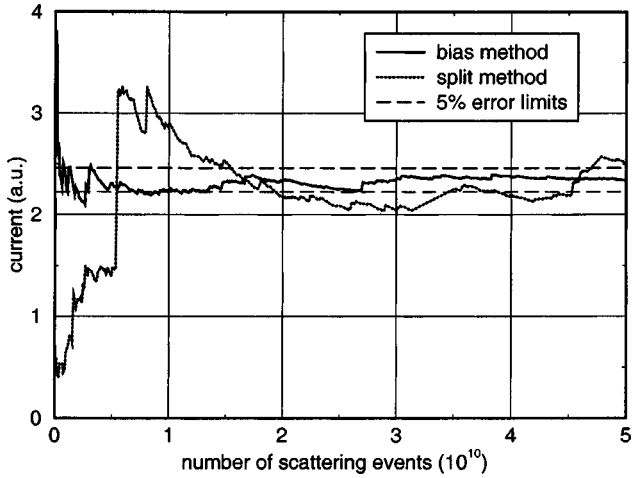

FIG. 4. Evolution of the device current for extremely long simulation times, evaluated every 10^8 scattering events. W-MC converges faster and shows better stability than the split method.

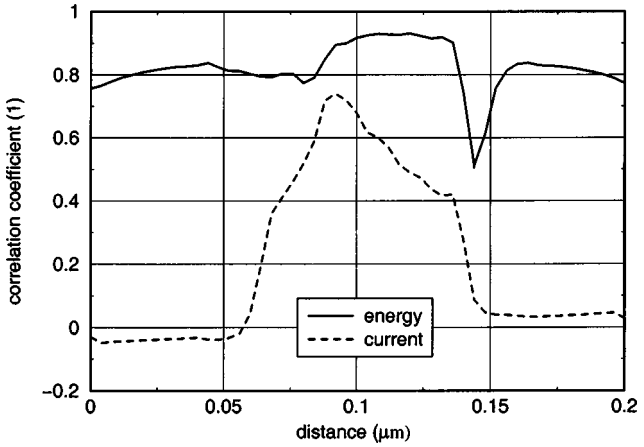

FIG. 5. Correlation coefficients of the the energy density $n\langle\varepsilon\rangle$ and n, and of the particle current $n\langle v_x\rangle$ and n.

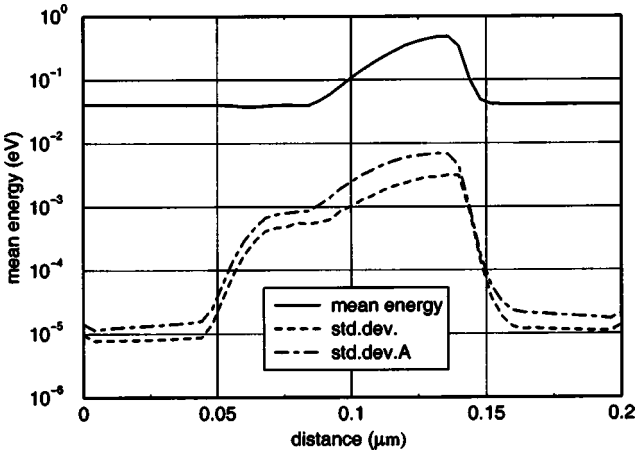

FIG. 6. Mean energy and its standard deviation, including the fluctuations of energy density and particle density (std. dev.), and neglecting the fluctuation of particle density (std. dev. A).

will evolve to extremely different values. Because of the large spreading of the particle weights the recorded averages will again show a large variance. Reasonable values found for the considered structure are $M_1=2$ and $M_2=2$.

The described behavior of the event-biasing technique suggests the usage of additional variance reduction techniques.[7] The general goal must be a reduction of the spreading of the weights. Such techniques are not subject of this study. Here, we investigate the event-bias algorithm, which is predominantly used to govern the evolution of the particle weight. Explicit measures are taken only to prevent weights from getting extremely high or low. The rare event in which a particle gains a very large weight is treated by splitting that particle. On the other hand, when a particle weight falls below a predefined limit, event biasing is disabled such that the weight is not further changed.

The event-biasing method has been compared with a simple particle-split method. To first order such a comparison is fair since the lightweight particles generated with either method are not further recycled. Figure 1 demonstrates for the mean energy that, with event biasing, the correct physical mean values are reproduced. Also shown is the mean energy of the simulated particles, which is considerably higher than the actual physical mean energy.

In the simulation shown in Fig. 2, a biased boundary distribution is also assumed. Electrons are injected from the emitter contact with a Maxwellian distribution at five times the lattice temperature. Again, the correct physical mean energy is obtained. In Fig. 3, the electron concentration and the standard deviations of the two MC methods are depicted. In the quasineutral base region (75–90 nm) event biasing gives a standard deviation reduced by more than one order of magnitude. Figure 4 shows the superior convergence of the event-biasing technique. Because of the poor convergence of the split method for the 0.8-eV barrier structure, 1.5×10^{10} scattering events needed to be processed to permit realistic comparisons.[8]

TABLE I. Examples of physical quantities used in Eq. (38).

