Monte Carlo Methods Appl. Vol. 13 No. 4 (2007), pp. 299–331
DOI 10.1515/MCMA.2007.017

Mixed initial-boundary value problem in particle modeling of microelectronic devices

M. Nedjalkov, D. Vasileska, I. Dimov and G. Arsov

Abstract. The Boltzmann equation in presence of boundary and initial conditions, which describes the general case of carrier transport in microelectronic devices is analysed in terms of Monte Carlo theory. The classical Ensemble Monte Carlo algorithm which has been devised by merely phenomenological considerations of the initial and boundary carrier contributions is now derived in a formal way. The approach allows to suggest a set of event-biasing algorithms for statistical enhancement as an alternative of the population control technique, which is virtually the only algorithm currently used in particle simulators. The scheme of the self-consistent coupling of Boltzmann and Poisson equation is considered for the case of weighted particles. It is shown that particles survive the successive iteration steps.

Keywords. Boltzmann equation, carrier transport in semiconductors, event biasing, integral equations.

1. Introduction

As semiconductor feature sizes shrink into the nanometer scale regime, even conventional device behavior becomes increasingly complicated as new physical phenomena at short dimensions occur, and limitations in material properties are reached. In addition to the problems related to the understanding of actual operation of ultra-small devices, the reduced feature sizes require more complicated and time-consuming manufacturing processes. This fact signifies that a pure trial-and-error approach to device optimization will become impossible since it is both too time consuming and too expensive. Since computers are considerably cheaper resources, simulation is becoming an indispensable tool for the device engineer. Besides offering the possibility to test hypothetical devices which have not (or could not) yet been manufactured, simulation

First author: This work has been supported by Österreichische Forschungsgemeinschaft (ÖFG), Project MOEL239.

offers unique insight into device behavior by allowing the observation of phenomena that can not be measured on real devices. Computational Electronics in this context refers to the physical simulation of semiconductor devices in terms of charge transport and the corresponding electrical behavior. It is related to, but usually separate from process simulation, which deals with various physical processes such as material growth, oxidation, impurity diffusion, etching, and metal deposition inherent in device fabrication leading to integrated circuits.

Device simulation can be thought of as one component of technology for computer-aided design, which provides a basis for device modeling. The goal is to provide simulation tools with the necessary level of sophistication to capture the essential physics while at the same time minimizing the computational burden so that results may be obtained within a reasonable time frame.

There are two main kernels, which must be solved self-consistently with one another, at any level of semiconductor device simulation: the transport equations governing charge flow, and the fields driving charge flow. Both are coupled strongly to one another, and hence must be solved simultaneously. The fields arise from external sources, as well as the charge and current densities which act as sources for the time varying electric and magnetic fields obtained from the solution of Maxwell's equations. Under appropriate conditions, only the quasi-static electric fields described by the solution of Poisson's equation are necessary. The fields, in turn, are driving forces for charge transport modeled at various levels of approximation within a hierarchical structure ranging from compact modeling to an exact quantum mechanical description. At the very beginnings of semiconductor technology, the electrical device characteristics could be estimated using simple analytical models relying on the drift-diffusion (DD) formalism. Various approximations had to be made to obtain closed-form solutions, but the resulting models captured the basic features of the devices. These approximations include simplified doping profiles and device geometries. With the ongoing refinements and improvements in technology, these approximations lost their basis and a more accurate description was required. This goal could be achieved by solving the DD equations numerically. Furthermore, as semiconductor devices were scaled into the submicrometer regime, the assumptions underlying the DD model lost their validity. Therefore, the transport models have been continuously refined and extended to more accurately capture transport phenomena occurring in these devices. The need for refinement and extension is primarily caused by the ongoing feature size

reduction in state-of-the-art technology. As the supply voltages can not be scaled accordingly without jeopardizing the circuit performance, the electric field inside the devices has increased. A large electric field which rapidly changes over small length scales gives rise to non-local and hot-carrier effects which begin to dominate device performance. An accurate description of these phenomena is required and is becoming a primary concern for industrial applications. To overcome some of the limitations of the DD model, extensions have been proposed which basically add an additional balance equation for the average carrier energy. However, such tools do not have predictive capability for ultra-small structures, for which it is necessary to consider the Boltzmann transport equation. This equation provides the most complete classical description of the carrier transport, and is inherently associated with evolution of particles and thus with the Monte Carlo method. It will be introduced in details in the next section. Here, to give a more complete picture we add few words about the transport models beyond the Boltzmann equation: the quantum models. Moving downwards to the quantum area in the hierarchical map of transport models, at the very bottom we have the Green's function approach. The latter is the most exact, but at the same time the most difficult of all. One level above are approaches based on the density matrix or the defined via the Fourier transform of the latter Wigner function. These are Markovian in time. In contrast, the Green's functions method allows one to consider simultaneously correlations in space and time, both of which are expected to be important in nano-scale devices. However, the lack of a direct physical interpretation of the formal quantum models and the enormous computational burden needed for their implementation make the usefulness in understanding quantum effects in actual devices of limited values. For example, the only successful utilization of the Green's function approach commercially is the NEMO (Nano-Electronics Modeling) simulator which is effectively 1D and is primarily applicable to resonant tunneling diodes. Thus, effective computational methods for quantum transport are still in process of establishment. Many works utilize the alternative way, namely to incorporate quantum models on top of the Boltzmann equation.

Within the requirement of self-consistently solving the coupled transport-field problem in this emerging domain of device physics, there are several computational challenges, which limit this ability. One is the necessity to solve both the transport and the Poisson's equations over the full 3D domain of the device (and beyond if one includes radiation effects). As a result, highly efficient algorithms targeted to high-end

computational platforms (most likely in a multi-processor environment) are required to fully solve even the appropriate field problems. The appropriate level of approximation necessary to capture the proper non-equilibrium transport physics relevant to a future device model is an even more challenging problem both computationally and from a fundamental physics framework.

From the above discussion it follows that, one must to perform particle-based device simulations that solve the Boltzmann transport equation self-consistently with the Poisson equation, as the latter gives the fields that drive the carriers during the free-flight portion of the Monte Carlo method. A development of formal approaches to this problem gives novel opportunities for optimization and statistical enhancement and assists the incorporation of quantum models.

The Monte Carlo method for device simulations is so transparent from a physical point of view, that is frequently interpreted as emulation of the process of carrier transport as viewed by the classical Boltzmann picture. According to this picture point-like particles drift over Newton's trajectories under the action of the applied electric field. The drift processes are interrupted by scattering events which are due to lattice imperfections such as lattice vibrations - phonons, impurities and other defects. The physical model of the particular semiconductor provides information about the carrier effective masses, dispersion relations and the variety of interaction mechanisms between carriers and the lattice imperfections. This information is incorporated into physical probability functions generally depending on the phase space coordinates formed by carrier wave vector and position, and which govern the carrier evolution. As the Monte Carlo method simulates the probabilistic evolution of the individual carriers, it provides the carrier distribution satisfying the Boltzmann equation. This phenomenological result is in accordance with the fact that the equation itself is derived from phenomenological considerations, by accounting the number of carriers, exchanged due to the applied forces, velocities and scattering events between given small phase space region and the rest of the phase space. The formal proof that the used algorithms solve the equation was carried out later, for the homogeneous, single particle steady-state algorithm [1], for the inhomogeneous counterpart [2], and for the transient Ensemble Monte Carlo algorithm (EMC) for an initial condition problem [3].

The alternative way to approach the Boltzmann equation with the methods of the numerical Monte Carlo theory was reported in end of the 1980's [4], [5]. This mathematically-based approach lead to the development of new algorithms such as the Weighted

Ensemble Monte Carlo method [6], [7] and the Backward Monte Carlo method [4], [5]. These early works focused on the transient case of evolution of an initial distribution of carriers. The potential of this approach for treatment of quantum transport problems has been also recognized [8]. Indeed, only formal approaches are possible in this case because of the lack of direct physical probability functions which govern the quantum evolution.

The approach has been further applied to the stationary Boltzmann equation with boundary conditions, giving rise to the single particle counterparts of the weighted and backward algorithms, as well as event biasing schemes for statistical enhancement [9], [10]. Statistical enhancement aims at reduction of the time necessary for computation of the desired device characteristics. Enhancement algorithms are especially useful when rare events in the transport process control the device behavior.

As based on the linear Boltzmann equation (BE), the approach is relevant for frozen field transport, where Coulomb interactions between particles are neglected. If the latter are enabled, the problem becomes nonlinear through the electric force, which now becomes dependent on the distribution function. Recently also this problem has been resolved within an analysis of the iterative procedure of the self-consistent coupling between the Boltzmann and Poisson equations. The general transient problem in presence of initial and boundary conditions has been considered and event biasing has been applied for simulation of real structures. The obtained results as well as a discussion of the physical and application aspects of the approach are given in [11]. In this work we present the theoretical aspects of the approach. In particular a detailed derivation of the event-biasing scheme is provided, beginning with the integral form of the transport equation and the corresponding adjoint equation. It is shown that the separation of the contributions form the initial and boundary conditions becomes obsolete in the special case of the EMC. Event biasing is introduced and some basic biasing models are discussed. Finally the self-consistent scheme for event biasing is analyzed.

A short introduction to Monte Carlo evaluation of integrals and integral equations which aims to introduce the notations and to make the work self-contained is given in the next section.

2. Monte Carlo evaluation of integrals and integral equations

The expectation value E_ψ of a random variable ψ which takes values $\psi(Q)$ with probability density $p_\psi(Q)$ is given by the integral $E_\psi = \int dQ p_\psi(Q)\psi(Q)$ where Q is a

multi-dimensional point. The simplest Monte Carlo method evaluates E_ψ by performing N independent realizations of the probability density p_ψ. Generated are N points $Q_1, \ldots, Q_N$, called sampling points for the random variable ψ. The sample mean η estimates the expectation value E_ψ

$$E_\psi \simeq \eta = \frac{1}{N}\sum_{i=1}^{N}\psi(Q_i) \qquad \mathcal{P}\{|E_\psi - \eta| \leq \frac{3\sigma_\psi}{\sqrt{N}}\} \simeq 0.997 \tag{2.1}$$

with a precision which depends on the number of independent realizations N and the variance of σ_ψ of the random variable. According to the "rule of the three sigma": the probability $\mathcal{P}$ for η to be inside the interval $3\sigma_\psi/\sqrt{N}$ around E_ψ is very high (0.997).

The concept of the Monte Carlo approach for evaluation integrals is to present given integral as an expectation value:

$$I = \int f(Q)dy = \int p(Q)\frac{f(Q)}{p(Q)},\ p(Q) \geq 0,\ \int p(Q)dy = 1 \tag{2.2}$$

of the random variable $\psi = f/p$. The probability density function p can be arbitrary, but admissible for f: $p \neq 0$ if $f \neq 0$. Different random variables can be introduced, depending on the choice of p. All of them have the same expectation value I but different variance and higher moments. It can be shown that the lowest variance is obtained if p is chosen to be proportional to $|f|$.

