International Journal of High Speed Electronics and Systems
Vol. 13, No. 3 (2003) 727–769
© World Scientific Publishing Company

Particle Models for Device Simulation

HANS KOSINA and MIHAIL NEDJALKOV
Institute for Microelectronics, TU Vienna, Gusshausstrasse 27-29
Vienna, Austria A-1040
Hans.Kosina@TUWien.ac.at

A theoretical analysis of the Monte Carlo (MC) method for both semiclassical and quantum device simulation is presented. A link between physically-based MC methods for semiclassical transport calculations and the numerical MC method for solving integrals and integral equations is established. The integral representations of the transient and the stationary Boltzmann equations are presented as well as the respective conjugate equations. The structure of the iteration terms of the Neumann series and their evaluation by MC integration is discussed. Using this formal approach the standard MC algorithms and variety of new algorithms are derived, such as the backward and the weighted algorithms, and algorithms for linear small-signal analysis. Applying this theoretical framework to the Wigner-Boltzmann equation enables the development of particle models for quantum transport problems.

Keywords: Monte Carlo method; device simulation; Boltzmann equation; Wigner equation.

1. Introduction

The MC method is now well established for studying semiconductor devices and exploring material properties. The method simulates the motion of charge carriers in the six-dimensional phase space formed by position and momentum. Subjected to the action of an external force field, the point-like carriers follow trajectories governed by Newton's law and the carrier's dispersion relation. These drift processes are interrupted by scattering events, which are assumed local in space and instantaneous in time. The duration of a drift process, the type of scattering mechanism and the state after scattering are selected randomly from given probability distributions characteristic to the microscopic process. The method of generating sequences of drift processes and scattering events appears so transparent from a physical point of view that it is frequently interpreted as a direct emulation of the physical process rather than as a numerical method. In fact, the main MC algorithms employed in device simulation were originally devised from merely physical considerations, viewing a MC simulation as a simulated experiment. This approach works well as long as the charge carriers can be treated as semi-classical particles. Then the carrier system can be described by the Boltzmann equation (BE), and the aforementioned MC method can be shown to implicitly provide a solution to this kinetic equation.

However, when the scope of simulation is extended to quantum devices or highly miniaturized classical devices the wave-like behavior of the carriers can no longer be neglected and a quantum kinetic equation has to be used. For such equations there exist in general no purely physically-based, imitative MC methods. Instead, MC algorithms can only be derived in a more formal, mathematically-based way. First, the kinetic equation under consideration has to be transformed to integral form. The conjugate equation has to be formulated in order to obtain forward MC methods. The Neumann Series of the equation is derived, and the elements of the series are evaluated by means of MC integration.

Using this mathematically-based approach to the BE, in this work the well-known Ensemble MC and the Single-Particle MC methods are derived in a unified way. Additionally, new MC algorithms such as the weighted and the backward algorithm are found. Furthermore, we consider the Wigner-Boltzmann equation for quantum device simulation. Since the quantum distribution function has many properties of a classical distribution function, its equation of motion can be expected to resemble the semi-classical BE so much that part of the vast body of experience on interpreting and solving the BE can be carried over to the quantum case.

This paper is organized as follows. The next Section outlines the numerical MC method for solving integrals and integral equations. MC algorithms for the transient and the stationary BE are derived formally in Section 3 and Section 4, respectively. Section 5 deals with a MC algorithm for the linear small-signal AC analysis of the BE. A recently developed MC algorithm for solving the Wigner-Boltzmann equation, which is based on the theoretical framework described, is presented in Section 6.

2. The Numerical Monte Carlo Method

This section introduces the general scheme of the MC method and outlines its application to the solution of integrals and integral equations.

2.1. *General scheme*

To calculate some unknown value m by the MC method one has to find a random variable ξ whose expected value equals $E\{\xi\} = m$. The variance of ξ is designated $\mathrm{Var}\,\{\xi\} = \sigma^2$ with σ being the standard deviation.

Consider N independent random variables $\xi_1, \xi_2, \ldots, \xi_N$ with distributions identical to that of ξ. Consequently, their expected values and their variance coincide

$$E\{\xi_i\} = m, \qquad \mathrm{Var}\,\{\xi_i\} = \sigma^2, \qquad i = 1, 2, \ldots N \tag{1}$$

Expected value and variance of the sum of all these random variables are given by

$$E\{\xi_1 + \xi_2 + \ldots + \xi_N\} = E\{\xi_1\} + E\{\xi_2\} + \ldots + E\{\xi_N\} = Nm \tag{2}$$

$$\mathrm{Var}\,\{\xi_1 + \xi_2 + \ldots + \xi_N\} = \mathrm{Var}\,\{\xi_1\} + \mathrm{Var}\,\{\xi_2\} + \ldots + \mathrm{Var}\,\{\xi_N\} = N\sigma^2 \tag{3}$$

Using the properties $E\{c\xi\} = cE\{\xi\}$ and $\mathrm{Var}\,\{c\xi\} = c^2\mathrm{Var}\,\{\xi\}$, one obtains from (2) and (3)

$$E\{\frac{1}{N}(\xi_1 + \xi_2 + \ldots + \xi_N)\} = m \tag{4}$$

$$\mathrm{Var}\left\{\frac{1}{N}(\xi_1 + \xi_2 + \ldots + \xi_N)\right\} = \frac{\sigma^2}{N} \tag{5}$$

Therefore, the random variable

$$\overline{\xi} = \frac{1}{N}\sum_{i=1}^{N}\xi_i \tag{6}$$

has the same expected value as ξ and an N times reduced variance. A MC simulation of the unknown m consists of drawing one random number $\overline{\xi}$. Indeed, this is equivalent to drawing N values of the random variable ξ, and evaluating the sample mean (6).

The MC method gives an estimate of both the result and the error. According to the central limit theorem the sum $\rho_N = \xi_1 + \xi_2 + \ldots + \xi_N$ of a large number of identical random variables is approximately normal. For this reason, the following *three-sigma* rule holds only approximately

$$P\{|\rho_N - Nm| < 3\sqrt{N\sigma^2}\} \approx 0.997\,. \tag{7}$$

In this equation the expected value and the variance of ρ_N are given by (2) and (3), respectively. Dividing the inequality by N and using $\overline{\xi} = \rho_N/N$ we arrive at an equivalent inequality and the probability will not change:

$$P\left\{|\overline{\xi} - m| < 3\frac{\sigma}{\sqrt{N}}\right\} \approx 0.997 \tag{8}$$

This formula indicates that the sample mean $\overline{\xi}$ will be approximately equal to m. The error of this approximation will most probably not exceed the value $3\sigma/\sqrt{N}$. This error evidently approaches zero as N increases[1].

2.2. *Monte Carlo integration*

We apply the MC method to the evaluation of an integral.

$$m = \int_a^b \phi(x)\,\mathrm{d}x \tag{9}$$

For this purpose the integrand has to be decomposed into a product $\phi = p\,\psi$, where p is a density function, which means that p is non-negative and satisfies $\int_a^b p(x)\,\mathrm{d}x = 1$. Integral (9) becomes

$$m = \int_a^b p(x)\psi(x)\,\mathrm{d}x, \tag{10}$$

and denotes the expected value of some random variable Ψ: $m = E\{\Psi\}$. Now the general scheme described in the previous section can be applied. First, a sample $x_1, \ldots, x_N$ is generated from the density p. Then the sample $\psi_1, \ldots, \psi_N$ is obtained by evaluating the function ψ: $\psi_i = \psi(x_i)$. The sample mean

$$m \simeq \overline{\psi} = \frac{1}{N} \sum_{i=1}^{N} \psi_i \tag{11}$$

approximates the expected value. To employ the error estimation (8) the variance of Ψ can be approximately evaluated by the sample variance

$$\sigma^2 \simeq \overline{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\psi_i - \overline{\psi})^2 \tag{12}$$

Since the factorization of the integrand is not unique different random variables can be introduced depending on the choice of the density p. All of them have the same expected value, but different variance.

2.3. *Integral equations*

The kinetic equations considered in this work can be reformulated in terms of integral equations of the form

$$f(x) = \int f(x') K(x', x) \, \mathrm{d}x' + f_0(x) \,, \tag{13}$$

where the kernel K and the free term f_0 are given functions. Equations of this form are known as *Fredholm integral equations of the second kind*. In the particular case of the Boltzmann equation the unknown function f represents the distribution function. The multi-dimensional variable x stands for $(\mathbf{k}, \mathbf{r}, t)$ in the transient case and for $(\mathbf{k}, \mathbf{r})$ in the steady state.

Substituting (13) recursively into itself gives the *Neumann series*, which is a formal solution to the integral equation[2].

$$f = f^{(0)} + f^{(1)} + f^{(2)} + \ldots \tag{14}$$

The iteration terms are defined recursively beginning with $f^{(0)}(x) = f_0(x)$.

$$f^{(n+1)}(x) = \int f^{(n)}(x') K(x', x) \, \mathrm{d}x' \,, \qquad n = 0, 1, 2, \ldots \tag{15}$$

The series (14) yields the function value in some given point x. However, in many cases one is interested in mean values of f rather than in a point-wise evaluation. Such a mean value represents a linear functional and can be expressed in terms of an inner product.

$$(f, A) = \int f(x) A(x) \mathrm{d}x \tag{16}$$

It is to note that (13) is a backward equation. The corresponding forward equation is given by the conjugate equation,

$$g(x') = \int g(x) K^{\dagger}(x, x') \, \mathrm{d}x + A(x'), \tag{17}$$

where the kernel is defined by $K^{\dagger}(x, x') = K(x', x)$. Multiplying (13) by $g(x)$ and (17) by $f(x')$, and integrating over x and x', respectively, results in the equality

$$(f, A) = (g, f_0) \, . \tag{18}$$

By means of (18) one can calculate a statistical mean value not only from f, but also from g, the solution of the conjugate equation. The given function A has to be used as the free term of the conjugate equation. The link with the numerical MC method is established by evaluating the terms of the Neumann series by MC integration, as pointed out in the previous section.

Note that usage of (18) precludes a point-wise evaluation of the distribution function using a forward algorithm, because $A(x) = \delta(x)$ cannot be treated by the MC method. The probability for a continuous random variable x' to assume a given value x is zero. Only the probability of finding x' within a small but finite volume around x is non-zero.

3. The Transient Boltzmann Equation

On a semi-classical level the transport of carriers in semiconductors can be well described by the BE. For device simulation the time- and position-dependent BE needs to be considered.

$$\left(\frac{\partial}{\partial t} + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} + \mathbf{F}(\mathbf{r}, t) \cdot \nabla_{\mathbf{k}} \right) f(\mathbf{k}, \mathbf{r}, t) = Q[f](\mathbf{k}, \mathbf{r}, t) \, , \qquad \mathbf{r} \in D \tag{19}$$

This equation is posed in the simulation domain D and has to be supplemented by boundary and initial conditions. In semiconductor physics the distribution function is commonly normalized as

$$\frac{1}{4\pi^3} \int\limits_{D} \mathrm{d}\mathbf{r} \int \mathrm{d}\mathbf{k} \, f(\mathbf{k}, \mathbf{r}, t) = N_D(t), \tag{20}$$

where N_D denotes the number of carriers contained in the semiconductor domain of volume V_D. This normalization is based on the notion of discrete states in $\mathbf{k}$-space having a of density $2V_D/(2\pi)^3$, such that f can be viewed both as an occupation probability of the discrete state $\mathbf{k}$ and a density function in the continuous $\mathbf{k}$-space. In both cases, however, with respect to $\mathbf{r}$, f is to be interpreted as a density function.

In (19) the carrier's group velocity $\mathbf{v}$ is related to the band energy $\epsilon(\mathbf{k})$ by $\mathbf{v} = \hbar^{-1} \nabla_k \epsilon(\mathbf{k})$. The force field $\mathbf{F}$ takes into account electric and magnetic fields. If only an electric field $\mathbf{E}$ is present, the force field is given by $\mathbf{F} = q\mathbf{E}/\hbar$, where q

is the charge of the carrier. The scattering operator $Q = Q_g - Q_l$ consists of a gain and a loss term, respectively. If many-body effects such as carrier-carrier scattering and degeneracy are neglected, the scattering operator will be linear, an assumption that is crucial for the presented approach. The two components of Q are

$$Q_g[f](\mathbf{k}, \mathbf{r}, t) = \int f(\mathbf{k}', \mathbf{r}, t) S(\mathbf{k}', \mathbf{k}, \mathbf{r}, t) \, \mathrm{d}\mathbf{k}' \,, \tag{21}$$

$$Q_l[f](\mathbf{k}, \mathbf{r}, t) = \lambda(\mathbf{k}, \mathbf{r}, t) f(\mathbf{k}, \mathbf{r}, t) \,, \tag{22}$$

with $\lambda(\mathbf{k}, \mathbf{r}, t) = \int S(\mathbf{k}, \mathbf{k}', \mathbf{r}, t) \, \mathrm{d}\mathbf{k}'$ denoting the total scattering rate.

3.1. *Transient MC algorithms*

In a bulk semiconductor transient transport phenomena occur when the carrier system evolves from an initial to some final distribution. Accordingly, the evolution of an ensemble of test particles has to be simulated starting from a given initial condition. Macroscopic quantities of interest are obtained by calculating ensemble averages, giving rise to the name *ensemble MC* (EMC) method[3,4]. For example, the distribution function in a given phase space point at a given time is estimated by the relative number of particles in a small volume around the point.

For a space-dependent problem the EMC algorithm has to take into account boundary conditions. A homogeneous Neumann boundary is realized by simply reflecting particles at the boundary. Physical models for ohmic contacts typically enforce local charge neutrality[5]. A proof that the EMC algorithm solves the transient BE has been given by Reklaitis[6] and can be found also in Ref.[7].

In the field of semiconductor transport, the formal way to use the BE as a starting point and to formulate MC algorithms for its solution has been reported by end of the 1980's. In [8,9] the BE is transformed into an integral equation which is then iteratively substituted into itself. For the evaluation of the resulting iteration series a new MC technique is proposed, called MC backward (MCB) since the trajectories are followed back in time. All trajectories start from a chosen phase space point, and their number is freely adjustable and not controlled by the physical process. The MCB method allows for the evaluation of the distribution function in a given point with a desired precision. The algorithm offers advantages when rare events have to be simulated or when the distribution function is needed only in a small phase space domain.

In the *weighted ensemble* MC (WEMC) method arbitrary probabilities can be used for trajectory construction in order to guide particles towards a region of interest[10,11]. The method evaluates the iteration series of an integral form of the BE originally given by Chambers[12]. The unbiased estimator for the distribution function contains a product of weights which are given by the ratio of the real and the modified probabilities of the selected events.

With the iteration approach[8] the MCB and the WEMC algorithms are stated in a unified way. The common EMC algorithm is obtained as a particular case, in

which the numerical trajectories coincide with the physical carrier trajectories, for bulk transport[13] and space-dependent conditions[14]. Moreover, a convergence proof for the underlying Neumann Series[15] and variance estimates for the EMC method[16] have been reported.

3.2. *Integral form of the Boltzmann equation*

In this section the BE is transformed from integro-differential form to integral form. The equations of motion in phase space are given by Newton's law (23) and the carrier's group velocity (24).