Quantity	C	A	B
Carrier concentration	n	$(N_D/V_p)W_p(\mathbf{r})$	1
Current density	$\mathbf{j}$	$(qN_D/V_p)\mathbf{v}(\mathbf{k})W_p(\mathbf{r})$	1
Mean velocity	$\langle\mathbf{v}\rangle$	$\mathbf{v}(\mathbf{k})W_p(\mathbf{r})$	$W_p(\mathbf{r})$
Mean energy	$\langle\varepsilon\rangle$	$\varepsilon(\mathbf{k})W_p(\mathbf{r})$	$W_p(\mathbf{r})$

B. Variance of distributed quantities

The variances of the distributed quantities summarized in Table I have been calculated, using the single-particle MC method without statistical enhancement. The emitter–base barrier of the $n-p-n$ diode was set to 0.2 eV. The simulation of 5×10^8 scattering events resulted in the construction of $N=5.48\times10^7$ trajectories.

For the sake of simplicity, the nearest grid point scheme is used for charge assignment.[9] The assignment function $W_p(x)$ for grid point x_p is given by the indicator function of the finite box associated with that grid point. The box volume is $V_p=\int W_p(\mathbf{r})\mathrm{d}\mathbf{r}$.

When computing the mean velocity using Eq. (38), $\langle\langle A\rangle\rangle$ represents the particle current density and $\langle\langle B\rangle\rangle$ the particle density. The correlation factor of the MC estimate of these two densities is plotted in Fig. 5 (dashed line). In the base region, a large positive correlation exists.

The energy density and the particle density show a large positive correlation throughout the device (Fig. 5, solid line). According to Eq. (42), such a correlation gives a significant reduction in standard deviation. This means that a mean value per carrier has less variance than a mean value per unit volume, as demonstrated in Fig. 6. Accounting for the fluctuations of both the energy density and the particle density gives the lower standard deviation (curve std. dev.), while by neglecting the fluctuation of the particle density the standard deviation is clearly increased (curve std. dev. A).

The presented method of variance estimation is parameter free. Neither has the particle's history to be divided into subhistories of some artificially predefined length,[6] nor does an unknown parameter of a stochastic process, such as the correlation time, need to be estimated.[10] Instead, the total history is divided naturally into independent subhistories at

the times when the particle reenters the simulation domain. A process with such a property is referred to as regenerative stochastic process.[5]

C. Variance of terminal currents

A stationary terminal current can be expressed as an integral over the device domain D:

$$I_l = \int_D \mathbf{j}(\mathbf{r}) \cdot \nabla h_l(\mathbf{r})\, d\mathbf{r}, \tag{46}$$

where h_l is a test function, which assumes 1 at the lth contact and 0 at all other contacts.[11] In this section, we discuss how the choice of the test function and MC estimators affect the variance of the estimated current.

For the one-dimensional structure considered here, it can easily be shown that calculating the current density by the time-averaging method is equivalent to counting particles passing through a normal plane. We consider a control volume formed by the normal planes $x=x_p$ and $x=x_p+\Delta x$. The time integral $\int \mathbf{v}[\mathbf{k}(\tau)]\, d\tau$ over that part of a trajectory which is inside the control volume can take on only three values: $+\Delta x$ if the particle enters through the left plane and exits through the right one, $-\Delta x$ in the opposite case, and 0 if the particle leaves and exits through the same plane. Since only a constant Δx is added or subtracted from the estimator, this method corresponds to particle counting. In the limit $\Delta x \to 0$ particles passing through a single plane are counted.

To determine the current at contact l, we define a random variable which has the realizations ψ_l:

$$\psi_l = \begin{cases} 1 & \text{trajectory starts at contact } l \text{ and terminates at } m \neq l, \\ -1 & \text{trajectory starts at contact } m \neq l \text{ and terminates at } l, \\ 0 & \text{trajectory starts and terminates at contact } l, \\ 0 & \text{trajectory starts at contact } m \neq l \text{ and terminates at } n \neq l. \end{cases} \tag{47}$$

The current is given by the ratio

$$\overline{I_l} = qN_D \frac{\overline{\psi_l}}{\overline{T}}, \quad \overline{T} = \frac{1}{N} \sum_{i=1} NT_i, \tag{48}$$

where $\overline{T}$ is the mean time of a trajectory. Neglecting the variance of $\overline{T}$, the current and its variance can be estimated as

$$\overline{I_l} = \frac{qN_D}{\overline{T}} (P_l^i - P_l^a), \tag{49}$$

$$\sigma^2\{\overline{I}\} = \left(\frac{qN_D}{\overline{T}}\right)^2 \frac{1}{N-1} [P_l^i + P_l^a - (P_l^i - P_l^a)^2], \tag{50}$$