The idea is generalized for evaluation of integral equations. Consider a Fredholm integral equation of the second kind with a kernel K and a free term f_0

$$f(Q) = \int dQ' f(Q')K(Q', Q) + f_0(Q) \tag{2.3}$$

The solution is expanded into a Neumann series by iteratively replacing the equation into itself. The terms of the series are consecutive iterations of the kernel on the initial condition. Each term in the series is a multiple integral of the type of (2.2) and can be evaluated by a Monte Carlo method. Now assume that we are not interested in the solution itself, but in the inner product of f with a given function A:

$$\langle A \rangle = (A, f) = \int dQ A(Q) f(Q)$$

Useful in this case becomes the adjoint to (2.3) equation:

$$g(Q') = \int dQ K(Q', Q)g(Q) + A(Q') \tag{2.4}$$

If (2.3) is multiplied by g, (2.4) is multiplied by f, the two equations are integrated and compared, it is obtained that $(A, f) = (f_0, g)$:

$$\langle A\rangle = \int dQA(Q)f(Q) = \int dQ' f_0(Q')g(Q') = \sum_i \langle A\rangle_i \tag{2.5}$$

The last equality is obtained with the help of the Neumann expansion of (2.4).

$$g(Q') = A(Q') + \sum_{n=1}^{\infty} K^n(Q', Q))A(Q);\ \ K^n(Q', Q) = \int dQ_1 K(Q', Q_1)K^{n-1}(Q_1, Q) \tag{2.6}$$

This gives rise to the series expansion $\langle A\rangle = \sum_i \langle A\rangle_i$. For example, the integrand of $\langle A\rangle_2$ is a product of f_0 with two iterations of the kernel:

$$\langle A\rangle_2 = \int dQ' dQ_1 dQ_2 P_0(Q')P(Q', Q_1)P(Q_1, Q_2)\frac{f_0(Q')K(Q', Q_1)K(Q_1, Q_2)A(Q_2)}{P_0(Q')P(Q', Q_1)P(Q_1, Q_2)} \tag{2.7}$$

$\langle A\rangle_3$ with three iterations, etc. Actually (2.7) has been augmented with the help of two probabilities P_0 and P in order to resemble (2.2). These probabilities are used to build the so called numerical trajectories:

- $P_0(Q')$ selects the initial point Q' of the trajectory. P_0 must take nonzero values in the points where f_0 is different from zero.
- The transition probability $P(Q', Q)$ uses the value of the selected on the previous step point Q' to select the next trajectory point Q. In order P to be a probability, it is required that $\int dQP(Q', Q) = 1 \forall Q'$. Furthermore P must be different from zero where K is nonzero.

The random variable in $\langle A\rangle_2$ is a product of factors $\frac{f_0}{P_0}$, $\frac{K}{P}$ evaluated at each selected point in the sequence $Q_0 \to Q_1 \to Q_2$. The latter is obtained by application of the probabilities $P_0 \to P \to P$. The mean value of N realizations of the r.v., calculated over the trajectories $(Q' \to Q_1 \to Q_2)_i$, $i = 1, \ldots, N$, evaluates $\langle A\rangle_2$. Here we used the iterative character of the multiple integral (2.7) to introduce a consecutive procedure for construction of the trajectory. In this way the trajectory can be continued by consecutive applications of P. Such trajectory can be used to evaluate simultaneously all terms (2.5), i.e. to evaluate directly $\langle A\rangle$. Our goal is to find a stochastic model where the concrete choice of P_0 and P optimizes the computation of $\langle A\rangle$. We continue with an analysis of the Boltzmann equation in the spirit of the above Monte Carlo theory.

3. The Boltzmann equation with initial and boundary conditions

3.1. The transport model

The time-dependent Boltzmann equation formulated in the complete six-dimensional phase space $\mathcal{T}$, composed by the wave vector $\mathbf{k}$ and position $\mathbf{r}$, is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}(\mathbf{k})\nabla_r + \mathbf{F}(\mathbf{r})\nabla_k\right) f(\mathbf{k},\mathbf{r},t) = \int d\mathbf{k}'\, S(\mathbf{k}',\mathbf{k},\mathbf{r}) f(\mathbf{k}',\mathbf{r},t) - \lambda(\mathbf{k},\mathbf{r}) f(\mathbf{k},\mathbf{r},t) \tag{3.1}$$

The function f gives the distribution of the semiconductor carriers in the phase space: Carriers are accelerated over Newton's trajectories by the electric field $\mathbf{E}$ and are scattered between them by the existing semiconductor lattice imperfections. The scattering is assumed local in position and time. Only the wave vector is changed instantaneously as described by the scattering rate $S(\mathbf{k}',\mathbf{k},\mathbf{r})$ giving the frequency for scattering from the before-scattering state $\mathbf{k}'$ to the after-scattering state $\mathbf{k}$. λ is the out-scattering rate:

$$\lambda(\mathbf{k},\mathbf{r}) = \int d\mathbf{k}' S(\mathbf{k},\mathbf{k}',\mathbf{r}) \qquad \mathbf{F}(\mathbf{r}) = q\mathbf{E}(\mathbf{r})/\hbar \tag{3.2}$$

q is the carrier charge and $\mathbf{v}$ the velocity. The force $\mathbf{F} = \mathbf{F}(\mathbf{r})$ is an explicit function of the position and does not depend on the solution f, so that carriers have negligible contribution to the build-in potential. This corresponding to a frozen field model assumption will be reconsidered later in this work. Both S and $\mathbf{F}$ can depend on the time which is not explicitly written, but is assumed in what follows.

In the general case of device transport the Boltzmann equation (3.1) is provided by both initial, and boundary conditions. The specified at time $t = 0$ initial condition f_i is assumed zero outside the device domain D, while the boundary condition f_b is defined at times $t > 0$ on the device boundaries ∂D. f_b is zero on the reflecting parts of the boundaries, and is usually assumed to be the equilibrium distribution function at the highly doped contacts where carriers flow in and out of the device.

The first term Sf on the right hand side accounts for the events of scattering at time t from all phase space points $\mathbf{k}', \mathbf{r}$ into the phase space point of interest $\mathbf{k}, \mathbf{r}$, so it is called in-scattering term, while the second term λf, called out-scattering term, accounts for the opposite events of scattering out of $\mathbf{k}, \mathbf{r}$.

The mean value $\langle A \rangle(t)$ of a generic physical quantity A are obtained with the help

of the inner product of A with the distribution function f:

$$\langle A\rangle = \int d\mathbf{k} \int_D d\mathbf{r} A(\mathbf{k},\mathbf{r}) f(\mathbf{k},\mathbf{r},t) = \int d\mathbf{k} \int_D d\mathbf{r} \int_0^\infty dt' A(\mathbf{k},\mathbf{r})\delta(t-t') f(\mathbf{k},\mathbf{r},t') \tag{3.3}$$

The reason for the extension with the time integral will become clear in what follows. Here f is assumed normalized to the total number $N(t) = \langle\theta_D\rangle$ of carriers in D (θ_D is explained below). In this way, in a homogeneous semiconductor of volume V_D, the occupation number of a state $\mathbf{k}$ is $4\pi^3 f(\mathbf{k})/V_D$. The averaged, or mean value A per carrier is then given by the ratio $\langle A\rangle/\langle\theta_D\rangle$. The mean value in given domain $\Omega \in D$ is obtained with the help of the domain indicator θ_Ω, which is unity if the phase space point belongs to Ω and is zero otherwise. For example $N(t,\Omega) = (\theta_\Omega, f)$ is the actual number of carriers in Ω.

3.2. Trajectories

The characteristics of the Liouville operator on the left hand side of (3.1) are determined by the Newton equations, which describe the free carrier dynamics:

$$\frac{d\mathbf{k}}{dt} = \mathbf{F}(\mathbf{r}) \qquad \frac{d\mathbf{r}}{dt} = \mathbf{v}(\mathbf{k}) \tag{3.4}$$

Note that F may depend on the time which is not explicitly written. A solution of (3.4), called trajectory, is uniquely determined by specifying an initialization phase space point and time. A trajectory initialized by $\mathbf{k},\mathbf{r},t$ is denoted as follows:

$$\mathbf{K}(t';\mathbf{k},\mathbf{r},t) = \mathbf{k} - \int_{t'}^{t} \mathbf{F}(\mathbf{R}(y))\mathrm{d}y \qquad \mathbf{R}(t';\mathbf{k},\mathbf{r},t) = \mathbf{r} - \int_{t'}^{t} \mathbf{v}(\mathbf{K}(y))\mathrm{d}y \tag{3.5}$$

The compact notation $\mathbf{K}(t'),\mathbf{R}(t')$ will be used; when needed to avoid ambiguity the initialization will be specified explicitly. t' parameterizes the trajectory "backward" in time if $0 < t' < t$ and "forward" in time if $t' > t$. Using (3.5), we can write for a trajectory segment between two given times $t_2 < t_1$:

$$\begin{aligned}
\mathbf{K}(t_e) &= \mathbf{K}(t_2) + \int_{t_2}^{t_e} \mathbf{F}(\mathbf{R}(y))\mathrm{d}y = \mathbf{K}(t_1) - \int_{t_e}^{t_1} \mathbf{F}(\mathbf{R}(y))\mathrm{d}y \\
\mathbf{R}(t_e) &= \mathbf{R}(t_2) + \int_{t_2}^{t_e} \mathbf{v}(\mathbf{K}(y))\mathrm{d}y = \mathbf{R}(t_1) - \int_{t_e}^{t_1} \mathbf{v}(\mathbf{K}(y))\mathrm{d}y
\end{aligned} \tag{3.6}$$

Since the uniqueness of the solution, the above relations show two possible ways for initialization: using $\mathbf{K}(t_2),\mathbf{R}(t_2),t_2$, $t_e > t_2$ or using $\mathbf{K}(t_1),\mathbf{R}(t_1),t_1$, $t_e < t_1$. At

evolution time t_e both ways of initialization provide the same point $\mathbf{K}(t_e), \mathbf{R}(t_e)$. We refer to the first way as to forward initialization and to the second as to backward initialization.

According the Liouville theorem, the trajectories conserve the phase space volume: $d\mathbf{k}d\mathbf{r} = d\mathbf{K}(t')d\mathbf{R}(t')$. In particular:

$$\int d\mathbf{k}d\mathbf{r}\phi(\mathbf{k},\mathbf{r},\mathbf{K}(t'),\mathbf{R}(t')) = \int d\mathbf{k}'d\mathbf{r}'\phi(\mathbf{K}'(t),\mathbf{R}'(t),\mathbf{k}',\mathbf{r}') \tag{3.7}$$

Here ϕ is a given function, $\mathbf{K}(\tau), \mathbf{R}(\tau)$ is a backward trajectory initialized by $\mathbf{k}, \mathbf{r}$ at t, while $\mathbf{K}'(\tau), \mathbf{R}'(\tau)$ is a forward trajectory initialized by $\mathbf{k}' = \mathbf{K}(t'), \mathbf{r}' = \mathbf{R}(t')$. Over a trajectory the Liouville operator becomes a total time derivative. Thus, provided that there are no scattering events, the value of f is conserved over trajectories. In order to use the rules for Monte Carlo integration we first need the integral form of (3.1).