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{K}(t) = \mathbf{F}(\mathbf{R}(t), t) \tag{23}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{R}(t) = \mathbf{v}(\mathbf{K}(t)) \tag{24}$$

A phase space trajectory with the initial condition $\mathbf{K}(t_0) = \mathbf{k}_0$ and $\mathbf{R}(t_0) = \mathbf{r}_0$ is obtained by formal integration.

$$\mathbf{K}(\tau) = \mathbf{k}_0 + \int_{t_0}^{\tau} \mathbf{F}\left(\mathbf{R}(y), y\right) \mathrm{d}y$$

$$\mathbf{R}(\tau) = \mathbf{r}_0 + \int_{t_0}^{\tau} \mathbf{v}\left(\mathbf{K}(y)\right) \mathrm{d}y \tag{25}$$

Assume a particle to be located in the phase space point $\mathbf{k}$, $\mathbf{r}$ at time t. The given point $\mathbf{k}, \mathbf{r}, t$ determines uniquely a phase space trajectory $\mathbf{K}(t')$ and $\mathbf{R}(t')$. The left-hand side of (19) represents the total derivative of the function

$$\hat{f}(t') = f(\mathbf{K}(t'), \mathbf{R}(t'), t')$$

with respect to time, which allows the BE to be rewritten as an ordinary differential equation of first order:

$$\frac{\mathrm{d}}{\mathrm{d}t'}\hat{f}(t') + \hat{\lambda}(t')\hat{f}(t') = \hat{Q}_g[f](t'). \tag{26}$$

The structure of the BE is more compact if multiplied by an integrating factor of the form $\exp(\int_0^{t'} \hat{\lambda}(y)\mathrm{d}y)$.

$$\frac{\mathrm{d}}{\mathrm{d}t'} \exp\left(\int_0^{t'} \hat{\lambda}(y)\mathrm{d}y\right) \hat{f}(t') = \exp\left(\int_0^{t'} \hat{\lambda}(y)\mathrm{d}y\right) \hat{Q}_g[f](t') \tag{27}$$

This equation can be integrated straight forwardly in the bounds 0 and t.

$$\hat{f}(t) = \int_0^t \hat{Q}_g[f](t') \exp\left(-\int_{t'}^{t} \lambda(y)\mathrm{d}y\right) \mathrm{d}t' + \hat{f}(0) \exp\left(-\int_0^t \lambda(y)\mathrm{d}y\right) \tag{28}$$

Substituting $\hat{f}(0) = f(\mathbf{K}(0), \mathbf{R}(0), 0)$ and $\hat{f}(t) = f(\mathbf{K}(t), \mathbf{R}(t), t) = f(\mathbf{k}, \mathbf{r}, t)$ one obtains the integral form of the BE.

$$f(\mathbf{k},\mathbf{r},t) = \int_0^t \mathrm{d}t' \int \mathrm{d}\mathbf{k}'\ f(\mathbf{k}', \mathbf{R}(t'), t') \times S(\mathbf{k}', \mathbf{K}(t'), \mathbf{R}(t'), t') \exp\left(- \int_{t'}^{t} \lambda(\mathbf{K}(y), \mathbf{R}(y), y) \mathrm{d}y \right) + f(\mathbf{K}(0), \mathbf{R}(0), 0) \exp\left(- \int_0^t \lambda(\mathbf{K}(y), \mathbf{R}(y), y) \mathrm{d}y \right) \quad (29)$$

This equation represents the generalized form of Chamber's path integral[12].

In (29) the term $S(\mathbf{k}', \mathbf{k}_f)\mathrm{d}\mathbf{k}'\mathrm{d}t'$ denotes the probability of a transition from an initial state $\mathbf{k}'$ within the volume element $\mathrm{d}\mathbf{k}'$ to the given final state $\mathbf{k}_f$ during the interval $\mathrm{d}t'$. This probability will in general be different from the forward transition probability $S(\mathbf{k}', \mathbf{k}_f)\mathrm{d}\mathbf{k}_f\mathrm{d}t'$, where the initial state is fixed and the final state is within some volume element $\mathrm{d}\mathbf{k}_f$. To obtain a forward MC algorithm, one has to change from integration over initial states to integration over final states, a task calling for the introduction of the conjugate equation. The derivation of this equation begins with a transformation of (29) to standard form (13) by augmenting the kernel.

$$f(\mathbf{k},\mathbf{r},t) = \int_0^\infty \mathrm{d}t' \int \mathrm{d}\mathbf{k}' \int \mathrm{d}\mathbf{r}'\ f(\mathbf{k}', \mathbf{r}', t') K(\mathbf{k}', \mathbf{r}', t', \mathbf{k}, \mathbf{r}, t) + f_0(\mathbf{k}, \mathbf{r}, t) \quad (30)$$

$$K(\mathbf{k}', \mathbf{r}', t', \mathbf{k}, \mathbf{r}, t) = S(\mathbf{k}', \mathbf{K}(t'), \mathbf{r}', t') \exp\left(- \int_{t'}^{t} \lambda(\mathbf{K}(y), \mathbf{R}(y), y) \mathrm{d}y \right) \times \delta(\mathbf{r}' - \mathbf{R}(t')) H(t - t') \quad (31)$$

$$f_0(\mathbf{k},\mathbf{r},t) = f_i(\mathbf{K}(0), \mathbf{R}(0)) \exp\left(- \int_0^t \lambda(\mathbf{K}(y), \mathbf{R}(y), y) \mathrm{d}y \right) \quad (32)$$

H denotes the unit step function and f_i the initial distribution. The integral form (29) is immediately recovered from (30) by performing the $\mathbf{r}'$ integration and replacing the upper bound of the time integral by t, thus eliminating the unit step function.

Using the kernel (31) in the defining equation (17) the conjugate equation evaluates to

$$g(\mathbf{k}',\mathbf{r}',t') = \int_{t'}^{\infty} \mathrm{d}\tau \int \mathrm{d}\mathbf{k}_a S(\mathbf{k}',\mathbf{k}_a,\mathbf{r}',t') \exp\Big(-\int_{t'}^{\tau} \lambda(\mathbf{K}(y),\mathbf{R}(y),y)\mathrm{d}y\Big) \times g(\mathbf{K}(\tau),\mathbf{R}(\tau),\tau) + g_0(\mathbf{k}',\mathbf{r}',t')\,. \quad (33)$$

To obtain this equation, in (31) variables are changed, $\mathbf{k}_a = \mathbf{K}(t')$, $\mathbf{r}'' = \mathbf{R}(t')$. According to the Liouville theorem the volume element does not change over a trajectory such that $\mathrm{d}\mathbf{k}\mathrm{d}\mathbf{r} = \mathrm{d}\mathbf{k}_a\mathrm{d}\mathbf{r}''$. The $\mathbf{r}''$ integration is then carried out with the help of the δ-function.

The solution g for free term $g_0 = \delta(\mathbf{k}-\mathbf{k}')\delta(\mathbf{r}-\mathbf{r}')\delta(t-t')$ represents the Green's function of the BE. From (18) is follows that the solution f of (30) is given by the scalar product

$$f(\mathbf{k},\mathbf{r},t) = \int_0^{\infty} \mathrm{d}t' \int \mathrm{d}\mathbf{k}' \int \mathrm{d}\mathbf{r}'\; g(\mathbf{k}',\mathbf{r}',t';\mathbf{k},\mathbf{r},t) f_0(\mathbf{k}',\mathbf{r}',t')\,. \quad (34)$$

3.3. *The ensemble MC method*

Assume we are interested in the integral of f over some phase space sub-domain Ω at time t:

$$f_\Omega(t) = \int_0^{\infty} \mathrm{d}t' \int \mathrm{d}\mathbf{k}' \int \mathrm{d}\mathbf{r}'\; f(\mathbf{k}',\mathbf{r}',t')\delta(t-t')\theta_\Omega(\mathbf{k}',\mathbf{r}')\,, \quad (35)$$

where θ_Ω denotes the indicator function of the sub-domain. Considering this integral as scalar product $f_\Omega(t) = (f, g_0)$ it follows that $g_0 = \delta(t-t')\theta_\Omega(\mathbf{k}',\mathbf{r}')$. Using the Neumann series of (33), $g = \sum_0^\infty g^{(i)}$, we obtain

$$f_\Omega(t) = (f_0, g) = f_\Omega^{(0)}(t) + f_\Omega^{(1)}(t) + f_\Omega^{(2)}(t) + \ldots \quad (36)$$

The meaning of the iteration terms can be understood from their structure. The second iteration term, for example, can be expressed as

$$\begin{aligned} f_\Omega^{(2)}(t) = \int_0^t \mathrm{d}t_2 \int_{t_2}^t \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_2^a \int \mathrm{d}\mathbf{k}_1^a \int \mathrm{d}\mathbf{k}_i \int \mathrm{d}\mathbf{r}_i\; f_i(\mathbf{k}_i,\mathbf{r}_i) \\ \times \exp\Big(-\int_0^{t_2} \lambda(\mathbf{K}_2(y),\mathbf{R}_2(y),y)\mathrm{d}y\Big) S(\mathbf{K}_2(t_2),\mathbf{k}_2^a,\mathbf{R}_2(t_2),t_2) \\ \times \exp\Big(-\int_{t_2}^{t_1} \lambda(\mathbf{K}_1(y),\mathbf{R}_1(y),y)\mathrm{d}y\Big) S(\mathbf{K}_1(t_1),\mathbf{k}_1^a,\mathbf{R}_1(t_1),t_1) \\ \times \exp\Big(-\int_{t_1}^{t} \lambda(\mathbf{K}_0(y),\mathbf{R}_0(y),y)\mathrm{d}y\Big) \theta_\Omega(\mathbf{K}_0(t),\mathbf{R}_0(t)) \end{aligned} \quad (37)$$

Initial conditions for the **k**-space trajectories are given first by $\mathbf{K}_2(0) = \mathbf{k}_i$ and then by the after-scattering states $\mathbf{K}_1(t_2) = \mathbf{k}_2^a$ and $\mathbf{K}_0(t_1) = \mathbf{k}_1^a$ (see Fig. 1). The real space trajectory starts initially from $\mathbf{R}_2(0) = \mathbf{r}_i$ and is continuous at the time of scattering: $\mathbf{R}_2(t_2) = \mathbf{R}_1(t_2)$, $\mathbf{R}_1(t_1) = \mathbf{R}_0(t_1)$.

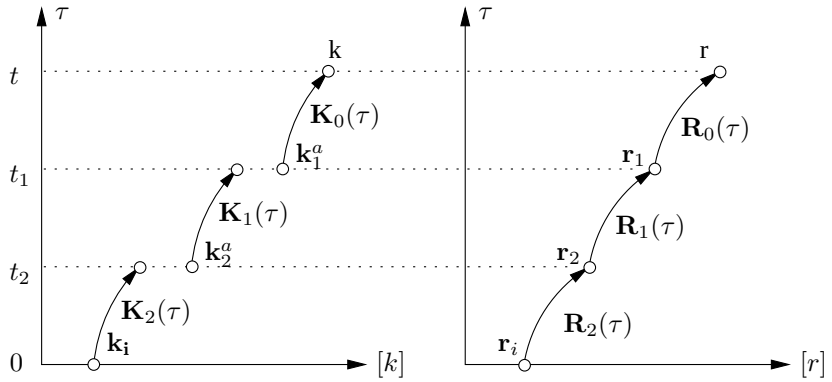

Figure 1: Sketch of a forward trajectory which reaches the evolution time t after three free flights. The symbols used in (37) are shown.

The iteration term (37) describes the contribution of all particles that undergo two scattering events when they propagate from time 0 to t. At time t we find the particles on their third free-flight path. Analogously, $f_\Omega^{(0)}$ represents the contribution of all particles which propagate without scattering from 0 to t, $f_\Omega^{(1)}$ the contribution of all particles that propagate with one scattering event, and so forth.

The next task is to separate the integrand in (37) into a probability density p and a random variable ψ according to (9). To accomplish this task the integrand is augmented in two steps. First, the term $\exp(-\int_{t_1}^{t})$, which represents the probability that the particle drifts without scattering from t_1 to t, is expressed as an integral over the corresponding probability density. For the sake of brevity the time and position-dependence of the scattering rates is not written explicitly.

$$\exp\left(-\int_{t_1}^{t}\lambda(\mathbf{K}_0(y))\mathrm{d}y\right) = \int_{t}^{\infty}\lambda(\mathbf{K}_0(t_0))\exp\left(-\int_{t_1}^{t_0}\lambda(\mathbf{K}_0(y))\mathrm{d}y\right)\mathrm{d}t_0 \qquad (38)$$

Second, products of the form $\exp(-\int) S$ in (37) are multiplied by $\lambda\lambda^{-1}$. These changes yield the following expression for the second iteration term.

$$
\begin{aligned}
f_\Omega^{(2)} = &\int_0^t \mathrm{d}t_2 \int_{t_2}^t \mathrm{d}t_1 \int_t^\infty \mathrm{d}t_0 \int \mathrm{d}\mathbf{k}_2^a \int \mathrm{d}\mathbf{k}_1^a \int \mathrm{d}\mathbf{k}_i \int \mathrm{d}\mathbf{r}_i \Big\{ f_i(\mathbf{k}_i, \mathbf{r}_i) \Big\} \\
&\times \Big\{ \lambda(\mathbf{K}_2(t_2)) \, \exp\Big(- \int_0^{t_2} \lambda(\mathbf{K}_2(y)) \mathrm{d}y \Big) \Big\} \Big\{ \frac{S(\mathbf{K}_2(t_2), \mathbf{k}_2^a)}{\lambda(\mathbf{K}_2(t_2))} \Big\} \\
&\times \Big\{ \lambda(\mathbf{K}_1(t_1)) \, \exp\Big(- \int_{t_2}^{t_1} \lambda(\mathbf{K}_1(y)) \mathrm{d}y \Big) \Big\} \Big\{ \frac{S(\mathbf{K}_1(t_1), \mathbf{k}_1^a)}{\lambda(\mathbf{K}_1(t_1))} \Big\} \\
&\times \Big\{ \lambda(\mathbf{K}_0(t_0)) \, \exp\Big(- \int_{t_1}^{t_0} \lambda(\mathbf{K}_0(y)) \mathrm{d}y \Big) \Big\} \theta_\Omega(\mathbf{K}_0(t), \mathbf{R}_0(t))
\end{aligned} \tag{39}
$$

Probability densities assigned to elementary events, such as generation of an initial state, of a free-flight time or an after scattering state are enclosed in curly brackets for easier recognition. Note that $\lambda^{-1}S$ is the normalized distribution of the after-scattering states, since $\int \lambda^{-1}(\mathbf{k}) S(\mathbf{k}, \mathbf{k}') \mathrm{d}\mathbf{k}' = 1$ for all $\mathbf{k}$. The free-flight time distributions are normalized on semi-infinite time intervals. For instance, for t_1 we get

$$
\int_{t_2}^\infty p(t_1) \, \mathrm{d}t_1 = \int_{t_2}^\infty \lambda(\mathbf{K}_1(t_1)) \exp\Big(- \int_{t_2}^{t_1} \lambda(\mathbf{K}_1(y)) \mathrm{d}y \Big) \, \mathrm{d}t_1 = 1 \tag{40}
$$

None of the distributions for t_0, t_1, t_2 is normalized in the time intervals given in (39) which reflects the simple fact that $f_\Omega^{(2)}$ does not represent the whole solution f_Ω but only a partial contribution.