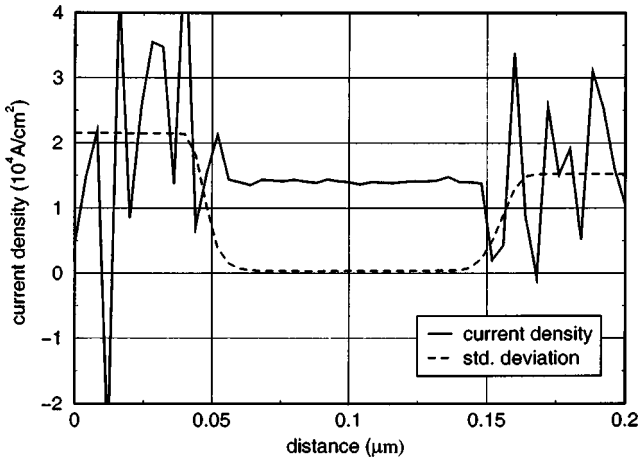

FIG. 7. Current density and standard deviation obtained from the before-scattering method.

with $N_l^i = P_l^i N$ and $N_l^a = P_l^a N$ denoting the numbers of particles injected and absorbed at contact l, respectively.

Particles can be counted at any normal plane inside the device. Since all trajectories start and terminate at the contacts, the estimated current and the variance will be the same, independently of the location of the plane. This means, that the time integration method strictly conserves the current. Since in this example the variance of the current density is constant in the device, the variance of the terminal current cannot be influenced by the choice of the test function in Eq. (46). A general proof of the independence of the variance of the terminal current from the test function can be found in Ref. 12.

On the other hand, the before-scattering method conserves current on the average only. Using this method to estimate the statistical averages of A and B in Table I gives the current density shown in Fig. 7. The variance of the current density is much higher in the emitter/collector regions than in the base region. Therefore, the variance of the terminal current can be minimized in this case by a proper choice of the test function.

Table II shows the relative standard deviation of the terminal current obtained from different methods. The lowest stochastic error is obtained using the simple method of par-

TABLE II. Relative standard deviation of terminal current for different estimators

Estimator	σ_I / I
Before-scattering, average over whole device	28.7×10^{-3}
Before-scattering, average over base only	2.36×10^{-3}
Time-integration (particle counting)	2.19×10^{-3}

ticle counting at the contact, since this method implies strict current conservation. A higher stochastic error is obtained using the before-scattering method, because it conserves current only on the average. The stochastic error depends on the test function, and is always larger than that from the time-integration method.

VI. CONCLUSION

For the single-particle Monte Carlo method, it has been demonstrated that event biasing is a competitive statistical enhancement technique. It can be used as a stand-alone method or in combination with other, presently used variance reduction techniques. Implementation of the method does not require structural changes of an existing code. The independent, identically distributed random variables underlying the single-particle Monte Carlo method have been identified. Based on these random variables, the stochastic error is estimated. Without variance estimation, the statistics can be collected after each free flight of the test particle, whereas variance estimation requires that over one particle trajectory, a substatistics is collected, which is added to the total statistics when the trajectory terminates at the domain boundary. If mean values per particle on a mesh must be computed, correlation of the mean value per volume and the particle number per volume has to be taken into account, which leads to partial cancellation of statistical fluctuations. For the estimation of the stationary terminal current, the simple particle counting method is best suited as it gives lower variance than the before-scattering estimator and does not require the use of weight functions.

ACKNOWLEDGMENT

This work has been supported in part by the IST program, project NANOTCAD, IST-1999-10828.

[1] M. Gray, T. Booth, T. Kwan, and C. Snell, IEEE Trans. Electron Devices **45**, 918 (1998).
[2] P. Poli, L. Rota, and C. Jacoboni, Appl. Phys. Lett. **55**, 1026 (1989).
[3] M. Nedjalkov and P. Vitanov, Solid-State Electron. **33**, 407 (1990).
[4] C. Jacoboni, Tech. Dig. - Int. Electron Devices Meet. **1989**, 469.
[5] R. Rubinstein, Simulation and the Monte Carlo Method (Wiley, New York, 1981).
[6] C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).
[7] C. Wordelman, T. Kwan, and C. Snell, IEEE Trans. Comput.-Aided Des. **17**, 1230 (1998).
[8] H. Kosina, M. Nedjalkov, and S. Selberherr, in Proceedings of the Modeling and Simulation of Microsystems, MSM 2001, edited by M. Laudon and B. Romanowicz (Computational Publications, Boston, 2001), pp. 11–14.
[9] R. Hockney and J. Eastwood, Computer Simulation Using Particles Hilger, Bristol, 1988).
[10] C. Jungemann, S. Yamaguchi, and H. Goto, in Proceedings, Simulation of Semiconductor Processes and Devices, Cambridge, MA, 1997, pp. 209-212.
[11] H. Kim, H. Min, T. Wang, and Y. Park, Solid-State Electron. **34**, 1251 (1991).
[12] C. Jungemann and B. Meinerzhagen, Tech. Dig. - Int. Electron Devices Meet. **2000**, 109.