3.3. Integral form

We consider equation (3.1), written for a backward trajectory (3.5) initialized by $\mathbf{k}, \mathbf{r}, t$:

$$\left(\frac{d}{dt'} + \lambda(\mathbf{K}(t'),\mathbf{R}(t'))\right) f(\mathbf{K}(t'),\mathbf{R}(t'),t') = \int d\mathbf{k}' S(\mathbf{k}',\mathbf{K}(t'),\mathbf{R}(t')) f(\mathbf{k}',\mathbf{R}(t'),t') \tag{3.8}$$

After a multiplication by the integrating factor $e^{-\int_{t'}^{t}\lambda(\mathbf{K}(y),\mathbf{R}(y))dy}$ the left hand side becomes a total time derivative:

$$\frac{d}{dt'} e^{-\int_{t'}^{t}\lambda(\mathbf{K}(y),\mathbf{R}(y))dy} f(\mathbf{K}(t'),\mathbf{R}(t'),t') =$$

$$\int d\mathbf{k}'\, S(\mathbf{k}',\mathbf{K}(t'),\mathbf{R}(t')) f(\mathbf{k}',\mathbf{R}(t'),t') e^{-\int_{t'}^{t}\lambda(\mathbf{K}(y),\mathbf{R}(y))dy}$$

The obtained equation can be integrated in the limits (t_0, t). According to equation (3.5), $\mathbf{K}(t' = t) = \mathbf{k}$ and $\mathbf{R}(t' = t) = \mathbf{r}$. The time t_0 can be appropriately chosen to provide an initial or boundary point, where the value of f is known. That is t_0 is 0 if at $t' = 0$ all points of the trajectory segment between $t' = t$ and $t' = 0$ belong to D. If the trajectory (evolving backwards in time) crosses the boundary ∂D at time $t_b > 0$, then $t_0 = t_b$. Note that in stationary conditions the interval $\tau_b = t - t_b$ depends only on the phase space point $\mathbf{k}, \mathbf{r}$. It is convenient to utilize the indicator θ_D of the device domain D, which takes values 1 if the real space component of the trajectory belongs to D and 0 otherwise. This allows to set the bottom time limit on the right

to 0. Moreover, since the domains of definition of f_i and f_b are complementary, the integral form can be stated in a way which includes both, the initial and the boundary terms:

$$f(\mathbf{k},\mathbf{r},t)=\int_0^t dt'\theta_D(\mathbf{R}(t'))\int d\mathbf{k}' f(\mathbf{k}',\mathbf{R}(t'),t')S(\mathbf{k}',\mathbf{K}(t'),\mathbf{R}(t'))e^{-\int_{t'}^{t}\lambda(\mathbf{K}(y),\mathbf{R}(y))dy}$$
$$+e^{-\int_0^t\lambda(\mathbf{K}(y),\mathbf{R}(y))dy}f_i(\mathbf{K}(0),\mathbf{R}(0))+e^{-\int_{t_b}^{t}\lambda(\mathbf{K}(y),\mathbf{R}(y))dy}f_b(\mathbf{K}(t_b),\mathbf{R}(t_b),t_b) \quad (3.9)$$

The obtained equation has a transparent physical meaning. The term $e^{-\int_{t'}^{t}\lambda(\mathbf{K}(y),\mathbf{R}(y))dy}$ is the probability for a carrier starting at time t' from the phase space point $\mathbf{K}(t'),\mathbf{R}(t')$ to remain over the trajectory until reaching $\mathbf{k},\mathbf{r}$ at time t, provided that the out-scattering rate is λ. Hence, the value of $f(\mathbf{k},\mathbf{r},t)$ is formed by two contributions: The first one is the fraction of the initial (boundary) value which survives over the trajectory despite the scattering events. The second contribution is accumulated during the evolution of the system by the following events taking place in the trajectory part before the point $\mathbf{k},\mathbf{r},t$: At time t' electrons from the whole momentum space, but having real space coordinate equal to $\mathbf{R}(t')$, are scattered with rate S in the proper momentum state $\mathbf{K}(t')$ (this ensures that they will follow the trajectory which passes through $\mathbf{k},\mathbf{r}$ at time t). The number of such events is proportional to the electron distribution $f(\mathbf{k}',\mathbf{R}(t'),t'))$. The exponent gives the probability the properly scattered electrons to remain on the trajectory until time t is reached.

An application of the Monte Carlo theory to equation (IF) gives rise to backward evolution algorithms. The natural forward evolution is introduced by the adjoint equation.

3.4. The adjoint equation

The function $f(Q) = f(\mathbf{k},\mathbf{r},t)$ in (3.9) is defined on a seven dimensional point Q. However, the integral on the right hand side is only four dimensional, which shows that the kernel is degenerate. To obtain a formal analogy with (2.3) we first augment the right hand side with an integration on $\mathbf{r}'$. Furthermore the upper limit of the time integral is augmented to ∞ and the kernel is multiplied by a term, which aims to keep

the value of the multidimensional integral unchanged:

$$f(\mathbf{k},\mathbf{r},t) = \int_0^\infty dt' \int d\mathbf{k}' \int d\mathbf{r}' f(\mathbf{k}',\mathbf{R}(t'),t') S(\mathbf{k}',\mathbf{K}(t'),\mathbf{r}') e^{-\int_{t'}^{t} \lambda(\mathbf{K}(y),\mathbf{R}(y))dy} \times$$

$$\theta_D(\mathbf{r}')\delta(\mathbf{r}' - \mathbf{R}(t'))\theta(t-t') + f_0(\mathbf{k},\mathbf{r},t) \tag{3.10}$$

Note that the free term in (3.9), shortly denoted by f_0 is a function of $\mathbf{k},\mathbf{r},t$ only, since the later determine t_0 and $\mathbf{K}(t_0),\mathbf{R}(t_0)$. Here $\theta(t-t')$ is the Heaviside function. According to (2.4), the adjoint equation with a given free term g_0 has the same kernel, but the integration is carried over the unprimed variables:

$$g(\mathbf{k}',\mathbf{r}',t') = \int_0^\infty dt \int d\mathbf{k} \int d\mathbf{r} g(\mathbf{k},\mathbf{r},t) S(\mathbf{k}',\mathbf{K}(t'),\mathbf{r}') e^{-\int_{t'}^{t} \lambda(\mathbf{K}(y),\mathbf{R}(y))dy} \times$$

$$\theta_D(\mathbf{r}')\delta(\mathbf{r}' - \mathbf{R}(t'))\theta(t-t') + g_0(\mathbf{k}',\mathbf{r}',t') \tag{3.11}$$

It is now convenient to apply the Liouville theorem (3.7). The integration variables $\mathbf{k},\mathbf{r}$ can be changed to $\mathbf{K}(t'),\mathbf{R}(t')$ which are further denoted as $\mathbf{k}^a = \mathbf{K}(t'), r'' = \mathbf{R}(t')$. The latter at time t' provide a forward initialization of the trajectory. The equation reads:

$$g(\mathbf{k}',\mathbf{r}',t') = \int_0^\infty dt \int d\mathbf{k}^a \int d\mathbf{r}'' g(\mathbf{K}(t),\mathbf{R}(t),t) S(\mathbf{k}',\mathbf{k}^a,\mathbf{r}') e^{-\int_{t'}^{t} \lambda(\mathbf{K}(y),\mathbf{R}(y))dy} \times$$

$$\theta_D(\mathbf{r}')\delta(\mathbf{r}' - \mathbf{r}'')\theta(t-t') + g_0(\mathbf{k}',\mathbf{r}',t')$$

Now it is possible to account for the δ function and the Heaviside function θ. As t has been used to denote the time for evaluation of (3.3), it is convenient to change the time integration variable to τ.

$$g(\mathbf{k}',\mathbf{r}',t') = \tag{3.12}$$

$$\int_{t'}^\infty d\tau \int d\mathbf{k}^a \theta_D(\mathbf{r}') S(\mathbf{k}',\mathbf{k}^a,\mathbf{r}') e^{-\int_{t'}^{\tau} \lambda(\mathbf{K}(y),\mathbf{R}(y))dy} g(\mathbf{K}(\tau),\mathbf{R}(\tau),\tau) + g_0(\mathbf{k}',\mathbf{r}',t')$$

A comparison with (2.4), (2.5) and (3.3) shows that the free term g_0 must be chosen as:

$$g_0(\mathbf{k}',\mathbf{r}',t') = A(\mathbf{k}',\mathbf{r}')\delta(t-t')$$

The Neumann expansion (2.6) of g is obtained by a recursive replacement of (3.12) into itself. For example the second term $g^{(2)}$ is:

$$g^{(2)}(\mathbf{k}', \mathbf{r}', t') = \int_{t'}^{t} \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_1^a \int \mathrm{d}\mathbf{k}_2^a S(\mathbf{k}', \mathbf{k}_1^a, \mathbf{r}')\theta_D(\mathbf{r}') \\ \mathrm{e}^{-\int_{t'}^{t_1} \lambda(\mathbf{K}_1(y), \mathbf{R}_1(y))\mathrm{d}y} S(\mathbf{K}_1(t_1), \mathbf{k}_2^a, \mathbf{R}_1(t_1))\theta_D(\mathbf{R}_1(t_1)) \\ \mathrm{e}^{-\int_{t_1}^{t} \lambda(\mathbf{K}_2(y), \mathbf{R}_2(y))\mathrm{d}y} A(\mathbf{K}_2(t), \mathbf{R}_2(t)) \tag{3.13}$$

where the trajectory $\mathbf{K}_1, \mathbf{R}_1$ is initialized by $\mathbf{k}_1^a, \mathbf{r}', t'$ and $\mathbf{K}_2, \mathbf{R}_2$ is initialized by $\mathbf{k}_2^a, \mathbf{R}_1(t_1), t_1$.

The physical meaning of equation (3.12) can be displayed by choosing A to be the indicator of a phase space sub-domain $\Omega \in D : A = \theta_\Omega(\mathbf{k}, \mathbf{r})$. It will be shown that in this case the quantity

$$G(\theta_\Omega, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0) = \int_0^\infty dt' \mathrm{e}^{-\int_0^{t'} \lambda(\mathbf{K}_\alpha(y), \mathbf{R}_\alpha(y))\mathrm{d}y} g(\mathbf{K}_\alpha(t'), \mathbf{R}_\alpha(t')) \tag{3.14}$$

is the probability of a carrier, initially located at $\mathbf{k}_\alpha, \mathbf{r}_\alpha, 0$ (these values initialize the trajectory $\mathbf{K}_\alpha(y), \mathbf{R}_\alpha(y)$) to appear in Ω at time t without leaving D during the evolution. In particular consider the contribution of the term (3.13) to (3.14) in the case $A = \theta_\Omega$:

$$G^{(2)} = \int_0^\infty dt' \int_{t'}^{t} \mathrm{d}t_1 \int \mathrm{d}\mathbf{k}_1^a \int \mathrm{d}\mathbf{k}_2^a \mathrm{e}^{-\int_0^{t'} \lambda(\mathbf{K}_\alpha(y), \mathbf{R}_\alpha(y))\mathrm{d}y} S(\mathbf{K}_\alpha(t'), \mathbf{k}_1^a, \mathbf{R}_\alpha(t'))\theta_D(\mathbf{R}_\alpha(t')) \\ \times \mathrm{e}^{-\int_{t'}^{t_1} \lambda(\mathbf{K}_1(y), \mathbf{R}_1(y))\mathrm{d}y} S(\mathbf{K}_1(t_1), \mathbf{k}_2^a, \mathbf{R}_1(t_1))\theta_D(\mathbf{R}_1(t_1)) \\ \times \mathrm{e}^{-\int_{t_1}^{t} \lambda(\mathbf{K}_2(y), \mathbf{R}_2(y))\mathrm{d}y} \theta_\Omega(\mathbf{K}_2(t), \mathbf{R}_2(t)) \tag{3.15}$$

We recall two classical transport probability densities, p_λ and p_S. $p_t dt$ is the probability for a carrier at $\mathbf{k}, \mathbf{r}, t_0$ to experience a free flight in the time interval (t_0, t) and to be scattered during $(t, t + dt)$ provided that the out-scattering rate is λ:

$$p_t(t; \mathbf{k}, \mathbf{r}, t_0)dt = \mathrm{e}^{-\int_{t_0}^{t} \lambda(\mathbf{K}(y), \mathbf{R}(y))\mathrm{d}y} \lambda(\mathbf{K}(t), \mathbf{R}(t))dt$$

The proof of the normalization of p_t is straightforward [1].