When the multiple integrals of the iteration terms are evaluated by MC integration the well-known EMC algorithm is recovered. The separation of the integrand into a p and ψ follows quite naturally from (39).

$$
x = (\mathbf{k}_i, \mathbf{r}_i, t_2, \mathbf{k}_2^a, t_1, \mathbf{k}_1^a, t_0) \tag{41}
$$

$$
p = \{f_i\} \{\lambda \exp(-\textstyle\int)\} \{\lambda^{-1} S\} \{\lambda \exp(-\textstyle\int)\} \{\lambda^{-1} S\} \{\lambda \exp(-\textstyle\int)\} \tag{42}
$$

$$
\psi = \theta_\Omega(\mathbf{K}_0(t), \mathbf{R_0}(\mathbf{t})) \tag{43}
$$

To evaluate (39) by MC integration one has to generate N realizations of the multi-dimensional variable x, which are referred to as numerical trajectories. The factors in (42) denote conditional probability densities, except f_i, which is unconditional. Therefore, one would first generate a phase space point $(\mathbf{k}_i, \mathbf{r}_i)$ from the initial distribution f_i, then choose t_2 from the free-flight time distribution, select $\mathbf{k}_2^a$ with density $\lambda^{-1}S$, and so forth. Finally, at time t the indicator function θ_Ω, which plays the role of ψ in (9), needs to be evaluated. The result will be simply one or zero. Doing so for N trajectories corresponds to counting the number of particles found in Ω at time t.

The generated times are of ascending order, $0 < t_2 < t_1 < t_0$, which means that the trajectory is followed forward in time. With a forward algorithm it is only possible to evaluate an average of the distribution function over some sub-domain, but not the exact value in a given point.

3.4. *The weighted EMC method*

In the WEMC method instead of the physical densities in (39), which follow in a natural way from the kernel, arbitrary densities are used for numerical trajectory construction. The ratio of the physical density over the numerical density determines the weight of the numerical trajectory. The basic ideas can be explained by rewriting the random variable x, the density p, and the dependent random variable $\psi(x)$ for the n-th iteration term in a formal way as

$$x = (x_0, x_1, \dots x_n) \tag{44}$$

$$p = f_0(x_0) K(x_0, x_1) \dots K(x_{n-1}, x_n) \tag{45}$$

$$\psi = \theta_\Omega \tag{46}$$

where K stands for the kernel of the conjugate equation. One can now 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.

$$p = p_0(x_0) P(x_0, x_1) \dots P(x_{n-1}, x_n) \tag{47}$$

$$\psi = \frac{f_0(x_0) K(x_0, x_1) \dots K(x_{n-1}, x_n)}{p_0(x_0) P(x_0, x_1) \dots P(x_{n-1}, x_n)} \theta_\Omega \tag{48}$$

The numerical initial distribution p_0 and the numerical transition probability P have to be non-zero where the physical counterparts are non-zero, i.e., $p_0(x_0) \neq 0$ if $f_0(x_0) \neq 0$ and $P(x_i, x_j) \neq 0$ if $K(x_i, x_j) \neq 0$. Furthermore, only normalized densities are considered, $\int p_0(x_0)\,\mathrm{d}x_0 = 1$ and $\int P(x_i, x_j)\,\mathrm{d}x_j = 1$ for all x_i.

Consequently, whenever in the process of numerical trajectory construction a random variable is selected from a numerical density rather than from a physical density, the weight of the trajectory changes by the ratio of the two densities.

As an example the WEMC method is applied to compute the energy distribution of electrons in Si. The used semiconductor model[5] assumes an analytical, non-parabolic band-structure characterized by $m_n^* = 0.32\, m_0$, $\alpha = 0.5\,\mathrm{eV}^{-1}$. To increase the probability for electrons to gain energy and thus to populate the high energy tail the probability for phonon absorption has been increased at the expense of phonon emission. Fig. 2 shows the result of a simulation of $4{\cdot}10^7$ electrons at $E = 30\,\mathrm{kV/cm}$ for $t = 1\,\mathrm{ps}$. The initial distribution is a Maxwellian at lattice temperature, chosen as $T_L = 300\,\mathrm{K}$. For the given particle number the EMC method can resolve not more than 7 decades of the energy distribution, while the WEMC method gives reasonable accurate results within 17 decades. However, it is to note that the variance of the

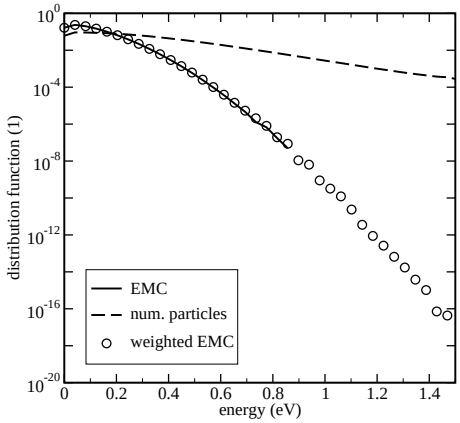

Figure 2: Electron energy distribution functions obtained by the EMC and WEMC algorithms for $E = 30\,\mathrm{kV/cm}$.

WEMC method increases with increasing evolution time. The dashed line in Fig. 2 represents the distribution of the endpoints of the numerical trajectories.

3.5. *The backward MC method*

In the previous sections forward algorithms were formally derived from the Neumann series of the conjugate equation. If the Neumann series of the integral form of the BE is approached by the MC method, backward algorithms will be obtained[8].

As an instructive example, we consider the term $f^{(2)}$ of the Neumann series of (29).

$$f^{(2)}(\mathbf{k},\mathbf{r},t) = \int_0^t \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_1 \int_0^{t_1} \mathrm{d}t_2 \int \mathrm{d}\mathbf{k}_2$$
$$f_i(\mathbf{K}_2(0),\mathbf{R}_2(0)) \exp\left(-\int_0^{t_2} \lambda(\mathbf{K}_2(y),\mathbf{R}_2(y),y)\mathrm{d}y\right)$$
$$\times S(\mathbf{k}_2,\mathbf{K}_1(t_2),\mathbf{R}_1(t_2),t_2) \exp\left(-\int_{t_2}^{t_1} \lambda(\mathbf{K}_1(y),\mathbf{R}_1(y),y)\mathrm{d}y\right)$$
$$\times S(\mathbf{k}_1,\mathbf{K}_0(t_1),\mathbf{R}_0(t_1),t_1) \exp\left(-\int_{t_1}^{t} \lambda(\mathbf{K}_0(y),\mathbf{R}_0(y),y)\mathrm{d}y\right) \quad (49)$$

Final conditions for the **k**-space trajectories are given first by $\mathbf{K}_0(t) = \mathbf{k}$ and then by the before-scattering states $\mathbf{K}_1(t_1) = \mathbf{k}_1$ and $\mathbf{K}_2(t_2) = \mathbf{k}_2$ (See Fig. 3). The real space trajectory ends at final time t in the given point $\mathbf{R}_0(t) = \mathbf{r}$ and is continuous at the time of scattering: $\mathbf{R}_1(t_1) = \mathbf{R}_0(t_1)$, $\mathbf{R}_2(t_2) = \mathbf{R}_1(t_2)$.

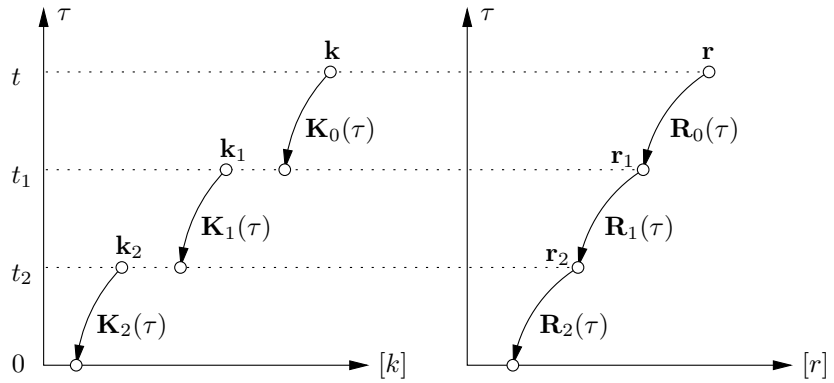

Figure 3: Sketch of a backward trajectory starting at time t and reaching time 0 after three free flights. The symbols used in (49) are shown.

As in the forward case, the integrand of (49) is augmented in two steps. The probability $\exp(-\int_0^{t_2})$ is expressed as an integral over the corresponding density.

$$\exp\Big(-\int_0^{t_2}\lambda(\mathbf{K}_2(y))\mathrm{d}y\Big)=\int_{-\infty}^{0}\lambda(\mathbf{K}_2(t_3))\exp\Big(-\int_{t_3}^{t_2}\lambda(\mathbf{K}_2(y))\mathrm{d}y\Big)\,\mathrm{d}t_3 \tag{50}$$

Second, the normalization of the distribution of the before-scattering states has to be introduced.

$$\lambda^*(\mathbf{k})=\int S(\mathbf{k}',\mathbf{k})\,\mathrm{d}\mathbf{k}' \tag{51}$$

From (51) it follows that $\int \lambda^*(\mathbf{k})^{-1}S(\mathbf{k}',\mathbf{k})\,\mathrm{d}\mathbf{k}'=1$ for all $\mathbf{k}$. Products of the form $S\exp(-\int)$ in (49) are augmented using both λ and λ^* as shown in the following expression.

$$\begin{aligned}f^{(2)}(\mathbf{k},\mathbf{r},t)=\int_0^t\mathrm{d}t_1\int_0^{t_1}\mathrm{d}t_2\int_{-\infty}^{0}\mathrm{d}t_3\int\mathrm{d}\mathbf{k}_1\int\mathrm{d}\mathbf{k}_2\\ f_i(\mathbf{K}_2(0),\mathbf{R}_2(0))\Big\{\lambda(\mathbf{K}_2(t_3))\exp\Big(-\int_{t_3}^{t_2}\lambda(\mathbf{K}_2(y))\mathrm{d}y\Big)\Big\}\\ \times\frac{\lambda^*(\mathbf{K}_1(t_2))}{\lambda(\mathbf{K}_1(t_2))}\Big\{\frac{S(\mathbf{k}_2,\mathbf{K}_1(t_2))}{\lambda^*(\mathbf{K}_1(t_2))}\Big\}\Big\{\lambda(\mathbf{K}_1(t_2))\exp\Big(-\int_{t_2}^{t_1}\lambda(\mathbf{K}_1(y))\mathrm{d}y\Big)\Big\}\\ \times\frac{\lambda^*(\mathbf{K}_0(t_1))}{\lambda(\mathbf{K}_0(t_1))}\Big\{\frac{S(\mathbf{k}_1,\mathbf{K}_0(t_1))}{\lambda^*(\mathbf{K}_0(t_1))}\Big\}\Big\{\lambda(\mathbf{K}_0(t_1))\exp\Big(-\int_{t_1}^{t}\lambda(\mathbf{K}_0(y))\mathrm{d}y\Big)\Big\}\end{aligned} \tag{52}$$

Writing the position and time-dependence of the scattering rates has been omitted for the sake of brevity. To evaluate (52) by MC integration the integrand is separated into p and ψ.

$$x = (t_1, \mathbf{k}_1, t_2, \mathbf{k}_2, t_3) \tag{53}$$

$$p = \{\lambda \exp(-\textstyle\int)\}\{S/\lambda^*\}\{\lambda \exp(-\textstyle\int)\}\{S/\lambda^*\}\{\lambda \exp(-\textstyle\int)\} \tag{54}$$

$$\psi = \frac{\lambda^*(t_1)}{\lambda(t_1)} \frac{\lambda^*(t_2)}{\lambda(t_2)} f_i \tag{55}$$

N realizations of the multi-dimensional variable x have to be generated. Since $\mathbf{k}, \mathbf{r}, t$ are given, the construction of the numerical trajectory starts at this point by choosing a random variable t_1, which obviously is less than t. The next random variable to be chosen is $\mathbf{k}_1$, a before-scattering state, and so forth. The selected times are of descending order, $t < t_1 < t_2 < t_3$, which means that the numerical trajectory is followed back in time. During trajectory construction the product $\prod \lambda^*(t_i)/\lambda(t_i)$ has to be recorded. At time zero the product is multiplied by the initial distribution f_i evaluated at the reached phase space point to give the random variable ψ. After construction of N numerical trajectories the sample mean of ψ is formed.

$$f(\mathbf{k}, \mathbf{r}, t) \approx \frac{1}{N} \sum_{j=1}^{N} \psi_j \tag{56}$$

The MCB method allows to evaluate the distribution function at given phase space points with desired accuracy.

In analogy with the forward case event biasing can be applied leading to weighted backward algorithms.

Fig. 4 compares the energy distributions of electrons in Si as computed by the MCB and the WEMC methods. Conditions assumed are $E = 10\,\mathrm{kV/cm}$ and $t = 3\,\mathrm{ps}$. The initial distribution and the number of particles for the WEMC simulation are as in Fig. 2. The MCB method is used to evaluate the energy distribution at discrete points above 800 meV. The statistical uncertainty of the result is controlled by the number of numerical trajectories starting from each point. In the simulation 10^7 backward trajectories are computed for each point. Using the MCB method the high energy tail is obtained with high precision as shown in Fig 4. The depicted range of 30 decades is out of reach even for the here considered variant of WEMC method.

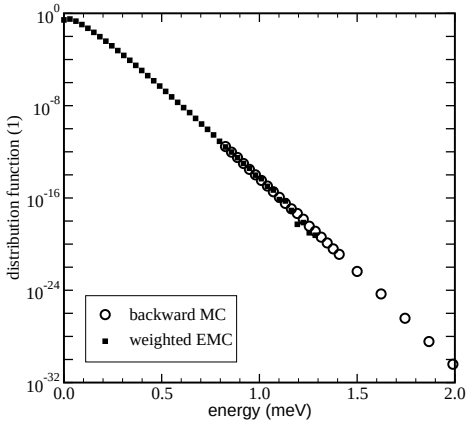

Figure 4: Electron energy distribution functions obtained by the WEMC amd MCB algorithms for $E = 10\,\mathrm{kV/cm}$.

4. The Stationary Boltzmann Equation

Aiming at steady-state device simulation, the position-dependent and time-invariant BE is to be considered. The force field and all material properties are independent of time.

$$[\mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} + \mathbf{F}(\mathbf{r}) \cdot \nabla_{\mathbf{k}}]\, f(\mathbf{k}, \mathbf{r}) = Q[f](\mathbf{k}, \mathbf{r}), \qquad \mathbf{r} \in D \tag{57}$$

This equation, which is posed in the simulation domain D, is supplemented by boundary conditions modeling the interaction of the device with the environment. The distribution function is normalized as (see also Section 3)

$$\frac{1}{4\pi^3} \int\limits_D \mathrm{d}\mathbf{r} \int \mathrm{d}\mathbf{k}\, f(\mathbf{k}, \mathbf{r}) = N_D\,. \tag{58}$$

In the scattering operator $Q = Q_g - Q_l$ the scattering rate S is independent of time:

$$Q_g[f](\mathbf{k}, \mathbf{r}) = \int f(\mathbf{k}', \mathbf{r}) S(\mathbf{k}', \mathbf{k}, \mathbf{r})\, \mathrm{d}\mathbf{k}'\,, \tag{59}$$

$$Q_l[f](\mathbf{k}, \mathbf{r}) = \lambda(\mathbf{k}, \mathbf{r}) f(\mathbf{k}, \mathbf{r})\,, \qquad \text{with} \qquad \lambda(\mathbf{k}, \mathbf{r}) = \int S(\mathbf{k}, \mathbf{k}', \mathbf{r})\, \mathrm{d}\mathbf{k}' \tag{60}$$

To describe a time-invariant system an absolute time scale is not needed. Only the time difference between two consecutive events is significant. A phase space trajectory with the initial condition $\mathbf{K}(t_0) = \mathbf{k}_0$ and $\mathbf{R}(t_0) = \mathbf{r}_0$ is obtained by

formal integration of the equations of motion.

$$\mathbf{K}(t;t_0,\mathbf{k}_0,\mathbf{r}_0) = \mathbf{k}_0 + \int_{t_0}^{t} \mathbf{F}\left(\mathbf{R}(y;t_0,\mathbf{k}_0,\mathbf{r}_0)\right) \mathrm{d}y$$

$$\mathbf{R}(t;t_0,\mathbf{k}_0,\mathbf{r}_0) = \mathbf{r}_0 + \int_{t_0}^{t} \mathbf{v}\left(\mathbf{K}(y;t_0,\mathbf{k}_0,\mathbf{r}_0)\right) \mathrm{d}y \tag{61}$$

In addition to the time argument of the functions $\mathbf{K}$ and $\mathbf{R}$, the parameters t_0, $\mathbf{k}_0$, $\mathbf{r}_0$ describing the initial condition of the phase space trajectory are stated explicitly. Expressions (61) can be read as the phase space position of a particle at time t that passes through $\mathbf{k}_0$ and $\mathbf{r}_0$ at time t_0. In this regard, the order of t_0 and t is irrelevant. For $t \leq t_0$ the meaning of $\mathbf{k}_0$ and $\mathbf{r}_0$ would be that of a final condition.