The quantity

$$p_S(\mathbf{k},\mathbf{k}',\mathbf{r})d\mathbf{k}' = \frac{S(\mathbf{k},\mathbf{k}',\mathbf{r})}{\lambda(\mathbf{k},\mathbf{r})}d\mathbf{k}'$$

is the probability for scattering per unit time from a phase space point $\mathbf{k},\mathbf{r}$ in the volume $d\mathbf{k}'$ around $\mathbf{k}',\mathbf{r}$. Only the wave vector component changes, showing that the scattering is regarded instantaneous in time and local in real space. The function p_S is normalized to unity as λ is defined from equation (3.2).

The densities p_t and p_S appear explicitly in equation (3.15) if the latter is augmented by a multiplication of each exponent with λ and a division of the next S by the same λ. For convenience the time variables are altered according to $t' \to t_1$, $t_1 \to t_2$,

$$G^{(2)} = \int_0^\infty dt_1 \int_{t_1}^\infty \mathrm{d}t_2 \int \mathrm{d}\mathbf{k}_1^a \int \mathrm{d}\mathbf{k}_2^a \tag{3.16}$$

$$\left\{ \mathrm{e}^{-\int_0^{t_1} \lambda(\mathbf{K}_\alpha(y),\mathbf{R}_\alpha(y))\mathrm{d}y} \lambda(\mathbf{K}_\alpha(t_1),\mathbf{R}_\alpha(t_1)) \right\} \left\{ \frac{S(\mathbf{K}_\alpha(t_1),\mathbf{k}_1^a,\mathbf{R}_\alpha(t_1))}{\lambda(\mathbf{K}_\alpha(t_1),\mathbf{R}_\alpha(t_1))} \right\} \theta_D(\mathbf{R}_\alpha(t_1))$$

$$\left\{ \mathrm{e}^{-\int_{t_1}^{t_2} \lambda(\mathbf{K}_1(y),\mathbf{R}_1(y))\mathrm{d}y} \lambda(\mathbf{K}_1(t_2),\mathbf{R}_1(t_2)) \right\} \left\{ \frac{S(\mathbf{K}_1(t_2),\mathbf{k}_2^a,\mathbf{R}_1(t_2))}{\lambda(\mathbf{K}_1(t_2),\mathbf{R}_1(t_2))} \right\} \theta_D(\mathbf{R}_1(t_2))$$

$$\mathrm{e}^{-\int_{t_2}^{t} \lambda(\mathbf{K}_2(y),\mathbf{R}_2(y))\mathrm{d}y} \theta_\Omega(\mathbf{K}_2(t),\mathbf{R}_2(t))\theta(t-t_2)$$

The obtained probability densities are enclosed in curly brackets. Note that the integral on t_2 is augmented to infinity, which is compensated by $\theta(t-t_2)$. $G^{(2)}$ is entirely decomposed into a product of conditional probability densities $p_t p_S p_t p_S$ which generate the natural process of evolution of the Boltzmann carrier system. Indeed we can recognize elementary events which compose the following process: The initial carrier drifts over the trajectory $\mathbf{K}_\alpha, \mathbf{R}_\alpha$ until time t_1 when a scattering event occurs. The carrier coordinates $\mathbf{K}_\alpha(t_1), \mathbf{R}_\alpha(t_1)$ at the end of the free flight are used in p_S to generate the after scattering state: The carrier scatters from $\mathbf{K}_\alpha(t_1), \mathbf{R}_\alpha(t_1), t_1$ to

[1] Here we assume the existence of both, trajectories and phonon scattering outside the device domain D. Alternatively we may assume that trajectories do not exist outside D, which is equivalent to assume λ infinity. In this case the time to the boundary becomes an upper limit of the normalization integral. The results of this section remain true after minor modifications

$\mathbf{k}_1^a, \mathbf{R}_\alpha(t_1), t_1$ The next p_t gives rise to a free flight in the time interval (t_1, t_2) over a trajectory $\mathbf{K}_1, \mathbf{R}_1$ initialized by $\mathbf{k}_1^a, \mathbf{R}_\alpha(t_1) = (\mathbf{r}_1, t_1)$. At time t_2 there is another scattering from the end point of the trajectory $\mathbf{K}_1(t_2), \mathbf{R}_1(t_2)$ to $\mathbf{k}_2^a, \mathbf{R}_1(t_2) = \mathbf{r}_2$. The latter initializes the trajectory $\mathbf{K}_2, \mathbf{R}_2$.

The last exponent is the probability for drift without scattering over $\mathbf{K}_2, \mathbf{R}_2$ until time t is reached. Indeed the equality

$$\mathrm{e}^{-\int_{t_2}^{t} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y} = \int_{t_2}^{\infty} dt_3 \mathrm{e}^{-\int_{t_2}^{t_3} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y} \lambda(\mathbf{K}(t_3), \mathbf{R}(t_3))\theta(t_3 - t)dt \quad (3.17)$$

shows that the exponent unifies all events having scattering time greater than t. Finally θ_Ω in 3.16 accounts only these processes, where the carrier is in Ω at time t.

The integrals on time and wave vectors take into account all possible elementary events comprising three free flights and two scattering events. The domain indicators reject all events where trajectories leave the device domain D before time t. The θ function takes care for the proper time ordering of the transport process. It discards all events which do not obey the condition $t_1 \leq t_2 \leq t$. Thus counted are only processes where the carrier: (i) is a subject of a two scattering evolution process; (ii) remains inside D during the evolution; (iii) resides in Ω at time t.

Equation (3.17) can be used to unify the terms with $i \leq 2$ into one expression.

$$\begin{aligned}\sum_{i=0}^{2} G^{(i)} = \int_0^\infty dt_1 \int_{t_1}^\infty \mathrm{d}t_2 \int_{t_2}^\infty \mathrm{d}t_3 \int \mathrm{d}\mathbf{k}_1^a \int \mathrm{d}\mathbf{k}_2^a \int \mathrm{d}\mathbf{k}_3^a \times \quad (3.18)\\ p_t(t_1; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0) p_S(\mathbf{K}_\alpha(t_1), \mathbf{k}_1^a, \mathbf{R}_\alpha(t_1)) \times \\ p_t(t_2; \mathbf{k}_1^a, \mathbf{R}_\alpha(t_1), t_1) p_S(\mathbf{K}_1(t_2), \mathbf{k}_2^a, \mathbf{R}_1(t_2)) \times \\ p_t(t_3; \mathbf{k}_2^a, \mathbf{R}_1(t_2), t_2) p_S(\mathbf{K}_2(t_3), \mathbf{k}_3^a, \mathbf{R}_2(t_3)) \times \\ \Big(\theta_\Omega(\mathbf{K}_\alpha(t), \mathbf{R}_\alpha(t))\theta(t_1 - t) + \theta(t - t_1)\theta_D(\mathbf{R}_\alpha(t_1))\theta_\Omega(\mathbf{K}_1(t), \mathbf{R}_1(t))\theta(t_2 - t) \\ + \theta(t - t_2)\theta_D(\mathbf{R}_\alpha(t_1))\theta_D(\mathbf{R}_1(t_2))\theta_\Omega(\mathbf{K}_2(t), \mathbf{R}_2(t))\theta(t_3 - t)\Big)\end{aligned}$$

This result can be generalized for the rest of the series, giving rise to an infinite integral and an infinite sum of theta functions which present the random variable Ψ associated

to the Boltzmann transport process.

$$G = \lim_{m\to\infty} \prod_{l=0}^{m} \int_{t_l}^{\infty} dt_{l+1} \int \mathrm{d}\mathbf{k}^a_{l+1}\; p_t(t_{l+1}; \mathbf{k}^a_l, \mathbf{R}_{l-1}(t_l), t_l) p_S(\mathbf{K}_l(t_{l+1}), \mathbf{k}^a_{l+1}, \mathbf{R}_l(t_{l+1})) \ldots$$

$$\times \sum_{i=0}^{m} \theta(t - t_i) \left(\prod_{j=0}^{i-1} \theta_D(\mathbf{R}_j(t_{j+1})) \right) \theta_\Omega(\mathbf{K}_i(t), \mathbf{R}_i(t)) \theta(t_{i+1} - t) \tag{3.19}$$

The following convention justifies the recursive notations:

$$t_0 = 0; \quad \mathbf{k}^a_0 = \mathbf{k}_\alpha; \quad \mathbf{R}_{-1} = \mathbf{r}_\alpha; \quad \mathbf{K}_0 = \mathbf{K}_\alpha; \quad \mathbf{R}_0 = \mathbf{R}_\alpha; \quad \prod_{j=0}^{-1} \theta_D = 1$$

Now consider the special case where the carrier evolves in the whole phase space $\mathcal{T}$, i.e. $\Omega = D = \mathcal{T}$. Then $\theta_\Omega = \theta_D = 1$ and the random variable in (3.19) reduces to

$$\psi = \sum_{i=0}^{m-1} \theta(t - t_i) \theta(t_{i+1} - t); \qquad 0 < t_1 < t_2 \ldots \tag{3.20}$$

The time axis is decomposed into intervals by the consecutive times of (3.19). For any time t and for any setting of the consecutive times t_i there is a certain interval $t_j < t < t_{j+1}$ which contains t. Then $\theta(t - t_j)\theta(t_{j+1} - t) = 1$, while the remaining terms in the sum of (3.20) become zero. The theta functions in time decompose the evolution into mutually complementary events whose probabilities sum up to unity, $\psi = 1$. In this case the value of (3.19) is easily evaluated:

$$G(\theta_\mathcal{T}, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0) = 1 \qquad \forall \quad t, \quad \mathbf{k}_\alpha, \; \mathbf{r}_\alpha \tag{3.21}$$

This property is called conservation of the mass, which is another way of saying that at any time t the carrier must be somewhere in the phase space. It should be noted that the traditional way of obtaining the conservation of the mass, equation (3.21), is as follows. From (2.5), (3.3) and (3.21) and by setting $\theta_\Omega = \theta_\mathcal{T} = 1$ it follows that:

$$G(\theta_\mathcal{T}, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0) = \int d\mathbf{k} \int d\mathbf{r} f(\mathbf{k}, \mathbf{r}, t)$$

By integrating (3.1) in the phase space $\mathcal{T}$, the left hand side becomes the total time derivative of G while the right hand side becomes 0. Hence

$$G(\theta_\mathcal{T}, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0) = G(\theta_\mathcal{T}, 0; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0) = 1$$

In the general case $\Omega \in D \neq \mathcal{T}$ the domain indicators discard some of the events of the evolution, which are counted in (3.20). The random variable ψ in (3.19) is now not fixed to 1, but has the 0 as an alternative value. Thus ψ counts only elementary events where the domain indicators are unity and discards the rest of the events which accomplish $G(\theta_\Omega, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0)$ to $G(\theta_\mathcal{T}, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0)$. It follows that $G(\theta_\Omega, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0)$ is the probability of a carrier, initially located at $\mathbf{k}_\alpha, \mathbf{r}_\alpha, 0$ to appear in Ω at time t without leaving D during the evolution. In what follows we call condition I the restriction that the trajectory must belong to D. With this experience we can generalize (3.19) to $G(A, t; \mathbf{k}_\alpha, \mathbf{r}_\alpha, 0)$ of any generic physical quantity A by simply replacing θ_Ω by A.

3.5. Initial and boundary conditions

According to equation (2.5), the mean value of the physical quantity A at time t is obtained by the inner product of the free term of (3.10) with the solution of the adjoint equation (3.12):

$$\langle A \rangle(t) = \int dQ' f_0(Q') g(Q') = \int d\mathbf{k}' \int d\mathbf{r}' \int_0^\infty dt' f_0(\mathbf{k}', \mathbf{r}', t') g(\mathbf{k}', \mathbf{r}', t') \quad (3.22)$$

The Neumann expansion of (3.12) inserted in (3.22) gives the desired series for a Monte Carlo approach. Since f_0 accounts for contributions of both, initial and boundary conditions, (3.22) must be evaluated for each contribution separately.

3.5.1. Initial condition

$$\langle A \rangle_i(t) = \int d\mathbf{k}' \int d\mathbf{r}' \int_0^\infty dt' f_i(\mathbf{K}(0), \mathbf{R}(0)) \mathrm{e}^{-\int_0^{t'} \lambda(\mathbf{K}(y), \mathbf{R}(y)) \mathrm{d}y} g(\mathbf{k}', \mathbf{r}', t')$$

We recall that the backward trajectory $(\mathbf{K}(y), \mathbf{R}(y))$ is initialized by $\mathbf{k}', \mathbf{r}', t'$. Using the same arguments as for (3.7) we change to forward initialization. By denoting $\mathbf{K}(0), \mathbf{R}(0)$ as $\mathbf{k}_i, \mathbf{r}_i$ the forward trajectory $\mathbf{K}_i(y), \mathbf{R}_i(y)$ is now initialized by $\mathbf{k}_i, \mathbf{r}_i$ at time 0. Then $\mathbf{k}' = \mathbf{K}_i(t')$ and $\mathbf{r}' = \mathbf{R}_i(t')$ so that

$$\langle A \rangle_i(t) = \int d\mathbf{k}_i \int d\mathbf{r}_i f_i(\mathbf{k}_i, \mathbf{r}_i) \int_0^\infty dt' \mathrm{e}^{-\int_0^{t'} \lambda(\mathbf{K}_i(y), \mathbf{R}_i(y)) \mathrm{d}y} g(\mathbf{K}_i(t'), \mathbf{R}_i(t'), t') \quad (3.23)$$

The interpretation of this expression follows the arguments after (3.15). The function f_i gives the distribution of the initial carriers, i.e. the points which initialize the trajectories at time 0. The rest of the expression can be easily identified as $G(A, t; \mathbf{k}_i, \mathbf{r}_i, 0)$.

Indeed the choice of $f_i(\mathbf{k}_i, \mathbf{r}_i) = \delta(\mathbf{k}_i - \mathbf{k}_\alpha)\delta(\mathbf{r}_i - \mathbf{r}_\alpha)$ along with $A = \theta_\Omega$ transforms (3.23) into (3.14).

3.5.2. Boundary conditions

The treatment of the boundary conditions is not straightforward. The reason is that the former are known on the boundaries only so that the volume integral on $\mathbf{r}'$ in

$$\langle A\rangle_b(t) = \int d\mathbf{k}' \int d\mathbf{r}' \int_0^\infty dt' f_b(\mathbf{K}(t_b), \mathbf{R}(t_b)) \mathrm{e}^{-\int_{t_b}^{t'} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y} g(\mathbf{k}', \mathbf{r}', t') \tag{3.24}$$

must be transformed into a boundary integral. A phase space point $(\mathbf{k}', \mathbf{r}')$ is bijectively mapped onto $(\mathbf{K}(t_b), \mathbf{r}_b = \mathbf{R}(t_b), t_b)$ where $\mathbf{r}_b \in \partial D$ if $t_b > -\infty$. This prompts that the transformation must replace one of the space integrals by a time integral. We consider the subspace Φ_b of points $(\mathbf{k}', \mathbf{r}')$ having finite t_b. Fortunately this is the relevant sub-domain for the integral (3.24), since in the complementary subspace the integrand vanishes. The required time integral is introduced as follows. The domain boundary can be formally defined by a function $B(\mathbf{r}) = 0$. This gives an implicit definition of the boundary time as a root of $B(\mathbf{R}(t'')) = 0$. Consider the equality

$$\delta(B(\mathbf{R}(t''))) = \frac{\delta(t'' - t_b)}{|\nabla_{\mathbf{R}} B(\mathbf{R})_{|\mathbf{R}=\mathbf{R}(t'')} \cdot \mathbf{v}(\mathbf{K}(t''))|} = \frac{\delta(t'' - t_b)}{|\nabla_{\mathbf{R}} B(\mathbf{R})_{|\mathbf{R}=\mathbf{R}(t_b)} |\mathbf{v}_\perp(\mathbf{K}(t_b))} \tag{3.25}$$

Here $\mathbf{v}_\perp$ is the velocity component normal to the domain boundary in the crossing point with the trajectory. The wave vectors $\mathbf{K}(t_b) \in K^+$ comprise the subspace K^+ of these wave vectors corresponding to velocities inwardly directed with respect to the domain D.

The equality can be used to augment (3.24) by a time integral on t'' in the limits $(0, t')$:

$$\begin{aligned}\langle A\rangle_b(t) = & \int d\mathbf{k}' \int d\mathbf{r}' \int_0^\infty dt' \int_0^{t'} dt'' \delta(B(\mathbf{R}(t''))) |\nabla_{\mathbf{R}} B(\mathbf{R})_{|\mathbf{R}=\mathbf{R}(t'')} \cdot \mathbf{v}(\mathbf{K}(t''))| \\ & \times f_b(\mathbf{K}(t''), \mathbf{R}(t''), t'') \mathrm{e}^{-\int_{t''}^{t'} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y} g(\mathbf{k}', \mathbf{r}', t')\end{aligned}$$

The time argument t_b in the second row has been replaced by t'' with the help of $\delta(t'' - t_b)$ in the first row of the expression. The obtained equation can be processed

within the same steps leading to (3.23). The integration variables $d\mathbf{k}', d\mathbf{r}'$ are changed to

$$d\mathbf{k}'d\mathbf{r}' = d\mathbf{K}(t''; \mathbf{k}', \mathbf{r}', t')d\mathbf{R}(t''; \mathbf{k}', \mathbf{r}', t') = d\mathbf{K}(t'')d\mathbf{R}(t'') = d\mathbf{k}''d\mathbf{r}''$$

Now $\mathbf{K}(y), \mathbf{R}(y)$ denote a forward trajectory initialized by $\mathbf{k}'', \mathbf{r}'', t''$ so that $\mathbf{k}' = \mathbf{K}(t')$ and $\mathbf{r}' = \mathbf{R}(t')$. The expression for $\langle A\rangle_b(t)$ becomes:

$$\begin{aligned}\langle A\rangle_b(t) &= \int d\mathbf{k}'' \int d\mathbf{r}'' \int_0^\infty dt' \int_0^{t'} dt'' \delta(B(\mathbf{r}''))|\nabla_{\mathbf{R}} B(\mathbf{R})_{|\mathbf{R}=\mathbf{r}''} \cdot \mathbf{v}(\mathbf{k}'')| \\ &\times f_b(\mathbf{k}'', \mathbf{r}'', t'')\mathrm{e}^{-\int_{t''}^{t'} \lambda(\mathbf{K}(y),\mathbf{R}(y))\mathrm{d}y} g(\mathbf{K}(t'), \mathbf{R}(t'), t') \qquad (3.26)\end{aligned}$$

The $\mathbf{r}''$ integral is transformed with the aid of the equality:

$$\int_D \delta(B(\mathbf{r}''))\phi(\mathbf{r}'')d\mathbf{r}'' = \oint \frac{\phi(\mathbf{r}_b)}{|\nabla_{\mathbf{r}} B(\mathbf{r})_{|\mathbf{r}=\mathbf{r}_b}|} d\sigma(\mathbf{r}_b) \qquad (3.27)$$

where ϕ is a test function. By using (3.27) and (3.25) into (3.26) it is obtained:

$$\begin{aligned}\langle A\rangle_b(t) &= \oint_{\partial D} d\sigma(\mathbf{r}_b)d\mathbf{k}'' \int_{K_+} \int_0^\infty dt' \int_0^{t'} dt'' \, v_\perp(\mathbf{k}'') f_b(\mathbf{k}'', \mathbf{r}_b, t'') \\ &\exp\left(-\int_{t''}^{t'} \lambda(\mathbf{K}(y), \mathbf{R}(y))dy\right) g(\mathbf{K}(t'), \mathbf{R}(t'), t')\end{aligned}$$

For consistence we denote the wave vectors $\mathbf{k}''$ of the boundary states by $\mathbf{k}_b$, the time t'' by t_b and rename the trajectory. Recalling that $\int_0^\infty dt' \int_0^{t'} dt_b = \int_0^\infty dt_b \int_{t_b}^\infty dt'$ we finally obtain:

$$\begin{aligned}\langle A\rangle_b(t) &= \int_0^t dt_b \oint_{\partial D} d\sigma(\mathbf{r}_b)d\mathbf{k}_b \int_{K_+} v_\perp(\mathbf{k}_b) f_b(\mathbf{k}_b, \mathbf{r}_b, t_b) \\ &\times \int_{t_b}^\infty dt' \exp\left(-\int_{t_b}^{t'} \lambda(\mathbf{K}_b(y), \mathbf{R}_b(y))dy\right) g(\mathbf{K}_b(t'), \mathbf{R}_b(t'), t') \qquad (3.28)\end{aligned}$$

The upper bound of the integral on t_b is set to t by the delta function in g_0. The term $v_\perp f_b$ has similar interpretation as f_i: it gives the distribution of the points which initialize the trajectories at time t_b. The rest of the expression can be easily identified

as $G(A, t; \mathbf{k}_b, \mathbf{r}_b, t_b)$. A physical insight to (3.28) provide the quantities:

$$
\begin{aligned}
j_\perp(\mathbf{r}_b, t_b) &= \int_{K_+(\mathbf{r}_b)} d\mathbf{k}_b \, v_\perp(\mathbf{k}_b) f_b(\mathbf{k}_b, \mathbf{r}_b, t_b), \\
\Gamma_D(t_b) &= \oint_{\partial D} j_\perp(\mathbf{r}_b, t_b) d\sigma(\mathbf{r}_b), \\
N_\Gamma(t) &= \int_0^t dt_b \Gamma_D(t_b)
\end{aligned}
\tag{3.29}
$$

They represent the normal component of the incident carrier current density, the total incident carrier current at time t_b and the total number of carriers injected until evolution time t. Then the integral with respect to t_b gives the contribution of all carriers injected at times $t_b \leq t$ to the value of $\langle A \rangle_b(t)$. Assuming $A = \theta_\Omega$, the term of the second row is identified as $G(\theta_\Omega, t; \mathbf{k}_i, \mathbf{r}_i, t_b)$. Usually the boundary conditions f_b are time independent so that the integrals in (3.29) become constants with respect to the time.