Invariance under time translation can be proven, provided that $\mathbf{F}$ does not depend explicitly on time.

$$\mathbf{K}(t+\tau;t_0+\tau,\mathbf{k}_0,\mathbf{r}_0) = \mathbf{K}(t;t_0,\mathbf{k}_0,\mathbf{r}_0) \tag{62}$$

$$\mathbf{R}(t+\tau;t_0+\tau,\mathbf{k}_0,\mathbf{r}_0) = \mathbf{R}(t;t_0,\mathbf{k}_0,\mathbf{r}_0) \tag{63}$$

This property will be used repeatedly in the following to adjust conveniently the time reference for each free flight.

4.1. *Stationary MC algorithms*

The bulk EMC algorithm can also be applied under stationary conditions. In this situation, for large evolution times the final distribution approaches a steady state, and the information introduced by the initial condition is lost entirely. Alternatively, the ergodicity of the process can be exploited to replace the ensemble average by a time average. Since it is sufficient to simulate one test particle for a long period of time, the algorithm is called *Single-Particle MC* (SPMC). The effect of the particle's initial state vanishes for long simulation times. The distribution function in a given phase space point is estimated by the time spent by the particle in a fixed, small volume around the point divided by the total time the trajectory was followed. Another method of obtaining steady-state averages has been introduced by Price[17]. With the *synchronous-ensemble* or *before-scattering* method averages are formed by sampling the trajectory at the end of each free flight which is in many cases easier a task than evaluating a path integral over each free flight.

For a space-dependent problem the SPMC algorithm has to take into account boundary conditions. Whenever in a simulation the particle leaves the simulation domain through a contact it is re-injected through one of the contacts, selected according to the probabilities of the underlying model. Proofs that the SPMC algorithm yields a distribution function which satisfies the stationary BE has been given by Fawcett et al. for homogeneous case[18], and by Baccarani et al. for the

inhomogeneous case[19]. For the SPMC algorithm, convergence proof for the iteration series and variance estimates have been reported recently[20,21].

4.2. *Integral form of the stationary Boltzmann equation*

In this section the BE is transformed from integro-differential form into integral form. Particular care is taken to account for the boundary condition.

Assume a given phase space point $\mathbf{k}$, $\mathbf{r}$. This point determines uniquely a phase space trajectory, for which the short cut notation $\mathbf{K}(t) = \mathbf{K}(t; 0, \mathbf{k}, \mathbf{r})$ and $\mathbf{R}(t) = \mathbf{R}(t; 0, \mathbf{k}, \mathbf{r})$ is used. The arbitrary initial time is set to $t_0 = 0$. The left-hand side of (57) represents the total time derivative of $\hat{f}(t) = f(\mathbf{K}(t), \mathbf{R}(t))$, which allows the BE to be rewritten as an ordinary differential equation of first order as shown in Section 3.2:

$$\frac{\mathrm{d}}{\mathrm{d}t} \exp\left(\int_0^t \hat{\lambda}(y)\mathrm{d}y \right) \hat{f}(t) = \exp\left(\int_0^t \hat{\lambda}(y)\mathrm{d}y \right) \hat{Q}_g[f](t) \tag{64}$$

This equation can be integrated straight forwardly. The upper bound of integration should be $t = 0$ to obtain $\hat{f}(0) = f(\mathbf{k}, \mathbf{r})$, the value of f at the given phase space point. The lower time bound has to be chosen such that the functions $\mathbf{K}(t)$ and $\mathbf{R}(t)$ take on values at which the distribution function is known. In the steady state the distribution function is known only at the domain boundary. An appropriate lower time bound is therefore the time, say t_b^-, at which the trajectory enters the simulation domain (see Fig. 5). Apparently, this time depends on the point $\mathbf{k}$, $\mathbf{r}$ under consideration.

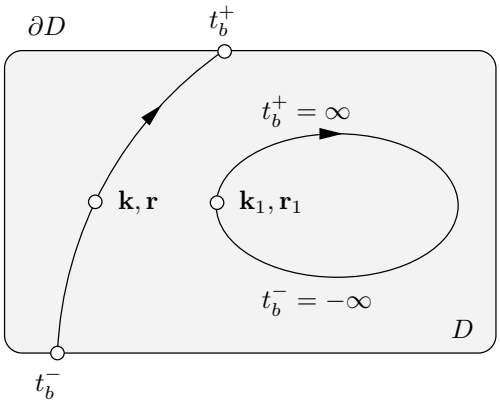

Figure 5: Illustration of the functions $t_b^-(\mathbf{k}, \mathbf{r})$ and $t_b^+(\mathbf{k}, \mathbf{r})$ which give the time at a trajectory's entry point and exit point, respectively. If $\mathbf{k}_1, \mathbf{r}_1$ is the initial point of a closed trajectory, the times are infinite, $t_b^\pm(\mathbf{k}_1, \mathbf{r}_1) = \pm\infty$.

If the real space trajectory $\mathbf{R}(t; 0, \mathbf{k}, \mathbf{r})$ never intersects the domain boundary, that is when the trajectory forms a closed loop, then $t_b^- = -\infty$ is an appropriate

choice. This means that a particle found at $t = 0$ on such a closed trajectory must have been scattered onto this trajectory at some time $t \in (-\infty, 0)$ in the past.

Integration of (64) in the time bounds discussed above results in the integral form of the stationary BE,

$$f(\mathbf{k},\mathbf{r}) = \int\limits_{t_b^-(\mathbf{k},\mathbf{r})}^{0} \mathrm{d}t' \int \mathrm{d}\mathbf{k}'\; f(\mathbf{k}',\mathbf{R}(t'))\; S(\mathbf{k}',\mathbf{K}(t'),\mathbf{R}(t')) \times \exp\!\left(-\int\limits_{t'}^{0} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y\right) + f_0(\mathbf{k},\mathbf{r}) \quad (65)$$

$$f_0(\mathbf{k},\mathbf{r}) = f_b\left(\mathbf{K}(t_b^-(\mathbf{k},\mathbf{r})),\mathbf{R}(t_b^-(\mathbf{k},\mathbf{r}))\right) \exp\!\left(-\int\limits_{t_b^-(\mathbf{k},\mathbf{r})}^{0} \lambda(\mathbf{K}(y),\mathbf{R}(y))\,\mathrm{d}y\right) \quad (66)$$

where f_b denotes the boundary distribution. The integral form is a bookkeeping equation for the probability $f(\mathbf{k},\mathbf{r})\mathrm{d}\mathbf{k}\,\mathrm{d}\mathbf{r}$ of finding a carrier in the volume element $\mathrm{d}\mathbf{k}\mathrm{d}\mathbf{r}$ about $\mathbf{k}$ and $\mathbf{r}$. The first summand in (65) describes the contribution of carriers that are scattered onto the considered trajectory at some time $t' \in (t_b^-, 0)$ and stay on it until time 0, whereas the second summand gives the contribution of carriers that stay from the time of entry t_b^- on the trajectory and have a collisionless free flight until time 0, reaching the point of interest $\mathbf{k}$, $\mathbf{r}$.

To complete the set of basic equations for the stationary transport problem the conjugate equation has to be found. Using the notation of Section 2.3, the conjugate equation has the same kernel as the integral equation, but integration is carried out over the unprimed variables. To apply this rule, the integral form of the BE first has to be transformed to standard form (13):

$$f(\mathbf{k},\mathbf{r}) = \int \mathrm{d}\mathbf{k}' \int \mathrm{d}\mathbf{r}'\; f(\mathbf{k}',\mathbf{r}')K(\mathbf{k}',\mathbf{r}',\mathbf{k},\mathbf{r}) + f_0(\mathbf{k},\mathbf{r}) \quad (67)$$

The required $\mathbf{r}'$ integration is introduced by augmenting the kernel by a δ-function

$$K(\mathbf{k}',\mathbf{r}',\mathbf{k},\mathbf{r}) = \int\limits_{t_b^-(\mathbf{k},\mathbf{r})}^{0} \mathrm{d}t'\;\; S(\mathbf{k}',\mathbf{K}(t'),\mathbf{r}')\;\exp\!\left(-\int\limits_{t'}^{0} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y\right)\delta(\mathbf{r}'-\mathbf{R}(t'))\,\theta_D(\mathbf{r}')\,, \quad (68)$$

where θ_D is the indicator function of the simulation domain. Since the integral equation (67) is posed in the six-dimensional phase-space it must contain a six-fold integral. The time integral in (65) therefore cannot stay in (67) and has to be assigned consequently to the kernel. This means that the kernel of the stationary BE is given by the kernel of the transient BE integrated over time.

After changing variables and reversing time the conjugate equation can be stated explicitly[20]:

$$g(\mathbf{k}',\mathbf{r}') = \int \mathrm{d}\mathbf{k}_a \int_0^{t_b^+(\mathbf{k}_a,\mathbf{r}')} \mathrm{d}\tau\, S(\mathbf{k}',\mathbf{k}_a,\mathbf{r}')\, \exp\!\left(-\int_0^{\tau} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y\right) \times g(\mathbf{K}(\tau),\mathbf{R}(\tau))\theta_D(\mathbf{r}') + A(\mathbf{k}',\mathbf{r}') \quad (69)$$

This equation has now the desired properties that integration is carried out over final states and that the time variable is positive. The iteration series of (69) will lead to forward MC algorithms.

4.3. *The Single-Particle MC method*

Assume we are interested in the mean value of some quantity $A(\mathbf{k},\mathbf{r})$.

$$\langle\langle \mathrm{A} \rangle\rangle = \int_D \mathrm{d}\mathbf{r} \int \mathrm{d}\mathbf{k}\; A(\mathbf{k},\mathbf{r}) f(\mathbf{k},\mathbf{r}) \quad (70)$$

A will typically be a product of some $\mathbf{k}$-dependent function and an $\mathbf{r}$-dependent charge assignment function[3]. The mean value per particle is obtained as $\langle \mathrm{A} \rangle = \langle\langle \mathrm{A} \rangle\rangle / \langle\langle 1 \rangle\rangle$, where the normalization constant evaluates to $\langle\langle 1 \rangle\rangle = 4\pi^3 N_D$ according to (20).

Equation (70) denotes an inner product (A, f), which can be transformed into (f_0, g) by means of (18):

$$\langle\langle \mathrm{A} \rangle\rangle = \int_D \mathrm{d}\mathbf{r}' \int \mathrm{d}\mathbf{k}'\; f_b(\mathbf{K}_b(t_b^-),\mathbf{R}_b(t_b^-)) \exp\!\left(-\int_{t_b^-}^{0} \lambda(\mathbf{K}_b(y),\mathbf{R}_b(y))\mathrm{d}y\right) g(\mathbf{k}',\mathbf{r}') \quad (71)$$

Here, t_b^- is a short cut for $t_b^-(\mathbf{k}',\mathbf{r}')$ and $\mathbf{K}_b$ and $\mathbf{R}_b$ is a phase space trajectory that passes through $\mathbf{k}'$ and $\mathbf{r}'$ at $t = 0$.

In (71) variables needs to be changed such that the arguments of f_b become integration variables. The new variables, $\mathbf{k}_b$ and $\mathbf{r}_b$, represent the initial state of a particle injected at the domain boundary. Since f_b is defined only at the boundary ∂D the transformation must lead from a volume to a boundary integral.

In deriving the transformation first the domains of the involved variables have to be analyzed. The integration domain Φ is the direct product of D and K, the $\mathbf{k}$-space. The following decomposition of Φ is considered.

$$\Phi = D \otimes K = \Phi_b \cup \overline{\Phi}_b \quad (72)$$

The sub-domain Φ_b is formed by all points for which t_b^- is finite. Each point $\mathbf{k}'$, $\mathbf{r}' \in \Phi_b$ is connected with a boundary point $\mathbf{k}_b$, $\mathbf{r}_b$ by a free-flight trajectory. The complementary sub-domain $\overline{\Phi}_b$ contains all points for which $t_b^- = -\infty$, that are

those points that lie on closed trajectories. The integrand of (71) vanishes for all points in $\overline{\Phi}_b$ because of the e-function, and so does integral over $\overline{\Phi}_b$. It is therefore sufficient to restrict the integration domain to Φ_b.

Another decomposition needed in the following is that of the **k**-space at a boundary point. If $\mathbf{n}(\mathbf{r}_b)$ denotes the outward directed normal vector in a point $\mathbf{r}_b$ at the domain boundary, the two subspaces are defined by

$$K_+(\mathbf{r}_b) = \{\mathbf{k} : \ \mathbf{v}(\mathbf{k}) \cdot \mathbf{n}(\mathbf{r}_b) < 0\} \tag{73}$$

$$K_-(\mathbf{r}_b) = \{\mathbf{k} : \ \mathbf{v}(\mathbf{k}) \cdot \mathbf{n}(\mathbf{r}_b) \geq 0\} \tag{74}$$

All **k**-points in $K_+(\mathbf{r}_b)$ have an inward directed component of the group velocity and are therfore initial points of trajectories entering the domain at $\mathbf{r}_b$. Conversely, points in K_- are endpoints of trajectories leaving the domain.

Each point $(\mathbf{k}', \mathbf{r}') \in \Phi_b$ can now be mapped one-to-one onto a boundary point $(\mathbf{k}_b,\ \mathbf{r}_b)$ and a positive time t_0, where $\mathbf{k}_b \in K_+(\mathbf{r}_b)$ and $\mathbf{r}_b \in \partial D$. The time $t_0 = -t_b^-(\mathbf{k}', \mathbf{r}')$ it takes a particle to drift from the boundary point $(\mathbf{k}_b,\ \mathbf{r}_b)$ to the inner point $(\mathbf{k}',\ \mathbf{r}')$. The volume element transforms as[20]

$$\mathrm{d}\mathbf{r}'\,\mathrm{d}\mathbf{k}' = |v_\perp(\mathbf{k}_b)|\mathrm{d}\sigma(\mathbf{r}_b)\,\mathrm{d}\mathbf{k}_b\,\mathrm{d}t_0\,, \tag{75}$$

where $\mathrm{d}\sigma(\mathbf{r}_b)$ is the surface element at $\mathbf{r}_b$, and (71) gets transformed into

$$\langle\!\langle \mathrm{A} \rangle\!\rangle = \oint\limits_{\partial D} \mathrm{d}\sigma(\mathbf{r}_b) \int\limits_{K_+(\mathbf{r}_b)} \mathrm{d}\mathbf{k}_b \int\limits_0^{t_b^+(\mathbf{k}_b,\mathbf{r}_b)} \mathrm{d}t_0 \ \ |v_\perp(\mathbf{k}_b)| f_b(\mathbf{k}_b, \mathbf{r}_b)$$

$$\times \exp\Big(- \int\limits_0^{t_0} \lambda(\mathbf{K}_b(y), \mathbf{R}_b(y))\mathrm{d}y \Big) g(\mathbf{K}_b(t_0), \mathbf{R}_b(t_0))\,. \tag{76}$$

The accomplished change from volume to boundary integration is a key step in the treatment of the boundary value problem. It proves that knowledge of the boundary distribution is sufficient to determine arbitrary volume integrals defined by (70) and therefore to determine uniquely f.