4. Monte Carlo approach

4.1. The EMC algorithm for device simulations

We derive the EMC algorithm from the basic rules of the numerical MC procedure. In the previous section we discussed the peculiarities of the kernel and the free term of the Boltzmann equation, which determine the expression for $\langle A \rangle(t)$. In particular for a point-like initial condition $\mathbf{k}_\alpha, \mathbf{r}_\alpha$ the series expansion (2.5) becomes (3.19) (with θ_Ω replaced by the concrete expression for A). The terms of the series, represented by (3.16) and (3.18) have the form of (2.7). This allows to identify a "natural" transition probability P^B as a product of the probabilities $p_t p_S$, incorporated in the Boltzmann kernel. P^B gives rise to the standard free-flight and after-scattering selection scheme utilized in device Monte Carlo simulators. The time of the next scattering event t_r is determined by the equality:

$$
\int_{t_0}^{t_r} p_t(t; \mathbf{k}, \mathbf{r}, t_0) dt = r
$$

where r is a random number evenly distributed in the interval $(0, 1)$. The next step is to select the after scattering state according to p_S. Usually S is comprised by several

kinds scattering mechanisms S_i:

$$S = \sum_i S_i; \qquad p_S = \frac{S}{\lambda} = \sum_i \frac{\lambda_i}{\lambda}\frac{S_i}{\lambda_i} = \sum_i \frac{\lambda_i}{\lambda} p_{S,i}; \qquad \sum_i^{i_r} \frac{\lambda_i}{\lambda} \le r < \sum_i^{i_r+1} \frac{\lambda_i}{\lambda} \tag{4.1}$$

The probability for selection of a given after-scattering state is the product of the conditional probabilities involved in the first equality in (4.1). Thus inequality in (4.1) is used first to select the concrete scattering mechanism i_r, where r is another random number evenly distributed in the interval $(0, 1)$. Then the after-scattering state wave vector is generated from the selected mechanism i_r. As $|\mathbf{k}'|$ is determined by the type of scattering, three random numbers are utilized to select according p_{S,i_r} the angles which establish the $\mathbf{k}'$ as a vector.

We establish a link between the numerical trajectories and the trajectories of the classical semiconductor carriers. A numerical trajectory, constructed with the help of P, resembles a possible real trajectory of a classical carrier. Apart from the requirement to fulfill the condition I, only the final point of the trajectory $\mathbf{k}, \mathbf{r}$, reached at time t, determines the value of the corresponding random variable ψ in (3.19). Indeed, for any concrete trajectory only one term in the infinite sum in (3.19) can survive and has the value $\psi = A(\mathbf{k}, \mathbf{r})$, or $\psi = 0$ if I is not satisfied. Thus, formally we can speak about a numerical particle which evolves in the phase space as a classical carrier. According to (2.1), N such particles can be used to evaluate the desired mean value at time t: $\langle A \rangle \simeq \sum_{n=1}^{N} \psi_n / N$. Note that this is a mean value corresponding to a point-like carrier, initially placed at $\mathbf{k}_\alpha, \mathbf{r}_\alpha$. If the renormalization factor N is assigned to the numerical particle, the latter can be viewed as carrying N^{-1}th fraction of the physical characteristics of the carrier. Or we can say that the weight of such particles is N^{-1}.

The initial conditions (3.23) and boundary conditions (3.28) introduce additional integrals which, as discussed in the previous section, initialize carrier distributions at time 0 and t_b respectively. We assume these distributions dense enough, i.e. there are enough carriers in any initialization point, to allow the choice $N = 1$. Due to this convention, there is one to one correspondence between numerical particles and semiconductor carriers, the weight is unity, so that Monte Carlo numerical particles can be regarded as real semiconductor carriers.

The mean value $\langle A \rangle = \langle A \rangle_i + \langle A \rangle_b$ is decomposed into two terms which are evaluated separately. The initial probability $P^B_{0,i}$ for (3.23) is constructed by taking into

account the physical nature of f_i: The latter is presented by $N(0) = (1, f_i)$ carriers so that $P^B_{0,i} = f_i/N(0)$. According to the above convention, the number of the Monte Carlo particles is also $N(0)$. As discussed, the evolution of each of the initialized $n = 1, \ldots, N(0)$ particles follows the rules of the EMC algorithm and contribute to the estimate of $\langle A \rangle_i$ with ψ_n. According to (2.1) and (2.7) the following relation holds:

$$\langle A \rangle_i \simeq \frac{1}{N(0)} \sum_{n=1}^{N(0)} \frac{f_i}{P^B_{0,i}} \psi_n = \sum_{n=1}^{N(0)} \frac{P^B_{0,i}}{P^B_{0,i}} \psi_n = \sum_{n=1}^{N(0)} \psi_n = \sum_{n=1}^{N_i(t)} A_n \tag{4.2}$$

where $N_i(t)$ is the number of the particles which remain in the device during the evolution until time t. We note that any other number N in the place of $N(0)$ will give rise to a weight $N(0)/N$ for all particles.

The boundary term (3.28) is approached with the help of the physical characteristics (3.29) of the charge injection, which are known quantities, external for the simulation. The probability density for injection from the boundary per unit time is chosen to be

$$P^B_{0,b} = \frac{\Gamma_D(t_b)}{N_\Gamma(t)} \times \frac{j_\perp(\mathbf{r}_b, t_b)}{\Gamma_D(t_b)} \times \frac{v_\perp(\mathbf{k}_b) f_b(\mathbf{k}_b, \mathbf{r}_b, t_b)}{j_\perp(\mathbf{r}_b, t_b)} \tag{4.3}$$

With this choice of $P^B_{0,b}$, the estimator for (3.28) becomes:

$$\langle A \rangle_b \simeq \frac{1}{N_\Gamma(t)} \sum_{n=1}^{N_\Gamma(t)} \frac{v_\perp f_b}{P^B_{0,b}} \psi_n = \sum_{n=1}^{N_\Gamma(t)} \psi_n = \sum_{n=1}^{N_b(t)} A_n \tag{4.4}$$

where $N_b(t)$ is the number of the injected from the boundary particles which remain in the device until time t.

We show that the density $P^B_{0,b}$ provides the well known steps used in the EMC algorithm for boundary injection. $P^B_{0,b}$ is a product of three conditional probability densities. The formal Monte Carlo interpretation of this density is as follows. For each particle: (i) select the injection time t_b from the first density, (ii) select the boundary point by the second density, (iii) select the value of the in-warding wave vector by the third density.

i According to this formal step the injection time t_b is chosen randomly and a single particle is injected. The process continues until the number of injected particles becomes equal to $N_\Gamma(t)$. The net effect is that for a time interval Δt_b around t_b the differential number of injected particles is $\Gamma_D(t_b)\Delta t_b/N_\Gamma(t)$. It is then equivalent (and more convenient) to inject the corresponding differential number of particles at the consecutive steps Δt_b of the evolution.

ii As a rule f_b is the equilibrium distribution f_e, a function which characterizes given contact, assumed to be a homogeneous region. Then $j_\perp$ is a constant, characterizing this region.

iii It can be shown that if the particle distribution in a homogeneous domain is the equilibrium one, the distribution on the domain boundaries is $v_\perp f_e$ and vice versa. Thus there are two ways to inject f_e-distributed particles in the device: either to fix the initial position of all particles at the boundary, or to distribute them evenly in a chosen domain near the boundary.

To this extend the scheme relies on the explicit knowledge of $j_\perp$, Γ_D and N_Γ which are calculated from f_e. Alternatively, it is convenient to determine these quantities implicitly. In the homogeneous domain of the boundaries the charge of the carriers must compensate the charge of the dopants. Thus at each step of the simulation as many particles, as needed for recovering the charge neutrality (usually disturbed during the preceding step), are injected into the contacts. In this way we obtain the EMC scheme for boundary injection.

The estimator for $\langle A \rangle$ is a sum of the initial and the boundary estimators:

$$\langle A \rangle(t) = \langle A \rangle_i + \langle A \rangle_b \simeq \sum_{n=1}^{N_i(t)} A_n + \sum_{n=1}^{N_b(t)} A_n = \sum_{n=1}^{N(t)} A_n \tag{4.5}$$

That is, all particles remaining in the device are counted in the same way, no matter whether related to the initial or to the boundary conditions.

To summarize: The particular choice of the initial probability P_0^B gives rise to the EMC scheme for initialization of particles at the initial time 0 inside D, and at the boundary, at time $t_b \in [0, t]$. Once initialized they follow the evolution governed by the natural probability densities p_t and p_S of the Boltzmann transport process, which comprises the evolution scheme of the EMC. The numerical trajectories obtained by using these densities are possible classical trajectories of the real semiconductor carriers. At time t only particles which remain during the evolution inside D contribute to the estimated value of A. A Monte Carlo particle has the same physical characteristics $A(\mathbf{k}, \mathbf{r})$ as has a real carrier located in the point $\mathbf{k}, \mathbf{r}$ of the phase space. The above considerations allow to view the particularly obtained algorithm, as emulation of the real transport process. The latter presents the usual understanding of the EMC algorithm for device simulations. The difference now is that the EMC algorithm, which accounts for both, initial and boundary conditions, is obtained by using the formal rules of the numerical Monte Carlo theory.

Choices of alternative initial and transition probabilities give rise to algorithms for event biasing.

4.2. Event biasing

Since the degeneracy of the kernel of the Boltzmann equation, the transition probabilities are not entirely arbitrary: The latter must respect all delta functions in the kernel, which means that the evolution must proceed over classical trajectory segments and the scattering between these segments must obey the energy conservation imposed by the scattering functions. In this sense these transition probabilities can be associated with generation of free flight and scattering events, having distributions, which differs from the classical ones. Or, equivalently, the application of the such probabilities can be regarded as biasing of the classical evolution. In this approach, called Weighted EMC (WEMC), the evolution (and thus the distribution and the number) of the biased particles differ from these of the EMC particles.

It will be shown that in the case of event biasing the EMC estimator (4.5) is generalized to

$$\langle A \rangle(t) \simeq \sum_{n=1}^{N^{\text{bias}}(t)} w_n A_n \tag{4.6}$$

where $N^{\text{bias}}(t)$ is the number of the biased particles and w_n is the weight of the nth particle. The weight can be viewed as a memory for the numerical trajectory which leads a particle to given phase space point. In contrast the EMC particles have no memory: their contribution to the estimators is independent of the way they arrive at that point. We consider some possible ways of biasing the Boltzmann evolution. The corresponding estimators are discussed and weights are determined by (4.6). While biasing of the initial or boundary distributions give rise to weights which conserve during the evolution, the biasing of the evolution probabilities produce weighs which are dynamically updated in time. Then w_n in (4.6) is the value of the weight at time t (similar to A_n).

4.2.1. Biasing the initial and boundary conditions

Initial conditions The biased initial distribution could be any positive function $P_{0,i}^{\text{bias}}$ normalized to unity in D. According (2.7) the biased initial distribution replaces $P_{0,i}^{B}$ in the denominator of (4.2). The weight of a numerical particle initialized at $(\mathbf{k}, \mathbf{r})$ is

then

$$w(\mathbf{k},\mathbf{r}) = \frac{P^B_{0,i}(\mathbf{k},\mathbf{r})}{P^{\text{bias}}_{0,i}(\mathbf{k},\mathbf{r})} \tag{4.7}$$

Depending on the concrete choice of $P^{\text{bias}}_{0,i}$ particles have different weight distribution. However, there is a general property that the mean weight remains unity in D:

$$\int d\mathbf{k}d\mathbf{r} P^{\text{bias}}_{0,i}(\mathbf{k},\mathbf{r}) w(\mathbf{k},\mathbf{r}) = \int d\mathbf{k}d\mathbf{r} P^{B}_{0,i}(\mathbf{k},\mathbf{r}) = 1 \simeq \frac{1}{N(0)} \sum_{l}^{N(0)} w_l$$

where the last term, the estimator of the weight, shows that $N(0)$ biased particles can represent $N(0)$ real carriers exactly as the EMC particles do. That is, both properties hold: the number of the biased particles is $N(0)$ and their weights sum up to $N(0)$. It is possible to use another number N in the place of $N(0)$. This will lead to a renormalization of the weight w in (4.7) by a factor $N(0)/N$.