Note that from f_b only the part in K_+ determines the boundary condition, whereas the part in K_- is unknown and is a result of the simulation.

Required for the purpose of normalization are the integrals

$$j_\perp(\mathbf{r}) = \int\limits_{K_+(\mathbf{r})} \mathrm{d}\mathbf{k}\,|v_\perp(\mathbf{k})| f_b(\mathbf{k}, \mathbf{r}), \qquad \mathbf{r} \in \partial D \tag{77}$$

$$\Gamma_D = \oint\limits_{\partial D} j_\perp(\mathbf{r})\mathrm{d}\sigma(\mathbf{r})\,. \tag{78}$$

Taking into account the normalization given in (20), $j_\perp/(4\pi^3)$ represents the normal component of the incident particle current density and $\Gamma_D/(4\pi^3)$ the total incident particle current.

Substituting the Neumann series of the conjugate equation, $g = \sum_0^\infty g^{(i)}$, into (76) results in a series for the mean value, for which the following notation is adopted.

$$\langle\!\langle \mathrm{A} \rangle\!\rangle = \sum_{i=0}^{\infty} \langle\!\langle \mathrm{A} \rangle\!\rangle_i \tag{79}$$

As an instructive example the term of second order of is stated explicitly.

$$\begin{aligned}
\langle\!\langle \mathrm{A} \rangle\!\rangle_2 = \oint_{\partial D} \mathrm{d}\sigma(\mathbf{r}_b) \int_{K_+(\mathbf{r}_b)} \mathrm{d}\mathbf{k}_b \int_0^{t_b^+(\mathbf{k}_b,\mathbf{r}_b)} \mathrm{d}t_0 \int \mathrm{d}\mathbf{k}_1 \int_0^{t_b^+(\mathbf{k}_1,\mathbf{r}_1)} \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_2 \int_0^{t_b^+(\mathbf{k}_2,\mathbf{r}_2)} \mathrm{d}t_2 \;\; |v_\perp(\mathbf{k}_b)| f_b(\mathbf{k}_b, \mathbf{r}_b) \\
\times \exp\!\left(- \int_0^{t_0} \lambda(\mathbf{K}_b(y), \mathbf{R}_b(y)) \mathrm{d}y \right) S(\mathbf{K}_b(t_0), \mathbf{k}_1, \mathbf{R}_b(t_0)) \\
\times \exp\!\left(- \int_0^{t_1} \lambda(\mathbf{K}_1(y), \mathbf{R}_1(y)) \mathrm{d}y \right) S(\mathbf{K}_1(t_1), \mathbf{k}_2, \mathbf{R}_1(t_1)) \\
\times \exp\!\left(- \int_0^{t_2} \lambda(\mathbf{K}_2(y), \mathbf{R}_2(y)) \mathrm{d}y \right) A(\mathbf{K}_2(t_2), \mathbf{R}_2(t_2))
\end{aligned} \tag{80}$$

Initial conditions for the **k**-space trajectories are given by $\mathbf{k}_b$ and the after-scattering states $\mathbf{k}_i$, respectively, as shown in Fig.6.

$$\mathbf{K}_b(0) = \mathbf{k}_b \,, \qquad \mathbf{K}_i(0) = \mathbf{k}_i, \qquad i = 1, 2, \ldots \tag{81}$$

The real space trajectory is continuous at the time of scattering. It holds $\mathbf{R}_b(0) = \mathbf{r}_b$ and $\mathbf{R}_i(t_i) = \mathbf{R}_{i+1}(0)$.

The iteration term (80) describes the contribution of all particles that propagate from the boundary to the interior of the device, having undergone two scattering events and finished the third free flight. Analogously, the i-th iteration term, $\langle\!\langle \mathrm{A} \rangle\!\rangle_i$, represents the contribution of all particles which propagate into the device with i scattering events and $i+1$ free flights.

Furthermore, the symbols $\mathbf{k}_i^b$ and $\mathbf{r}_i$ are introduced, which denote the before-scattering momentum and the particle position for the i-th scattering event, respectively. They are related to the trajectories by

$$\mathbf{r}_{i+1} = \mathbf{R}_i(t_i) \,, \qquad \mathbf{k}_{i+1}^b = \mathbf{K}_i(t_i) \,. \tag{82}$$

As a next step the integrand of the iteration term (80) needs to be decomposed into a probability density p and a random variable ψ, as shown in (9). For this purpose we repeat the well-known probability densities used in MC device simulation, which are the distribution of the free-flight time, p_t, and that of the state after

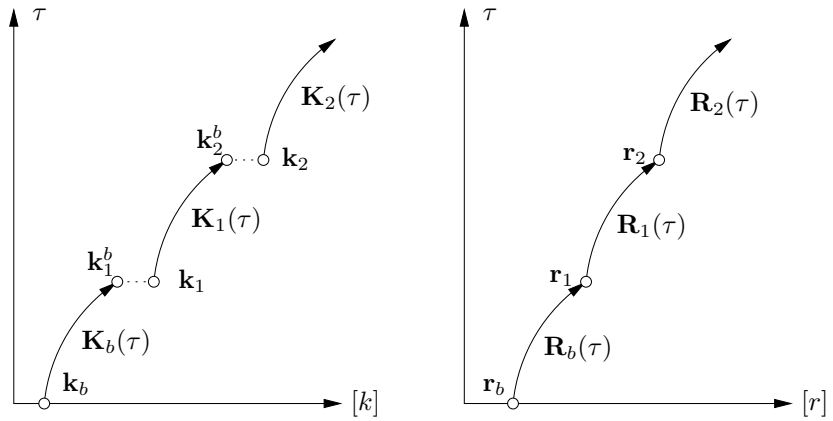

Figure 6: Sketch of a trajectory which starts at the boundary point $(\mathbf{k}_b, \mathbf{r}_b)$ and evolves until the third free flight. The symbols used in (80) are shown.

scattering, p_k.

$$p_t(t;\mathbf{k},\mathbf{r}) = \lambda(\mathbf{K}(t),\mathbf{R}(t)) \exp\left(-\int_0^t \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y\right) \tag{83}$$

$$p_k(\mathbf{k}';\mathbf{k},\mathbf{r}) = \frac{S(\mathbf{k},\mathbf{k}',\mathbf{r})}{\lambda(\mathbf{k},\mathbf{r})} \tag{84}$$

Both distributions are normalized as they satisfy for all $\mathbf{k}, \mathbf{r}$:

$$\int_0^\infty p_t(t;\mathbf{k},\mathbf{r})\,\mathrm{d}t = 1\,, \qquad \int p_{\mathbf{k}}(\mathbf{k}';\mathbf{k},\mathbf{r})\,\mathrm{d}\mathbf{k}' = 1\,. \tag{85}$$

In the integrand of (80) terms representing an unnormalized probability density are divided by the respective normalization factors. Beginning with the left-most term, the velocity-weighted boundary distribution $v_\perp f_b$, these factors are given by (78) and (77). Products of the form $\exp(-\int\lambda)\,S$ are multiplied by λ/λ in order to obtain the free-flight time distribution of the form $p_t = \lambda\exp(-\int\lambda)$ and the distribution of the after-scattering states, $p_k = S/\lambda$. The remaining product of the form $\exp(-\int\lambda)\,A$ can be treated in two different ways, leading to either the *synchronous ensemble* method or the *time-integration* method of average recording.

4.3.1. *The synchronous ensemble method*

One option is to multiply the remaining product $\exp(-\int\lambda)\,A$ by λ/λ to obtain a product of p_t and A/λ. For the sake of brevity, the position-dependence of the

scattering rate is suppressed in the following. Equation eqnA2-1 becomes:

$$\langle\!\langle \mathrm{A} \rangle\!\rangle_2 = \Gamma_D \oint\limits_{\partial D} \mathrm{d}\sigma \int\limits_{K_+} \mathrm{d}\mathbf{k}_b \int\limits_0^{t_{b0}^+} \mathrm{d}t_0 \int \mathrm{d}\mathbf{k}_1 \int\limits_0^{t_{b1}^+} \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_2 \int\limits_0^{t_{b2}^+} \mathrm{d}t_2$$
$$\left\{ \frac{j_\perp(\mathbf{r}_b)}{\Gamma_D} \right\} \left\{ \frac{|v_\perp(\mathbf{k}_b)| f_b(\mathbf{k}_b, \mathbf{r}_b)}{j_\perp(\mathbf{r}_b)} \right\}$$
$$\times \left\{ \lambda(\mathbf{K}_b(t_0)) \exp\left(- \int\limits_0^{t_0} \lambda(\mathbf{K}_b(y)) \mathrm{d}y \right) \right\} \left\{ \frac{S(\mathbf{K}_b(t_0), \mathbf{k}_1)}{\lambda(\mathbf{K}_b(t_0))} \right\}$$
$$\times \left\{ \lambda(\mathbf{K}_1(t_1)) \exp\left(- \int\limits_0^{t_1} \lambda(\mathbf{K}_1(y)) \mathrm{d}y \right) \right\} \left\{ \frac{S(\mathbf{K}_1(t_1), \mathbf{k}_2)}{\lambda(\mathbf{K}_1(t_1))} \right\}$$
$$\times \left\{ \lambda(\mathbf{K}_2(t_2)) \exp\left(- \int\limits_0^{t_2} \lambda(\mathbf{K}_2(y)) \mathrm{d}y \right) \right\} \frac{A(\mathbf{K}_2(t_2), \mathbf{R}_2(t_2))}{\lambda(\mathbf{K}_2(t_2))} \tag{86}$$

Each term representing a probability density is enclosed in curly brackets. In (86) we designate the multi-dimensional integration variable as x_2, the probability density as p_2, and a random variable as ψ_2: $\langle\!\langle \mathrm{A} \rangle\!\rangle_2 = \int \mathrm{d}x_2 \; p_2(x_2)\psi_2(x_2)$, where

$$x_2 = (\mathbf{r}_b, \mathbf{k}_b, t_0, \mathbf{k}_1, t_1, \mathbf{k}_2, t_2) \tag{87}$$
$$p_2(x_2) = \{j_\perp/\Gamma_D\}\{v_\perp f_b/j_\perp\}\{p_t\}\{p_k\}\{p_t\}\{p_k\}\{p_t\} \tag{88}$$
$$\psi_2(x_2) = A/\lambda \tag{89}$$

To evaluate (86) by MC integration one has to generate a sample $x_{2,1} \ldots x_{2,N}$ from the density p_2. A realization $x_{2,j}$ is referred to as a numerical trajectory, its generation as numerical trajectory construction.

Consider the construction of the j-th numerical trajectory, $x_{2,j}$. Since all factors in (88) except $j_\perp/\Gamma_D$ denote conditional probability densities, one first selects a boundary point $\mathbf{r}_{b,j}$ with the density $j_\perp/\Gamma_D$. Then generate $\mathbf{k}_{b,j}$ from the velocity-weighted boundary distribution, generate $t_{0,j}$ from the free-flight time distribution p_t, select $\mathbf{k}_{1,j}$ with density p_k, and so forth. Finally, at the end of the third free flight evaluate A/λ. After construction of N numerical trajectories the following sample mean is formed.

$$\langle\!\langle \mathrm{A} \rangle\!\rangle_2 \simeq \frac{\Gamma_D}{N} \sum_{j=1}^{N} \psi_2(x_{2,j}) \tag{90}$$

The described procedure contains all basic steps known from SPMC algorithm:

- generation of an initial state from the velocity-weighted boundary distribution[22], $v_\perp f_b$
- free-flight time generation from density p_t

- selection of the after scattering state with density p_k
- the synchronous ensemble method of average recording[17]

Note that in (86) the bounds of time integration are $(0, t_{bj}^+)$, where t_{bj}^+ can be either finite or infinite. On the other hand, the distribution of the free-flight time (83) is normalized in the bounds $(0, \infty)$. The issue of normalization is related to trajectories that terminate at the domain boundary and is discussed in more detail in Section 4.3.3.

4.3.2. *The time averaging method*

A second option is to process the t_2-integral in (80) by integration by parts.

$$\int_0^\infty \exp\Big(-\int_0^{t_2} \lambda(\mathbf{K}_2(y), \mathbf{R}_2(y))\mathrm{d}y\Big) H(t_{b2}^+ - t_2) A(\mathbf{K}_2(t_2), \mathbf{R}_2(t_2))\mathrm{d}t_2 =$$

$$\int_0^\infty \mathrm{d}t_2 \; \lambda(\mathbf{K}_2(t_2)) \exp\Big(-\int_0^{t_2} \lambda(\mathbf{K}_2(y))\mathrm{d}y\Big) \int_0^{t_2} H(t_{b2}^+ - \tau) A(\mathbf{K}_2(\tau), \mathbf{R}_2(\tau)\,\mathrm{d}\tau\,, \quad (91)$$

where H stands for the unit step function. On the left-hand side, $\exp(-\int)$ represents the probability that the particle drifts without scattering from 0 to t_2. Differentiating this probability gives the probability density p_t appearing on the right side. In this way the density p_2 defined by (88) is recovered, and the iteration term can be reformulated as

$$\langle\langle \mathrm{A} \rangle\rangle_2 = \Gamma_D \oint_{\partial D} \mathrm{d}\sigma \int_{K_+} \mathrm{d}\mathbf{k}_b \int_0^{t_{b0}^+} \mathrm{d}t_0 \int \mathrm{d}\mathbf{k}_1 \int_0^{t_{b1}^+} \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_2 \int_0^\infty \mathrm{d}t_2$$

$$p_2(\mathbf{r}_b, \mathbf{k}_b, t_0, \mathbf{k}_1, t_1, \mathbf{k}_2, t_2) \int_0^{\tau_2} A(\mathbf{K}_2(\tau), \mathbf{R}_2(\tau)\,\mathrm{d}\tau\,, \quad (92)$$

with $\tau_2 = \min(t_{b2}^+, t_2)$. The random variable ψ_2 in this expression is identified as the path integral over τ. As opposed to (86) the integration domain of t_2 is now $(0, \infty)$, which means that the random variable is nonzero regardless of the selected value for t_2. If $t_2 < t_{b2}^+$, the τ-integration is performed until the next scattering event occurs, otherwise until the boundary is reached.

The so obtained the time averaging method proves that the stochastic process under consideration is ergodic.

4.3.3. *MC evaluation of the iteration series*

One peculiarity of the SPMC method is that a realization of the iteration term of order i is not generated independently from that for the term of order $i - 1$.

Instead, a realization of x_i is generated by adding to the realization of x_{i-1} another after-scattering state $\mathbf{k}_i$ and another free-flight time t_i. The aim of this section is to find that random variable whose realizations are independent from each other.

Let us begin the analysis with the iteration term of order zero. Setting $g^{(0)} = A$ it follows from (76)

$$\langle\!\langle \mathrm{A} \rangle\!\rangle_0 = \Gamma_D \oint\limits_{\partial D} \mathrm{d}\sigma \int\limits_{K_+} \mathrm{d}\mathbf{k}_b \int\limits_0^{t_{b0}^+} \mathrm{d}t_0\, p_0(\mathbf{r}_b, \mathbf{k}_b, t_0) \frac{A(\mathbf{K}_b(t_0), \mathbf{R}_b(t_0))}{\lambda(\mathbf{K}_b(t_0))}\,, \tag{93}$$

with

$$p_0(\mathbf{r}_b, \mathbf{k}_b, t_0) = \left\{\frac{j_\perp(\mathbf{r}_b)}{\Gamma_D}\right\}\left\{\frac{|v_\perp(\mathbf{k}_b)| f_b(\mathbf{k}_b, \mathbf{r}_b)}{j_\perp(\mathbf{r}_b)}\right\}\left\{\lambda(\mathbf{K}_b(t_0)) \exp\left(-\int\limits_0^{t_0} \lambda(\mathbf{K}_b(y))\mathrm{d}y\right)\right\}. \tag{94}$$

The idea is now to add a scattering term and another free-flight term to the density p_0. The formal procedure is to multiply (93) by

$$\int \mathrm{d}\mathbf{k}_1 \left\{\frac{S(\mathbf{K}_b(t_0), \mathbf{k}_1)}{\lambda(\mathbf{K}_b(t_0))}\right\} \int\limits_0^\infty \mathrm{d}t_1 \left\{\lambda(\mathbf{K}_1(t_1)) \exp\left(-\int\limits_0^{t_1} \lambda(\mathbf{K}_1(y))\mathrm{d}y\right)\right\} = 1\,, \tag{95}$$

the product of two normalization integrals. In this way p_0 is multiplied by two factors such that the product gives p_1. This allows the partial sum of the first two iteration terms to be rewritten as one multiple integral.

$$\begin{aligned}\langle\!\langle \mathrm{A} \rangle\!\rangle_0 + \langle\!\langle \mathrm{A} \rangle\!\rangle_1 = \Gamma_D \oint\limits_{\partial D} \mathrm{d}\sigma \int\limits_{K_+} \mathrm{d}\mathbf{k}_b \int\limits_0^\infty \mathrm{d}t_0 \int \mathrm{d}\mathbf{k}_1 \int\limits_0^\infty \mathrm{d}t_1\, p_1(\mathbf{r}_b, \mathbf{k}_b, t_0, \mathbf{k}_1, t_1) \\ \times H(t_{b0}^+ - t_0)\left(\frac{A(\mathbf{k}_1^b, \mathbf{r}_1)}{\lambda(\mathbf{k}_1^b)} + H(t_{b1}^+ - t_1)\frac{A(\mathbf{k}_2^b, \mathbf{r}_2)}{\lambda(\mathbf{k}_2^b)}\right)\end{aligned} \tag{96}$$

Here the electron momentum before scattering, $\mathbf{k}_i^b$, is defined by (82). Since the integration domain of, for instance, t_1, is different in (86) and (96), time integration is generally carried out in $(0, \infty)$, while the integrand is set to zero above the actual time bound using the unit step function. Multiplying (96) by an integral similar to (95) and adding $\langle\!\langle \mathrm{A} \rangle\!\rangle_2$ gives the partial sum of the first three iteration terms expressed as one multiple integral. This procedure can be repeated to express the partial sum of any order n as one multiple integral

$$\langle\!\langle \mathrm{A} \rangle\!\rangle_0 + \langle\!\langle \mathrm{A} \rangle\!\rangle_1 + \ldots + \langle\!\langle \mathrm{A} \rangle\!\rangle_n = \int p_n(x_n)\psi^{[n]}(x_n)\mathrm{d}x_n\,, \tag{97}$$

using the recursive definitions

$$x_n = (x_{n-1}, \mathbf{k}_n, t_n) \tag{98}$$

$$p_n = p_{n-1} \times \left\{ \frac{S(\mathbf{k}_n^b, \mathbf{k}_n)}{\lambda(\mathbf{k}_n^b)} \right\} \left\{ \lambda(\mathbf{K}_n(t_n)) \exp\left(- \int_0^{t_n} \lambda(\mathbf{K}_n(y)) \mathrm{d}y \right) \right\} \tag{99}$$

$$\psi^{[n]} = \psi^{[n-1]} + \prod_{j=0}^{n} H(t_{bj}^+ - t_j) \frac{A(\mathbf{k}_{n+1}^b, \mathbf{r}_{n+1})}{\lambda(\mathbf{k}_{n+1}^b)} \tag{100}$$

At the beginning of the recursions are $x_0 = (\mathbf{k}_b, \mathbf{r}_b, t_0)$, p_0 given by (94), and $\psi^{[-1]} = 0$. Clearly, (98) and (99) are generalizations for arbitrary n of (87) and (88), respectively.

At the moment it is assumed that the series given by (97) is convergent, which means that there exists always some n such that the series of the truncated elements is below a desired limit. Assume n given. In general, to evaluate (97) by MC integration one has to generate N realizations of the random variable x_n. The initial state at the boundary, $\mathbf{r}_b$, $\mathbf{k}_b$, and the first free-flight time, t_0, have to be generated from p_0. If t_0 is less than t_{b0}^+ the unit step function in (100) evaluates to one and hence $\psi^{[0]} = A(\mathbf{k}_1^b, \mathbf{r}_1)/\lambda(\mathbf{k}_1^b)$. In this case numerical trajectory construction is continued by realizing a scattering event from $\mathbf{k}_1^b$ to $\mathbf{k}_1$ and by choosing t_1. Again, if $t_1 < t_{b1}^+$ compute $\psi^{[1]} = \psi^{[0]} + A(\mathbf{k}_2^b, \mathbf{r}_2)/\lambda(\mathbf{k}_2^b)$. In principle, this process should be continued until $\psi^{[n]}$ is obtained. However, if in the course of numerical trajectory construction a time $t_l > t_{bl}^+$ is generated, the unit step function in (100) evaluates to zero, such that the recursion terminates and the random variable keeps the value $\psi^{[l-1]}$. This is the realization of a numerical trajectory terminating at the boundary during the l-th free flight.

For a numerical trajectory of arbitrary ordering number $i \leq N$, which terminates after $l+1$ free-flight segments, the random variable takes on the value

$$\psi_i = \frac{A(\mathbf{k}_1^b, \mathbf{r}_1)}{\lambda(\mathbf{k}_1^b)} + \ldots + \frac{A(\mathbf{k}_l^b, \mathbf{r}_l)}{\lambda(\mathbf{k}_l^b)} \tag{101}$$

The ψ_i given by the sum (101) are summed up in the sample mean (11). This gives a double sum that can be replaced by one sum over all before-scattering states that have been generated.

$$\langle\!\langle \mathrm{A} \rangle\!\rangle \simeq \Gamma_D \frac{1}{N} \sum_{i=1}^{N} \psi_i = \Gamma_D \frac{1}{N} \sum_b \frac{A(\mathbf{k}^b, \mathbf{r}^b)}{\lambda(\mathbf{k}^b)} \tag{102}$$

To derive the expressions for the time-recording formalism the zero order iteration term is reformulated as

$$\langle\!\langle \mathrm{A} \rangle\!\rangle_0 = \Gamma_D \oint_{\partial D} \mathrm{d}\sigma \int_{K_+} \mathrm{d}\mathbf{k}_b \int_0^{\infty} \mathrm{d}t_0 \, p_0(\mathbf{r}_b, \mathbf{k}_b, t_0) \int_0^{\tau_0} A(\mathbf{K}_b(\tau), \mathbf{R}_b(\tau)) \, \mathrm{d}\tau \, . \tag{103}$$

This expression shows $\psi^{[0]}$ explicitly. The recursive definition of the random variable changes from (100) to

$$\psi^{[0]} = \int_0^{\tau_0} A(\mathbf{K}_b(\tau), \mathbf{R}_b(\tau))\,\mathrm{d}\tau \tag{104}$$

$$\psi^{[n]} = \psi^{[n-1]} + \prod_{j=0}^{n-1} H(t_{bj}^{+} - t_j) \int_0^{\tau_n} A(\mathbf{K}_n(\tau), \mathbf{R}_n(\tau))\,\mathrm{d}\tau\,, \tag{105}$$

where $\tau_j = \min(t_{bj}^{+}, t_j)$.

For a numerical trajectory comprising $l+1$ free-flight segments the random variable takes on the value

$$\psi_i = \int_0^{t_0} A\,\mathrm{d}\tau + \ldots \int_0^{t_{l-1}} A\,\mathrm{d}\tau + \int_0^{t_{bl}} A\,\mathrm{d}\tau\,. \tag{106}$$

This sum is over free flights and contains therefore one element more than (101) does.

Choosing n in (97) a priori implies that a numerical trajectory cannot contain more than $n+1$ free-flight segments. This restriction can be omitted by always following a numerical trajectory until it terminates at the boundary, permitting numerical trajectories with arbitrary many free-flight segments. In this case the infinite series representing $\langle\!\langle \mathrm{A} \rangle\!\rangle$ is evaluated rather than the partial sum given by (97).

4.3.4. *Normalization of the distribution function*

The normalization constant Γ_D needs not be evaluated from the theoretical definition (20). Instead, by setting $A = 1$ a relation between Γ_D and the total number of particles N_D is obtained

$$4\pi^3 N_D = \Gamma_D \frac{1}{N} \sum_b \lambda(\mathbf{k}^b)^{-1} \tag{107}$$

where N_D is usually known, for instance, from the constraint of total charge neutrality in the device. Note that N is the number of trajectories constructed. Using the time-integration scheme, in the special case $A = 1$, the realization (106) represents the total time of the i-th numerical trajectory, $\psi_i = T_i$. Denoting with $T = \sum T_i$ the total time the particle has been followed, one finds from (107):

$$T = \sum_b \lambda(\mathbf{k}^b)^{-1}\,. \tag{108}$$

The finite sum recorded during the simulation is an unbiased estimate of the total time the particle path is followed.

4.4. *The weighted single-particle MC method*

Expressing the solution of the conjugate equation by the Neumann series leads to the series expansion (79)of the statistical average of $A(\mathbf{k}, \mathbf{r})$. The explicite expression of a term of that series is given by (86). There exists a variable transformation such that the multiple integral gets expressed as

$$\langle\langle \mathrm{A} \rangle\rangle_n = \Gamma_D \int \mathrm{d}\xi_0 \ldots \mathrm{d}\xi_n \; h(\xi_0) K(\xi_0, \xi_1) K(\xi_1, \xi_2) \ldots K(\xi_{n-1}, \xi_n) \frac{A(\xi_n)}{\lambda(\xi_n)} \tag{109}$$

where the before-scattering states $\xi_i = (\mathbf{k}^b_{i+1}, \mathbf{r}_{i+1})$ are chosen as integration variables. The integrand of (109) contains an initial distribution h and the transition probability K given by the kernel of the conjugate equation, (69).

The integral (109) can be written as

$$\langle\langle \mathrm{A} \rangle\rangle_n = \Gamma_D \int p(y)\psi(y)\mathrm{d}y \tag{110}$$

$$y = (\xi_0, \xi_1, \ldots \xi_n) \tag{111}$$

$$p = h_0(\xi_0) K(\xi_0, \xi_1) \ldots K(\xi_{n-1}, \xi_n) \tag{112}$$

$$\psi = \frac{A(\xi_n)}{\lambda(\xi_n)} \,. \tag{113}$$

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 $\overline{\psi} = N^{-1} \sum \psi_i$.

In (109) the initial density h and the transition probability K reflect the physical properties of the system. Therefore, these physically-based 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. In that case one constructs numerical trajectories that are different from the physical ones. Using arbitrary probabilities aims 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.

$$p(y) = p_0(\xi_0) P(\xi_0, \xi_1) \ldots P(\xi_{n-1}, \xi_n) \tag{114}$$

$$\psi = \frac{h_0(\xi_0) K(\xi_0, \xi_1) \ldots K(\xi_{n-1}, \xi_n)}{p_0(\xi_0) P(\xi_0, \xi_1) \ldots P(\xi_{n-1}, \xi_n)} \frac{A(\xi_n)}{\lambda(\xi_n)} \tag{115}$$

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, $\int p_0(\xi_0)\, \mathrm{d}\xi_0 = 1$ and $\int P(\xi_i, \xi_j)\, \mathrm{d}\xi_j = 1$ for all ξ_i.

The ratio of the physical density over the numerical density determines the weight of a particle.

$$w_n = \frac{h_0(\xi_0)K(\xi_0,\xi_1)\dots K(\xi_{n-1},\xi_n)}{p_0(\xi_0)P(\xi_0,\xi_1)\dots P(\xi_{n-1},\xi_n)} \tag{116}$$

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 numerical density rather than from a physical density, the weight of the trajectory changes by the ratio of the two densities.

While (115) is an estimator for one iteration term, in the Single Particle MC method one uses estimators for the whole iteration series, such as (106) and (101). In these estimators each element of the sum has to be multiplied by the weight defined above. In the case of event biasing the before-scattering and the time-averaging estimators, (101) and (106), respectively, gets extended to:

$$\psi_i = \sum_j^l w_j \frac{A(\mathbf{k}_j^b, \mathbf{r}_j)}{\lambda(\mathbf{k}_j^b, \mathbf{r}_j)}\,, \tag{117}$$

$$\psi_i = \sum_j^l w_j \int_0^{t_j} A(\mathbf{K}(\tau;0,\mathbf{k}_j^a,\mathbf{r}_j),\mathbf{R}(\tau;0,\mathbf{k}_j^a,\mathbf{r}_j))\,\mathrm{d}\tau\,. \tag{118}$$

4.4.1. *Modified probabilities*

The purpose of the event-bias 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, including those for selecting the free-flight time, the scattering mechanism, the after scattering state, or the initial state at a contact.

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

$$\lambda_a' = \lambda_a + \lambda_e\left(1 - \frac{1}{M_1}\right)\,, \qquad \lambda_e' = \frac{\lambda_e}{M_1}\,. \tag{119}$$

If in the MC simulation phonon absorption is selected, the particle weight is to 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 are not changed.

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,

$$p'(\chi) = \begin{cases} \dfrac{1}{2M_2} & -1 \le \chi < \chi_0 \\ \dfrac{M_2}{2} & \chi_0 \le \chi < 1 \end{cases} \tag{120}$$

where $M_2 \geq 1$ 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 $r < P'(\chi_0)$

$$\chi_r = 2M_2 r - 1, \qquad \frac{p}{p'} = M_2,$$

and otherwise

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

This means that the particle weight is either reduced or increased by the factor M_2 whenever χ is generated from the density (120).

To support the formation of an artificially heated carrier distribution at the rising edge of an energy barrier one might inject particles from some nearby contact assuming some heated boundary distribution.

Consider a Maxwellian at lattice temperature T_0 and a heated Maxwellian at temperature $T' = M_3 T_0$ with $M_3 > 1$.

$$f_b(\mathbf{k}, \mathbf{r}) = C(\mathbf{r}) \exp\left(-\frac{\epsilon(\mathbf{k})}{k_B T_0}\right) \tag{121}$$

$$f'_b(\mathbf{k}, \mathbf{r}) = C'(\mathbf{r}) \exp\left(-\frac{\epsilon(\mathbf{k})}{k_B T'}\right) \tag{122}$$

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

$$j_\perp(\mathbf{r}) = -C(\mathbf{r}) \int\limits_{\mathbf{n}\cdot\mathbf{v}<0} \mathbf{n}(\mathbf{r}) \cdot v(\mathbf{k}) \exp\left(-\frac{\epsilon(\mathbf{k})}{k_B T_0}\right) \, \mathrm{d}^3 k \tag{123}$$

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

$$j_\perp(\mathbf{r}) = \frac{k_B}{\hbar} C(\mathbf{r}) T_0 \int\limits_{\mathbf{n}\cdot\mathbf{v}<0} \mathbf{n} \cdot \nabla \exp\left(-\frac{\epsilon(\mathbf{k})}{k_B T_0}\right) \, \mathrm{d}^3 k \tag{124}$$

$$= \frac{k_B}{\hbar} C(\mathbf{r}) T_0 \int\int \exp\left(-\frac{\epsilon(0, k_y, k_z)}{k_B T_0}\right) \, \mathrm{d}k_y \mathrm{d}k_z \tag{125}$$

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 non-zero.