Boundary conditions Two ways of biasing the boundary conditions by modifying the equilibrium boundary distribution $f_b = f_e$ to f^{bias} are considered. The term $v_\perp f_e$ in (3.28) is assigned to the random variable and the injection from the boundary is realized by a density $P^{\text{bias}}_{0,b}$:

$$P^{\text{bias}}_{0,b} = \frac{v_\perp f^{\text{bias}}}{N^{\text{bias}}_\Gamma(t)}, \qquad N^{\text{bias}}_\Gamma(t) = \int_0^t dt_b \oint d\sigma(\mathbf{r}_b) \int d\mathbf{k} v_\perp f^{\text{bias}}$$

where the N^{bias}_Γ is obtained in accordance with (3.29). The estimator for (3.28), written for the general choice of $N_{\text{num}}(t)$ injected biased particles becomes:

$$\langle A\rangle_b \simeq \frac{1}{N_{\text{num}}(t)} \sum_{n=1}^{N_{\text{num}}(t)} \frac{v_\perp f_e}{P^{\text{bias}}_{0,b}} \psi_n, \qquad w = \frac{v_\perp f_e}{P^{\text{bias}}_{0,b} N_{\text{num}}(t)} \tag{4.8}$$

A particular choice of $N_{\text{num}}(t) = N^{\text{bias}}_\Gamma(t)$ gives rise to a weight $w = f_e/f^{\text{num}}$.

$$\langle A\rangle_b \simeq \sum_{n=1}^{N^{\text{bias}}_\Gamma(t)} \frac{f_e}{f^{\text{bias}}} \psi_n = \sum_{n=1}^{N^{\text{bias}}(t)} w_n A_n$$

While the number of injected particles $N_{\text{num}}(t)$ (or $N^{\text{bias}}_\Gamma(t)$) is an externally imposed quantity, the number of biased particles $N^{\text{bias}}(t)$ is a quantity, controlled by the simulation.

A scheme which ensures the injection of $N_\Gamma^{\text{bias}}(t)$ biased particles is obtained with the help of (4.8): the number of injected real carriers is obtained by setting $\psi = 1$.

$$N_\Gamma(t) = \sum_{n=1}^{N_\Gamma^{\text{bias}}(t)} \frac{f_e}{f^{\text{bias}}} = \sum_{n=1}^{N_\Gamma^{\text{bias}}(t)} w_n$$

Thus the sum of the weights of the injected biased particles equals the number of the injected carriers at any evolution time t. Hence, if the biased particles are injected until the sum of their weights meets the requirement for a charge neutrality at any time step, in any point of the boundary domain, their number will be $N_\Gamma^{\text{bias}}(t)$ at any evolution time. By following this result, we avoid the need for explicit knowledge of the quantities in (3.29).

The averaged value of the weight per biased particle is estimated as:

$$\bar{w} = \int_0^t dt_b \oint d\sigma(\mathbf{r}_b) \int d\mathbf{k} w P_{0,b}^{\text{bias}} = \frac{N_\Gamma(t)}{N_\Gamma^{\text{bias}}(t)}$$

where we have used the equality $w v_\perp f^{\text{bias}} = v_\perp f_e$. Usually $\bar{w} \neq 1$ so that the number of biased particles is different from the EMC particles.

Two particular choices of the f^{bias} are considered below. By denoting the equilibrium distribution

$$f_e(\epsilon, T) = \frac{1}{\bar{\epsilon}} \mathrm{e}\{-\frac{\epsilon}{\bar{\epsilon}}\} \qquad \bar{\epsilon} = 1.5 k_B T \tag{4.9}$$

we first choose $f^{\text{bias}} = f_e(\epsilon, T_b)$ to be an equilibrium distribution, corresponding to a higher temperature T_b. Having higher kinetic energy, the numerical particles readily overcome the source potential barrier and enrich the statistics in the channel. The following peculiarity of the ratio f_e/f^{bias} can be observed. With the increase of T_b increases the spread of the weight further away from unity. Accordingly increases the variance of the physical averages, obtained by the mean of heavy and light particles. Thus, the choice of an the appropriate bias is a matter of compromise between the need for more particles in the channel and keeping the spread of the weight low.

The second choice of f^{bias} utilizes the idea of a weight control. A desired weight w_1 of the numerical particles with kinetic energy below given level ϵ_1 can be chosen. f^{bias} is obtained from f_e as follows:

$$f^{\text{bias}}(\epsilon) = \frac{f_e(\epsilon)}{w_1 \bar{\epsilon}}, \quad \epsilon \leq \epsilon_1, \qquad f^{\text{bias}}(\epsilon) = \frac{f_e(\epsilon)}{w_2 \bar{\epsilon}}, \quad \epsilon > \epsilon_1$$

The weight w_2 is obtained as a function of w_1 and ϵ_1 form the condition for normalization of f^{bias}:

$$w_2 = \frac{w_1 f_e(\epsilon_1)}{c_1 - 1 + f_e(\epsilon_1)} \tag{4.10}$$

The physical distribution (4.9) is recovered by setting $w_1 = 1$. A choice of $w_1 > 1$ effectively reduces the number of particles below ϵ_1 as compared to the unbiased case. Accordingly, in the energy region above ϵ_1, the number of the light particles increases. Indeed, from (4.10) it follows that $w_2 < 1$ as long as $w_1 > 1$. ϵ_1 can be adjusted in a way to enrich the statistics of the carriers travelling in the channel.

In both cases of biasing heavy particles which enter the channel perturb the statistics accumulated by a set of small weight particles. It is thus desirable to apply the technique of particle splitting in parallel to the biasing in order to minimize the spread of the weight.

4.2.2. Biasing the evolution

The evolution process is biased by a replacement of the Boltzmann probabilities comprising P^B in (3.19) by the corresponding counterpart P^{bias}. The ratio $w = P^B/P^{\text{bias}}$ is inserted (and assigned to the random variable) in order to retain the value of (3.19) unchanged. Each replacement determines a factor w which multiplies the random variable and thus updates the weight in the corresponding estimator. Thus we obtain a novel effect of the weight evolution: the weight is updated after each biased evolution step by multiplication with the corresponding factor w. In contrast the weight in all previous cases is established once and conserves during the evolution.

Free flight The free flight distribution can be biased to become p_t^{bias}, by replacing the physical total out-scattering rate λ in p_t by a numerical out-scattering rate λ_b.

By virtue of the ideas, illustrated by (2.7) the physical free flight distribution p_t in (3.19) is replaced by the biased distribution p_t^{bias}. Each replacement corresponds to a weight factor

$$w(t) = \frac{p(t; \mathbf{k}, \mathbf{r}, t_0)}{p_t^{\text{bias}}(t; \mathbf{k}, \mathbf{r}, t_0)}$$

In the special case $\lambda_b = 0$ the particle can travel without scattering over the trajectory, changing its weight according $w(t) = p(t; \mathbf{k}, \mathbf{r}, t_0)$. In this way particles with appropriate velocity may be "encouraged" to enter the desired region of interest.

Phonon scattering The phonon scattering can be biased by modifying any of the two step process of selection of the after-scattering state. Artificial carrier heating can be achieved by biasing of the phonon scattering rates. For a given scattering mechanism, the probability for phonon absorption is increased at the expense of phonon emission, Controlled by a parameter $w \geq 1$,

$$\lambda_a^{\text{bias}} = \lambda_a + \lambda_e \left(1 - \frac{1}{w}\right), \quad \lambda_e^{\text{bias}} = \frac{\lambda_e}{w}. \tag{4.11}$$

If in the course of the simulation a phonon absorption is selected, the particle weight must be updated by a multiplication with $\lambda_a/\lambda_a^{\text{bias}}$, otherwise with $\lambda_e/\lambda_e^{\text{bias}} = w$. The distribution of the flight time is not affected, because the sum of the emission and absorption rates remains unchanged.

Carrier can be guided towards preferred direction by modifying the distribution of the scattering angle. In particular, we consider isotropic processes, where the distribution of $\chi = \cos\theta$ is a constant: $p(\chi) = 1/2$ for $\chi \in (-1, 1)$. Here θ is defined as the angle between the after-scattering momentum and the desired direction. The following modified density function increases the probability for forward scattering at the expense of backscattering.

$$p^{\text{bias}}(\chi) = \begin{cases} \dfrac{1}{2w} & -1 \leq \chi < \chi_0 \\ \dfrac{w}{2} & \chi_0 \leq \chi < 1 \end{cases} \tag{4.12}$$

Here $w \geq 1$ is a given parameter, χ_0 is determined from the normalization:

$$\chi_0 = \frac{w-1}{w+1} \qquad P^{\text{bias}}(\chi_0) = \frac{\chi_0 + 1}{2w} = \frac{1}{1+w}$$

and $P^{\text{bias}}(\chi_0)$ is the cumulative probability. In this case if r is a random number evenly distributed between 0 and 1, and $r < P^{\text{bias}}(\chi_0)$ it holds:

$$\chi_r = 2wr - 1, \quad \frac{p}{p^{\text{bias}}} = w,$$

Otherwise

$$\chi_r = 1 - \frac{2(r-1)}{w}, \quad \frac{p}{p^{\text{bias}}} = \frac{1}{w}.$$

This means that the particle weight is either reduced or increased by the factor w whenever χ is generated from the density (4.12).

5. Self-consistent coupling with the Poisson equation

The Coulomb interaction between the semiconductor carriers is accounted by the Poisson equation:

$$\nabla(\epsilon\nabla V) = q(D+C) \qquad C(\mathbf{r},t) = \int d\mathbf{k} f(\mathbf{k},\mathbf{r},t) \qquad \mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}) \quad (5.1)$$

Here V is the electrical potential in the device and D is the concentration of the ionized donors. The equation links the electrical potential V with the distribution function f. The Boltzmann equation for Coulomb carriers becomes nonlinear via the electric force $\mathbf{F}(f)(\mathbf{r},t)$, now depending on the distribution function f. As the results of the previous section are based on the linearity of the integral equations involved, it is no longer possible to apply the steps used to derive event biasing schemes. The solution is sought in the iterative procedure of coupling of the EMC technique with the Poisson equation: the latter is discretized by a decomposition of the device region into mesh cells Ψ_l. The particle system is evolved in time intervals $\Delta t \simeq 0.1 fs$. At the end of each time step, at say time t, the charge density $qC(\mathbf{r}_l,t)$ is calculated and assigned to the corresponding grid points. We use the relation between C_l and the distribution function f, which is estimated with the help of (4.5) by introducing a mesh Φ_m in the wave vector space. Thus $\Omega_{m,l} = \Phi_m \Psi_l$, $\mathbf{k}_m, \mathbf{r}_l \in \Omega_{m,l}$, and:

$$\begin{aligned} f(\mathbf{k}_m,\mathbf{r}_l,t) &\simeq \frac{\sum_n^{N(t)} \theta_{\Omega_{m,l}}(n)}{V_{\Phi_m} V_{\Omega_l}}, \\ C(\mathbf{r}_l,t) &\simeq \sum_m f(\mathbf{k}_m,\mathbf{r}_l,t) V_{\Phi_m}, \\ N(t) &\simeq \sum_l C(\mathbf{r}_l,t) V_{\Psi_l} \end{aligned} \quad (5.2)$$

Here the argument of the θ function is a compact notation for the position of the n-th EMC particle. The above relations become equalities in the thermodynamic limit $N \to \infty$, $V_{\Phi_m}, V_{\Omega_l} \to 0$. The charge density C_l is used to find the solution of the Poisson equation, which provides an update for the electric force $\mathbf{F}(\mathbf{r},t)$ within the following scheme:

$$\begin{array}{ccccccccc} & \text{EMC} & & \text{PE} & \text{EMC} & & \text{PE} & \text{EMC} & \text{PE} \\ f_0 & \longrightarrow & f_{\Delta t} & \longrightarrow & f_{2\Delta t} & \cdots & \longrightarrow & & f_t \\ & \Delta t & & & 2\Delta t & & & t & \end{array}$$

The latter governs the trajectories evolving the particles in the next time interval $t, t + \Delta t$.