Note that (121) is the physical boundary distribution and hence the parameter $C(\mathbf{r})$ is known. The normalization factor $C'(\mathbf{r})$ in (122) is obtained from the condition $j'_\perp = j_\perp$. This requires evaluation of the double integral (125). For simple non-parabolic bands

$$\epsilon(1+\alpha\epsilon) = \frac{\hbar^2\mathbf{k}^2}{2m^*} \tag{126}$$

C' must satisfy

$$\frac{C(\mathbf{r})}{C'(\mathbf{r})} = \frac{T'^2(1+2\alpha T')}{T_0^2(1+2\alpha T_0)}\,. \tag{127}$$

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

$$w_0 = \frac{v_\perp f_b(\mathbf{k},\mathbf{r})}{v_\perp f'_b(\mathbf{k},\mathbf{r})} = M_3^2 \frac{1+2\alpha M_3 T_0}{1+2\alpha T_0} \exp\left(-\frac{\epsilon(\mathbf{k})}{k_B T_0}\left(1-\frac{1}{M_3}\right)\right) \tag{128}$$

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

4.4.2. *Evolution of the weights*

The particle weight (116) evolves randomly along a numerical trajectory. MC simulations show that at a given time the weights on different trajectories show a large spreading. Most of the weights evolve to extremely small values, such that new terms in the weighted sums (118) or (117) sooner or later becomes negligible.

This behavior can be investigated analytically for the simple case of the density function (120), 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}. The expected value of the selected multipliers equals $p_0 M + p_1 M^{-1} = 1$.

Consider a numerical 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 then be estimated as

$$w_B = M^{p_0 B} \cdot M^{-p_1 B} = \exp(-\alpha B) \tag{129}$$

$$\alpha = \frac{M-1}{M+1}\log M \tag{130}$$

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 physical meaning of the trajectory diminishes with increasing number of biased events, as the contributions to the estimator (117) continuously decreases.

As discussed in Section 4.3.4, setting $A = 1$ the before-scattering estimator (117) 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

$$T = \sum_n \frac{w_n}{\Gamma} = \frac{1}{\Gamma} \sum_{n=0}^{\infty} \exp(-\alpha n) = \frac{1}{\Gamma(1 - \exp(-\alpha))} \tag{131}$$

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.

4.4.3. *Results and discussion*

A one-dimensional *npn* structure has been analyzed. The three segments of the device are referred to as emitter, base, and collector. The semiconductor model of silicon assumes an analytical band structure[23].

The modified probabilities described in the previous Section have been used to simulate electron transport through the *npn*-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 so as 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. On the other hand, 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 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-bias technique suggests the usage of additional variance reduction techniques[24]. The general goal must be a reduction of the spreading of the weights. Such techniques are not used in this study. Instead, the evolution of the particle weight is governed predominantly by the event-bias algorithm. Explicit measures are taken only to prevent weights from getting extremely high or low. The rare event that 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 and only physical probabilities are used, such that the weight is not further changed.

The event-bias method has been compared with a simple particle split method. To first order such comparison is fair since the light-weight particles generated with either method are not further recycled. Fig.7 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 the physical mean energy.

In the simulation shown in Fig.8 a biased boundary distribution is also assumed.

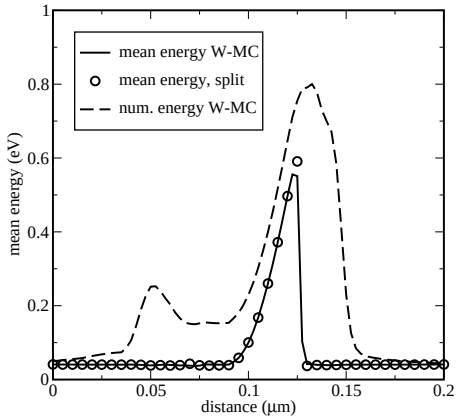

Figure 7: Mean energy of the physical system (mean) and of the simulated carriers (num.) in the *npn*-structure with 0.8 eV barrier height. Comparison of the event-bias method (W-MC) and a particle split method is shown.

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.9 the electron concentration and the standard deviations of the two MC methods are depicted. In the quasi-neutral base region (75–90nm) event biasing gives a standard deviation reduced by more than one order of magnitude.

4.5. *The single-particle backward MC algorithm*

Approaching the Neumann series of (65) by the MC method yields steady-state backward algorithms. Two algorithms are found. The first one allows to evaluate the distribution function at given points and is basically identical with the transient backward algorithm (Section 3.5) with the only difference that the numerical trajectory is followed in a variable time interval $(t_b^-(\mathbf{k},\mathbf{r}),0)$ rather than in a predefined one. Let us consider the problem of injection of channel hot carriers into the gate oxide. Using a backward algorithm carriers are launched at the semiconductor/oxide interface only at energies above the relevant energy threshold. In other words, only the rare events are simulated. Each high energetic carrier is followed back in time until it reaches an equilibrium region such as source or drain, where the distribution function is known. Since each trajectory is of different duration, a steady state formulation employing a variable time $t_b^-(\mathbf{k},\mathbf{r})$ and a boundary distribution appears appropriate.

The second algorithm can be viewed as the backward version of the Singe-Particle forward algorithm. N trajectories are constructed starting from an absorbing boundary. The weight of the particle changes by λ^*/λ at each scattering event, however, the weight remains undeterminated with respect to a scaling factor. Both the before-scattering and the time-recording method for average recording are

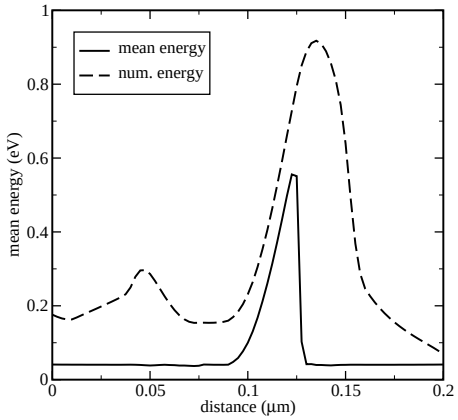

Figure 8: Mean energy in the 0.8 eV structure. In addition to Fig. 7, electrons are injected at the left contact at 1500 K, whereas at the right contact a Maxwellian at 300 K is assumed.

available, yet the current weight of the particle has to be taken into account. The particle weights are finally determined when the trajectory terminates at an injecting boundary, where the boundary distribution f_b evaluated at the reached phase space point gives the scaling factor.

5. Small-Signal MC Algorithms

Understanding the MC method as a versatile tool to solve integral equations enables its application to a class of problems which are not accessible by purely physically-based, imitative MC methods. One such class, which plays an important role in electrical engineering, is the linearized small signal analysis of nonlinear systems. Whether the linearized system is analyzed in the frequency or time domain is just a matter of convenience since the system responses obtained are linked by the Fourier transform.

At present, linear small signal analysis of semiconductor devices by the MC method is beyond the state of the art. The established technique to study the small signal AC characteristics consists of an EMC simulation followed by a Fourier transform of the step response currents. Since the EMC simulation captures the general nonlinear behavior of a device the voltage increment must be sufficiently small in order to stay in the linear response regime.

For small signal analysis of bulk carrier transport, however, various MC algorithms have been reported[25,26,27,28]. The formal approach pointed out in the previous Sections allows new and existing algorithms to be derived in a unified way[29].

Choosing a formulation in the time domain, a small perturbation $\mathbf{E}_1$ is superimposed to a stationary field $\mathbf{E}_s$. The stationary distribution function f_s will thus

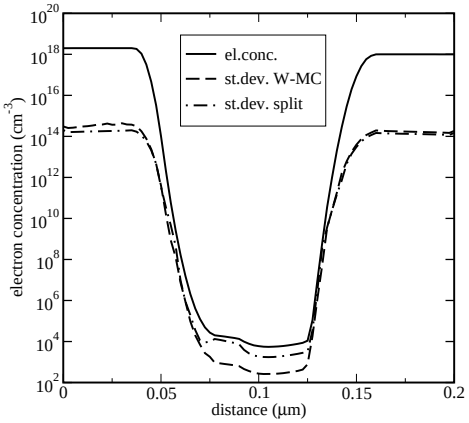

Figure 9: The electron concentration varies by more than 14 orders. In the base region the event-bias method gives significantly less variance than the split method.

be perturbed by some small quantity f_1.

$$\mathbf{E}(t) = \mathbf{E}_s + \mathbf{E}_1(t) \tag{132}$$

$$f(\mathbf{k},t) = f_s(\mathbf{k}) + f_1(\mathbf{k},t) \tag{133}$$

Inserting this Ansatz into the transient Boltzmann equation and retaining only first order perturbation terms yield a Boltzmann-like equation for f_1 which is linear in the perturbation $\mathbf{E}_1$.

$$\frac{\partial f_1(\mathbf{k},t)}{\partial t} + \frac{q}{\hbar}\mathbf{E}_s \cdot \nabla f_1(\mathbf{k},t) = Q[f_1](\mathbf{k},t) - \frac{q}{\hbar}\mathbf{E}_1(t) \cdot \nabla f_s(\mathbf{k}) \tag{134}$$

Compared with the common Boltzmann Equation, (134) has an additional term on the right hand side which contains f_s, the solution of the stationary Boltzmann Equation. The integro-differential type of equation, (134), is transformed into an integral form. Assuming an impulse-like excitation $\mathbf{E}_1(t) = \delta(t)\mathbf{E}_{\mathrm{im}}$ results in the following integral equation for the impulse response f_1,

$$f_1(\mathbf{k},t) = \int_0^t \mathrm{d}t' \int \mathrm{d}\mathbf{k}' f_1(\mathbf{k}',t') S(\mathbf{k}',\mathbf{K}(t')) \exp\Big(-\int_{t'}^{t} \lambda(\mathbf{K}(y))dy\Big) + G(\mathbf{K}(0)) \exp\Big(-\int_0^t \lambda(\mathbf{K}(y))dy\Big) \tag{135}$$

with

$$G(\mathbf{k}) = -\frac{q}{\hbar}\mathbf{E}_{\mathrm{im}} \cdot \nabla f_s(\mathbf{k})\,. \tag{136}$$

The free term of (135) is formally equivalent to the free term of the Boltzmann Equation. The only difference is that G takes on also negative values, and can therefore not be interpreted as an initial distribution. Various treatments of the term G can be devised giving rise to a variety of MC algorithms, all of which solve (135). G can be expressed as a difference of two positive functions, $G = G^+ - G^-$, an Ansatz which decomposes (135) into two common Boltzmann Equation for the unknowns f_1^+ and f_1^-. The initial conditions of these Boltzmann Equations are $f_1^\pm(\mathbf{k}, 0) = G^\pm(\mathbf{k}) \geq 0$. In this way the impulse response is understood in terms of the concurrent evolution of two carrier ensembles.

Using different methods to generate the initial distributions of the two ensembles gives rise to a variety of MC algorithms. Both existing and new MC algorithms are obtained in a unified way, and a transparent, physical interpretation of the algorithms is supported. For vanishing electric field, f_s is given by the equilibrium distribution and (136) can be evaluated explicitly. An efficient MC method for computation of the exact zero-field mobility is then obtained[30].

In the case that the stationary and the small signal field vectors are collinear, the stationary Boltzmann Equation can be used to express the distribution function gradient as

$$G(\mathbf{k}) = \frac{E_{\mathrm{im}}}{E_s} \left(\lambda(\mathbf{k}) f_s(\mathbf{k}) - \int f_s(\mathbf{k}') S(\mathbf{k}', \mathbf{k}) \mathrm{d}\mathbf{k}' \right) , \tag{137}$$

which gives a natural splitting of G into two positive functions. In the following we adopt the notation that terms which are employed in the respective algorithm as a probability density are enclosed in curly brackets.

From (137) we choose the initial distributions as

$$G^+(\mathbf{k}) = \frac{E_{\mathrm{im}}}{E_s} \langle \lambda \rangle_s \left\{ \frac{\lambda(\mathbf{k}) f_s(\mathbf{k})}{\langle \lambda \rangle_s} \right\} \tag{138}$$

$$G^-(\mathbf{k}') = \frac{E_{\mathrm{im}}}{E_s} \langle \lambda \rangle_s \int \left\{ \frac{\lambda(\mathbf{k}) f_s(\mathbf{k})}{\langle \lambda \rangle_s} \right\} \left\{ \frac{S(\mathbf{k}, \mathbf{k}')}{\lambda(\mathbf{k})} \right\} \mathrm{d}\mathbf{k} \tag{139}$$

where $\langle \lambda \rangle_s = \int f_s(\mathbf{k}) \lambda(\mathbf{k}) \mathrm{d}\mathbf{k}$ is introduced in the denominators to ensure normalization. $\langle \lambda \rangle_s$ is the inverse of the mean free-flight time, which can be seen immediately when evaluating the average by means of the 'before-scattering' method. The probability density $\lambda f_s / \langle \lambda \rangle_s$ represents the normalized distribution function of the before-scattering states. Consequently, the product of the two densities in (139) represents the normalized distribution function of the after-scattering states. Using the above expression the following algorithm can be formulated.

1) Follow a main trajectory for one free flight, store the before-scattering state in $\mathbf{k}_b$, and realize a scattering event from $\mathbf{k}_b$ to $\mathbf{k}_a$.

2) Start a trajectory $\mathbf{K}^+(t)$ from $\mathbf{k}_b$ and another trajectory $\mathbf{K}^-(t)$ from $\mathbf{k}_a$.

3) Follow both trajectories for time T. At equidistant times t_i add $A(\mathbf{K}^+(t_i))$ to a histogram ν_i^+ and $A(\mathbf{K}^-(t_i))$ to a histogram ν_i^-.

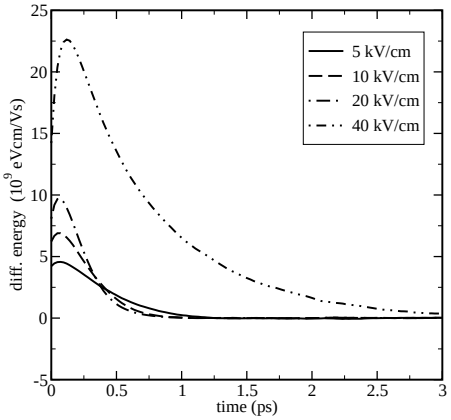

Figure 10: Impulse response of the differential energy.

4) Continue with the first step until N **k**-points have been generated.

5) Calculate the time discrete impulse response as $\langle A\rangle_{\rm im}(t_i) = \frac{E_{\rm im}\langle\lambda\rangle_s}{NE_s}(\nu_i^+ - \nu_i^-)$.

The mean free-flight time must be additionally calculated during the simulation. This algorithm shows in a transparent way the evolution of the P and M ensembles, as well as the generation of the initial states for those ensembles.

For electrons in Si the impulse response of mean energy and mean velocity has been calculated. Fig. 10 shows the response of the differential energy $\langle\epsilon\rangle_1/E_{\rm im}$ in the time domain for different field strengths. The frequency-dependent differential velocity obtained by a Fourier transform of the impulse response $\langle v\rangle_1/E_{\rm im}$ is plotted in Fig. 11 and Fig. 12. The low frequency limit of the real part tends to the static differential mobility $\partial\langle v\rangle_s/\partial E_s$.