A generalization towards event biasing becomes possible due to two features of the approach: (i) The electric field remains frozen during the successive steps of solving the Boltzmann equation; (ii) The Markovian character of the Boltzmann transport. The latter is formally proven as follows: Equation (3.9) is rewritten by splitting the time interval $(0, t)$ into $(0, \tau)$ and (τ, t)

$$f(\mathbf{k},\mathbf{r},t) = \int_\tau^t dt' \theta_D(\mathbf{R}(t')) \int \mathrm{d}\mathbf{k}' f(\mathbf{k}',\mathbf{R}(t'),t') S(\mathbf{k}',\mathbf{K}(t'),\mathbf{R}(t')) e^{-\int_{t'}^{t}\lambda(y)dy} +$$
$$e^{-\int_\tau^t \lambda(y)dy}\left[\int_0^\tau dt' \theta_D(\mathbf{R}(t')) \int \mathrm{d}\mathbf{k}' f(\mathbf{k}',\mathbf{R}(t'),t') S(\mathbf{k}',\mathbf{K}(t'),\mathbf{R}(t')) e^{-\int_{t'}^{\tau}\lambda(y)dy}\right.$$
$$\left. + e^{-\int_\tau^t \lambda(y)dy} e^{-\int_0^\tau \lambda(y)dy} f_i(\mathbf{K}(0),\mathbf{R}(0)) + e^{-\int_{t_b}^{\tau}\lambda(y)dy} f_b(\mathbf{K}(t_b),\mathbf{R}(t_b),t_b)\right]$$
$$+ e^{-\int_{t_b}^{t}\lambda(y)dy} f_b(\mathbf{K}(t_b),\mathbf{R}(t_b),t_b)$$

where the shortcut $\lambda(y) = \lambda(\mathbf{K}(y),\mathbf{R}(y))$ has been used. Here, by recalling the considerations above equation (3.9), the boundary term has been decomposed into two contributions: $t_b \leq \tau$ and $t_b > \tau$. The term in the square brackets is $f(\mathbf{K}(\tau),\mathbf{R}(\tau),\tau)$. Thus the solution f at time τ inside the device becomes an initial condition for the future evolution.

Assume a biased process, where at time $\tau = t - \Delta t$ the biased particles emerge with weights w_n. The carrier distribution function f and the distribution function of the numerical particles can be evaluated with the help of (4.6) and (5.2):

$$f(\mathbf{k}_m,\mathbf{r}_l,\tau) \simeq \frac{\sum_n^{N^{\mathrm{bias}}(\tau)} w_n \theta_{\Omega_{m,l}(n)}}{V_{\Phi_m} V_{\Omega_l}}, \qquad f^{\mathrm{bias}}(\mathbf{k}_m,\mathbf{r}_l,\tau) \simeq \frac{\sum_n^{N^{\mathrm{bias}}(\tau)} \theta_{\Omega_{m,l}(n)}}{V_{\Phi_m} V_{\Omega_l}}$$

Here the argument of the θ function is a compact notation for the position of the nth biased particle. Now we can use f to evaluate the physical carrier density at time τ. Accordingly the actual electric potential and force are provided by the Poisson equation. Now the question is what to do with the biased particles? Can they continue

further the evolution or a novel ensemble of particles with different weight and distribution must be introduced in order to correctly evaluate the carrier distribution at time $\tau + \Delta t = t$? The answer is that either way is possible: In the thermodynamic limit $N^{\text{bias}} \to \infty$, $V_{\Phi_m}, V_{\Omega_l} \to 0$, the values of f and f^{bias} become well defined numbers in any point of the phase space. We can formulate a biased initial distribution $P^{\text{bias}}_{\tau,i}$ and the carrier initial distribution $P_{\tau,i}$ (at time τ!):

$$P^{\text{bias}}_{\tau,i} = \frac{f^{\text{bias}}}{N^{\text{bias}}(\tau)}, \qquad P_{\tau,i} = \frac{f}{N(\tau)}, \qquad w(\mathbf{k},\mathbf{r}) = \frac{P_{\tau,i}}{P^{\text{bias}}_{\tau,i}} = \frac{f}{f^{\text{bias}}} \frac{N^{\text{bias}}(\tau)}{N(\tau)} \quad (5.3)$$

and proceed according the discussion after (4.7). The number of the biased particles which continue the evolution changes from $N^{\text{bias}}(\tau)$ to $N(\tau)$ - the actual number of carriers in the device. This approach renormalizes the original weights w_n but keeps the distribution proportional to f^{bias}. As $P_{\tau,i}$ is known, any other initial distribution, (which can be entirely different from f^{bias}) can also be used for event biasing.

Alternatively, a choice of $N^{\text{bias}}(\tau)$ particles will renormalize the weight in (5.3) by a factor $N(\tau)/N^{\text{bias}}(\tau)$:

$$w(\mathbf{k}_m,\mathbf{r}_l) = \frac{f(\mathbf{k}_m,\mathbf{r}_l,\tau)}{^{\text{bias}}(\mathbf{k}_m,\mathbf{r}_l,\tau)} \simeq \frac{\sum_n^{N^{\text{bias}}(\tau)} w_n \theta_{\Omega_{m,l}(n)}}{\sum_n^{N^{\text{bias}}(\tau)} \theta_{\Omega_{m,l}(n)}} = \frac{\sum_n^{N^{\text{bias}}(\tau)} w_n \theta_{\Omega_{m,l}(n)}}{N^{\text{bias}}_{m,l}(\tau)}$$

The orthodox Monte Carlo approach requires $N^{\text{bias}}_{m,l}(\tau)$ particles initiating from $\mathbf{k}_m, \mathbf{r}_l$ with a weight $w(\mathbf{k}_m,\mathbf{r}_l)$. However from the last term it follows that $w(\mathbf{k}_m,\mathbf{r}_l)$ is the mean weight per particle, averaged within the set of $N^{\text{bias}}_{m,l}(\tau)$ biased particles which reside in $(\mathbf{k}_m,\mathbf{r}_l)$. As the weights w_n sample w, we can use the original weights and states of the particles from this set to continue the evolution. It follows that biased particles, which emerge at time τ with weights w_n present a biased initial condition for the next step of the evolution:

$$\begin{array}{ccccccccccc} & \text{WEMC} & & \text{PE} & & \text{WEMC} & & \text{PE} & & & \\ f_0^{\text{bias}} & \longrightarrow & f^{\text{bias}}_{\Delta_t} & \to & f_{\Delta t} & \to f^{\text{bias}}_{\Delta_t} & \longrightarrow & f^{\text{bias}}_{2\Delta_t} & \to & f_{2\Delta t} & \to f^{\text{bias}}_{2\Delta_t} \cdots \\ & \Delta t & & & & 2\Delta t & & & & & \end{array}$$

Concluded is that the particle weights and states survive between the successive iteration steps, which completes the proof of the self-consistent WEMC scheme.

6. Conclusion

The Boltzmann-Poisson equations coupling scheme is considered for the general case of boundary and initial conditions in presence of event biasing. A self-consistent Weighted EMC method is derived and the basic steps of the method are analyzed in terms of the Monte Carlo theory. A set of algorithms for biasing the initial, and/or the boundary conditions and the evolution of the weighted particles is presented and their peculiarities with respect to optimization of the statistical enhancement are discussed. It is shown that weighted particles can survive the successive iteration steps of the self-consistent coupling with the Poisson equation. Alternatively, the algorithm of biasing can be changed after any step on the expense of a corresponding renormalization of the weight of the numerical particles.

We would like to note another aspect of the presented approach, which is related to the formal similarities between the Boltzmann and Wigner equations. Due to the latter the formal part of the analysis could be applied in the quantum case, so that this work can be viewed as a necessary step towards development of self-consistent quantum Monte Carlo algorithms.

References

1. W. Fawcett, A. Boardman, and S. Swain, *J. Phys. Chem. Solids* **31**, 1963 (1970).
2. G. Baccarani, C.Jacoboni, and A. M. Mazzone, *Solid-State Electron.* **20**, 5 (1977).
3. A. Reklaitis, *Physics Letters* **88A**, 367 (1982).
4. C. Jacoboni, P. Poli, and L. Rota, *Solid-State Electron.* **31**, 5523 (1988).
5. M. Nedjalkov and P. Vitanov, *Solid-State Electron.* **32**, 893 (1989).
6. L. Rota, C. Jacoboni, and P. Poli, *Solid-State Electron.* **32**, 1417 (1989).
7. M. Nedjalkov and P. Vitanov, *Solid-State Electron.* **33**, 407 (1990).
8. P. Vitanov, M. Nedjalkov, C. Jacoboni, F. Rossi, A. Abramo, in *Advances in Parallel Algorithms,*, ed. Bl.Sendov AND I.Dimov, IOS Press, 117–128, (1994).
9. H. Kosina, M. Nedjalkov, S. Selberherr, *Journal of Applied Physics* **93**, 6, 3553–3563, 2003
10. H. Kosina, M. Nedjalkov, S. Selberherr, *Journal of Applied Physics* **93**, 6, 3564–3571, 2003
11. M. Nedjalkov, S. Ahmed , D. Vasileska, *Journal of Computational Electronics*, **3**, 3–4, 305–309, 2004,

Received June 8, 2007; revised October 28, 2007

Author information

M. Nedjalkov, Institute for Microelectronics, TU Vienna, Gusshausstrasse 27–29 E360, 1040 Vienna, Austria.
Email: mixi@iue.tuwien.ac.at

D. Vasileska, Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-5706, USA.
Email: vasileska@asu.edu

I. Dimov, Institute for Parallel Processing, Bulgarian Academy of Sciences, 1113 Sofia, and ACET Centre, University of Reading, Whiteknights P.O. Box 217, Reading RG6 6AH, Bulgaria/UK.
Email: i.t.dimov@reading.ac.uk

G. Arsov, Faculty of Electrical Engineering and Information Technologies, SS "Cyril snd Methodius" University, Skopje, Karpos II b.b., P.O. Box 574, 1001 Skopje, Macedonia, Macedonia.
Email: g.arsov@ieee.org