6. The Stationary Wigner-Boltzmann Equation

At room temperature the electrical characteristics of nanoelectronic and highly down-scaled microelectronic devices are influenced simultaneously by semiclassical and quantum mechanical effects. A kinetic equation suitable for describing this mixed transport regime is given by the Wigner equation. This equation can be formulated in such a way that it simplifies to the semiclassical Boltzmann equation in those device regions where quantum effects are negligible. The MC method has proven to be a reliable and accurate numerical method for solving the Boltzmann equation. Therefore, it appears very promising to devise MC techniques also for the solution of the Wigner equation. The advantage of a particle method is that semiclassical scattering from various sources can be included in a straightforward way. The major problem to be overcome originates from the scattering kernel of the Wigner equation, which is, as opposed to the semiclassical case, no longer positive. A solution to this so-called negative-sign problem is presented in the following for

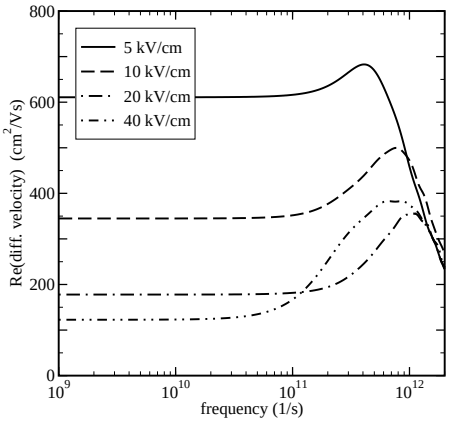

Figure 11: Real part of the differential velocity.

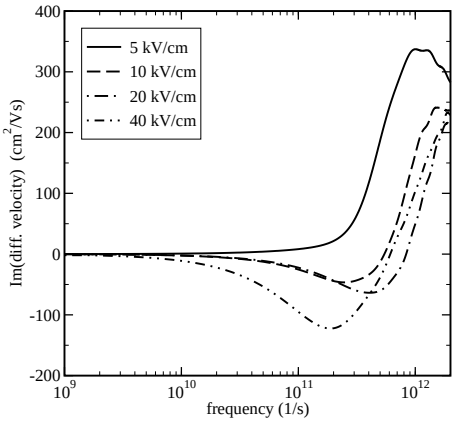

Figure 12: Imaginary part of the differential velocity.

the stationary case[31].

We consider the space-dependent Wigner equation, including semiclassical scattering via the Boltzmann collision operator $Q[f_w]$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla_r + q\mathbf{E}\cdot\nabla_k\right) f_w = Q[f_w] + \Theta_w[f_w]\,, \tag{140}$$

$$\Theta_w[f_w](\mathbf{k},\mathbf{r},t) = \int V_w(\mathbf{r},\mathbf{k}-\mathbf{k}')f_w(\mathbf{k}',\mathbf{r},t)d\mathbf{k}'\,. \tag{141}$$

The classical force term $q\mathbf{E}$ is separated from the Wigner potential[32]

$$V_w(\mathbf{r},\mathbf{k}) = \frac{1}{2\pi i\hbar}\int \Big(V(\mathbf{r}+\frac{\mathbf{s}}{2}) - V(\mathbf{r}-\frac{\mathbf{s}}{2}) + q\mathbf{s}\cdot\mathbf{E}\Big)\exp(-i\mathbf{k}\cdot\mathbf{s})\,d\mathbf{s}\,, \tag{142}$$

and thus appears in the Liouville operator on the left-hand side of (140). The kinetic equation (140) has now the form of a Boltzmann equation with an additional term caused by the Wigner potential. Whether the collision operator or the potential operator is dominant depends on the device under consideration. The chosen formulation of the Wigner equation ensures that in the classical limit the Boltzmann equation is obtained. Consequently, the MC method presented below simplifies gradually to the classical MC method when the Wigner potential vanishes. Therefore, an artificial separation of the simulation domain into a quantum and classical region and application of different numerical methods is avoided.

6.1. *The particle model*

Because the Wigner potential assumes positive and negative values, it cannot be used directly as a scattering probability. To permit a probabilistic interpretation,

V_w is expressed as a difference of two positive functions. Introducing the truncated Wigner potential

$$V_w^+(\mathbf{k}) = \begin{cases} V_w(\mathbf{k}), & V_w(\mathbf{k}) \geq 0 \\ 0, & V_w(\mathbf{k}) < 0 \end{cases} \tag{143}$$

and accounting for the antisymmetry of V_w with respect to $\mathbf{k}$, the potential operator can be expressed as

$$\Theta_w[f_w](\mathbf{k}) = \int V^+(\mathbf{q})[f_w(\mathbf{k}-\mathbf{q}) - f_w(\mathbf{k}+\mathbf{q})]\, d\mathbf{q}\,. \tag{144}$$

In terms of a particle model, the positive and negative terms of the integrand might be interpreted as in-scattering and out-scattering terms, respectively. However, the out-scattering operator is non-local in $\mathbf{k}$-space, whereas, for comparison, the semiclassical out-scattering operator is local. Therefore, (144) does not describe a scattering in the sense that an initial state is annihilated and a final state is created. Instead, (144) describes the creation of two new states, $\mathbf{k}-\mathbf{q}$ and $\mathbf{k}+\mathbf{q}$. When generating the second state, the sign of the statistical weight is changed.

$$\gamma(\mathbf{r}) = \int V_w^+(\mathbf{r},\mathbf{k})\, d\mathbf{k}\,. \tag{145}$$

It should be noted that the Wigner equation strictly conserves charge, as can be seen by taking the zero-order moment of (140)

$$\frac{\partial n}{\partial t} + \operatorname{div} \mathbf{J} = 0\,.$$

Looking at the number of particles regardless of their statistical weight, that is, counting each particle as positive, another potential operator needs to be considered.

$$\Theta_w^*[f_w](\mathbf{k}) = \int V^+(\mathbf{q})[f_w(\mathbf{k}-\mathbf{q}) + f_w(\mathbf{k}+\mathbf{q})]\, d\mathbf{q} \tag{146}$$

Using (146), a continuity equation for numerical particles is obtained.

$$\frac{\partial n^*}{\partial t} + \operatorname{div} \mathbf{J}^* = 2\gamma(\mathbf{r})n^* \tag{147}$$

The high generation rate in this equation is a direct consequence of the negative-sign problem. Not only can we expect cancellation effects in the estimators due to the positive and negative statistical weights, but also an exponential growth in time of certain quantities, such as particle number, particle weight or variance.

6.2. *Stationary MC method*

The particle model described in the previous sections provides a guideline for the development of new and the characterization of existing MC algorithms for solving the Wigner equation. Applying a formal approach, which employs the Neumann

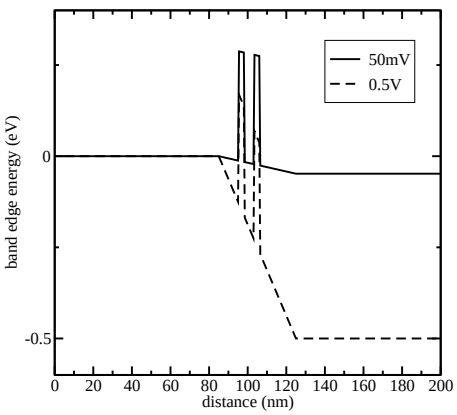

Figure 13: Conduction band edge of the RTD for different voltages.

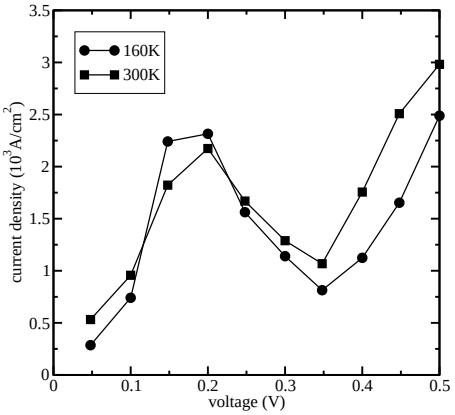

Figure 14: IV-characteristics of the RTD for two different temperatures.

series expansion of the underlying transport equation[33], leads to a MC algorithm with two significant properties: the number of numerical particles is conserved, and the particle weights increase exponentially in time. Using this algorithm it has been demonstrated that tunneling can be treated numerically by means of a particle model[34]. However, because of the exponential increase of the absolute value of the particle weight at the very short time scale $(2\gamma)^{-1}$ (see (147)), application of this algorithm turned out to be restricted to single-barrier tunneling and small barrier heights only.

For the simulation of double-barrier structures another MC algorithm has been designed, which now conserves the statistical weight. In return, particles are generated at the rate of 2γ. The kinetic equation (140) is interpreted as a Boltzmann equation augmented by a generation term Θ_w. Thus, in principle, any MC method for solving the Boltzmann equation can be employed, extended by a mechanism for generating particle pairs. The challenge of employing such algorithm is to handle the avalanche of numerical particles properly. This problem has been solved for stationary conditions. Particles of opposite weight and a sufficiently small distance in phase space are continuously removed in the course of a simulation.

A Resonant tunneling diode (RTD) has been simulated using the quantum MC algorithm (Fig.13). The temperature dependence of the current-voltage characteristics of the RTD is shown in Fig. 14. The resonance current is higher at low temperature due to the smaller spreading of the energy distribution, wheres the valley current increases with temperature.

Fig. 15 shows the electron concentration in the device at voltages below the resonance voltage. A classical behavior is found before and after the double barrier, whereas in the quantum well the behavior of the solution is non-classical. In the quantum well the concentration increases as the resonance is approached. After

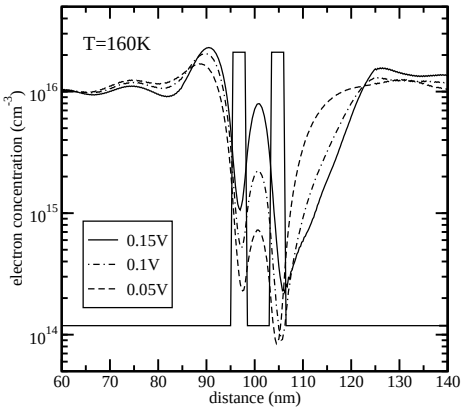

Figure 15: Electron concentration in the RTD for voltages less than the resonance voltage.

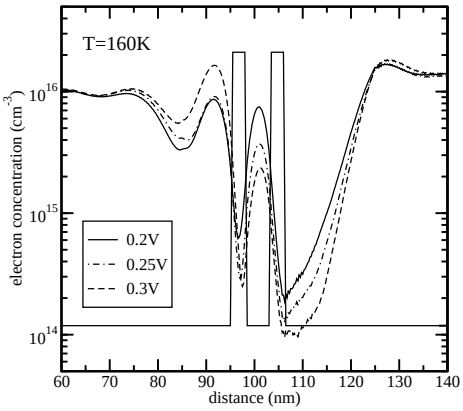

Figure 16: Electron concentration in the RTD for voltages greater than the resonance voltage.

the barrier a depletion layer forms, which grows with applied voltage. For voltages above the resonance voltage, the concentration in the well drops, while the depletion layer continues to grow.

7. Conclusion

Application of the numerical MC method for solving integral equations has been thoroughly studied for the cases of the transient and stationary Boltzmann equation. The well known Ensemble MC and the Single-Particle MC algorithms have been rederived in a unified way. In addition, specific algorithms which are very seldomly used have been discussed. The backward MC algorithm is well suited for the simulation of rare events, whereas the weighted MC method has been demonstrated to be an effective statistical enhancement technique. The same numerical approach proven to be successful for the semi-classical transport problem has been applied to solve the Wigner-Boltzmann equation. A MC method for the simulation of far-from-equilibrium transport in nano-structures has been developed. Treating the Wigner potential operator as a source of scattering is complicated by the so-called negative sign problem, which would lead to a run-away of variance, unless a proper variance reduction technique is used.

Acknowledgment

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

References

1. I. Sobol, *The Monte Carlo Method* (Mir Publishers, Moscow, 1984).

2. F. Byron and R. Fuller, *Mathematics of Classical and Quantum Physics* (Dover, New York, 1992).
3. R. Hockney and J. Eastwood, *Computer Simulation Using Particles* (Adam Hilger, Bristol and Philadelphia, 1988).
4. C. Moglestue, Computer Methods in Applied Mechanics and Engineering **30**, 173 (1982).
5. C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Wien-New York, 1989).
6. A. Reklaitis, Phys.Lett. **88A**, 367 (1982).
7. *Hot-Electron Transport in Semiconductors*, Vol. 58 of *Topics in Applied Physics*, edited by L. Reggiani (Springer Verlag, Berlin, Heidelberg, New York, Tokyo, 1985).
8. M. Nedjalkov and P. Vitanov, Solid-State Electron. **32**, 893 (1989).
9. C. Jacoboni, P. Poli, and L. Rota, Solid-State Electron. **31**, 523 (1988).
10. C. Jacoboni, in *Int.Electron Devices Meeting* (IEEE Electron Devices Society, Washington, D.C., 1989), pp. 469–472.
11. L. Rota, C. Jacoboni, and P. Poli, Solid-State Electron. **32**, 1417 (1989).
12. R. Chambers, Proc. Phys. Soc. (London) **A65**, 458 (1952).
13. M. Nedjalkov and P. Vitanov, Solid-State Electron. **33**, 407 (1990).
14. P. Vitanov and M. Nedjalkov, COMPEL **10**, 531 (1991).
15. M. Nedjalkov and I. Dimov, Mathematics and Computers in Simulations **47**, 383 (1998).
16. M. Nedjalkov and H. Kosina, Mathematics and Computers in Simulations **55**, 191 (2001).
17. P. Price, IBM Journal of Research and Development **14**, 12 (1970).
18. W. Fawcett, A. Boardman, and S. Swain, J. Phys. Chem. Solids **31**, 1963 (1970).
19. G. Baccarani, C.Jacoboni, and A. M. Mazzone, Solid-State Electron. **20**, 5 (1977).
20. H. Kosina, M. Nedjalkov, and S. Selberherr, J.Appl.Phys. **93**, 3553 (2003).
21. M. Nedjalkov, H. Kosina, and S. Selberherr, japlph **93**, 3564 (2003).
22. M. Lundstrom, *Fundamentals of Carrier Transport* (University Press, Cambridge, 2000).
23. C. Jacoboni and L. Reggiani, Review of Modern Physics **55**, 645 (1983).
24. C. Wordelman, T. Kwan, and C. Snell, IEEE Trans.Computer-Aided Design **17**, 1230 (1998).
25. P. Lebwohl, J.Appl.Phys. **44**, 1744 (1973).
26. J. Zimmermann, Y. Leroy, and E. Constant, J.Appl.Phys. **49**, 3378 (1978).
27. P. Price, J.Appl.Phys. **54**, 3616 (1983).
28. E. Starikov and P. Shiktorov, Sov. Phys. Semicond. **22**, 45 (1988).
29. H. Kosina, M. Nedjalkov, and S. Selberherr, J.Appl.Phys. **87**, 4308 (2000).
30. H. Kosina, M. Nedjalkov, and S. Selberherr, in *Large-Scale Scientific Computing*, edited by S. Margenov, J. Wasniewski, and P. Yalamov (Springer, Berlin Heidelberg, 2001), pp. 175–182.
31. H. Kosina, M. Nedjalkov, and S. Selberherr, in *Nanotech* (Computational Publications, San Francisco, 2003), pp. 190–193.
32. P. Bordone, A. Bertoni, R. Brunetti, and C. Jacoboni, Mathematics and Computers in Simulation **62**, 307 (2003).
33. H. Kosina, M. Nedjalkov, and S. Selberherr, IEEE Trans.Electron Devices **47**, 1898 (2000).
34. M. Nedjalkov, R. Kosik, H. Kosina, and S. Selberherr, in *Proc. Simulation of Semiconductor Processes and Devices* (Business Center for Academic Societies Japan, Kobe, Japan, 2002), pp. 187–190.