We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

15

Monte Carlo Device Simulations

Dragica Vasileska[1], Katerina Raleva[2] and Stephen M. Goodnick[1]
[1]*Arizona State University, Tempe AZ*
[2]*University Sts Cyril and Methodius, Skopje,*
[1]*USA*
[2]*Republic of Macedonia*

1. Introduction

As semiconductor devices are scaled into nanoscale regime, first velocity saturation starts to limit the carrier mobility due to pronounced intervalley scattering, and when the device dimensions are scaled to 100 nm and below, velocity overshoot starts to dominate the device behavior leading to larger ON-state currents. Alongside with the developments in the semiconductor nanotechnology, in recent years there has been significant progress in physical based modeling of semiconductor devices. First, for devices for which gradual channel approximation can not be used due to the two-dimensional nature of the electrostatic potential and the electric fields driving the carriers from source to drain, drift-diffusion models have been exploited. These models are valid, in general, for large devices in which the fields are not that high so that there is no degradation of the mobility due to the electric field. The validity of the drift-diffusion models can be extended to take into account the velocity saturation effect with the introduction of field-dependent mobility and diffusion coefficients. When velocity overshoot becomes important, drift diffusion model is no longer valid and hydrodynamic model must be used. The hydrodynamic model has been the workhorse for technology development and several high-end commercial device simulators have appeared including Silvaco, Synopsys, Crosslight, etc. The advantages of the hydrodynamic model are that it allows quick simulation runs but the problem is that the amount of the velocity overshoot depends upon the choice of the energy relaxation time. The smaller is the device, the larger is the deviation when using the same set of energy relaxation times. A standard way in calculating the energy relaxation times is to use bulk Monte Carlo simulations. However, the energy relaxation times are material, device geometry and doping dependent parameters, so their determination ahead of time is not possible. To avoid the problem of the proper choice of the energy relaxation times, a direct solution of the Boltzmann Transport Equation (BTE) using the Monte Carlo method is the best method of choice. That is why the focus of this review paper is on explaining basic Monte Carlo device simulator and then the focus will be shifted on the inclusion of various higher order effects that explain particular physical phenomena or processes.

The Monte Carlo book chapter is organized as follows. First, the idea behind the Monte Carlo technique is outlined by revoking the path integral method for the solution of the BTE. This approach naturally leads to the free-flight-scatter sequence that is used in solving the BTE using the Monte Carlo method. Various scattering mechanisms relevant for different

materials are given to completely specify the collision integral in the BTE. A discussion followed with the presentation of a generic flow-chart for implementing bulk Monte Carlo code is presented. Note that bulk Monte Carlo approach is suitable for the characterization of materials, but in order to study behavior of semiconductor devices coupling of the Monte Carlo transport kernel with a Poisson equation solver which gives the self-consistent field that moves the carriers around is needed. Important ingredients in describing particle-based device simulators are the particle-mesh coupling, treatment of the Ohmic contacts and calculation of the current. A generic flowchart of a particle-based device simulator is provided. The prospects of the Monte Carlo method for the solution of the Boltzmann transport equation, in the context of device simulations of nanoscale structures and of solar cells and power devices, are discussed at the end of the book chapter.

2. Importance of MC particle-based device simulations

2.1 Industry trends and the need for modeling and simulation

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 [1]. 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 offers unique insight into device behavior by allowing the observation of phenomena that can not be measured on real devices. *Computational Electronics* [2,3,4] 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 [5] leading to integrated circuits. Device simulation can be thought of as one component of technology for computer-aided design (TCAD), which provides a basis for device modeling, which deals with compact behavioral models for devices and sub-circuits relevant for circuit simulation in commercial packages such as SPICE [6]. The relationship between various simulation design steps that have to be followed to achieve certain customer need is illustrated in Figure 1.

The goal of *Computational Electronics* 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. Figure 2 illustrates the main components of semiconductor device simulation at any level. There are two main kernels, which must be solved self-consistently with one another, 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 arising from the solution of Poisson's equation are necessary.

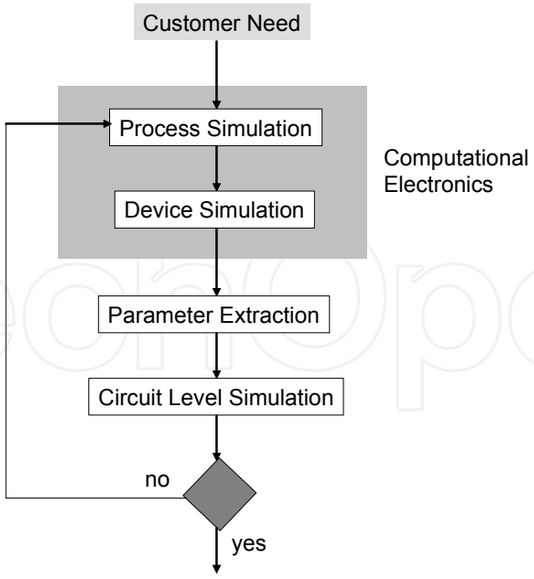


Fig. 1. Design sequence to achieve desired customer need.

The fields, in turn, are driving forces for charge transport as illustrated in Figure 3 for the various levels of approximation within a hierarchical structure ranging from compact modeling at the top to an exact quantum mechanical description at the bottom. At the very beginnings of semiconductor technology, the electrical device characteristics could be estimated using simple analytical models (gradual channel approximation for MOSFETs) 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 [7]. 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. Numerical simulation of carrier transport in semiconductor devices, dates back to the famous work of Scharfetter and Gummel [8], who proposed a robust discretization of the DD equations, which is still in use today.

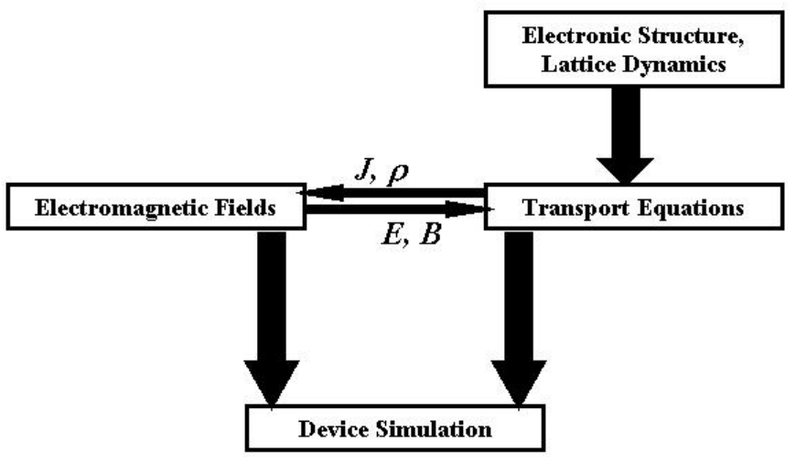


Fig. 2. Schematic description of the device simulation sequence.

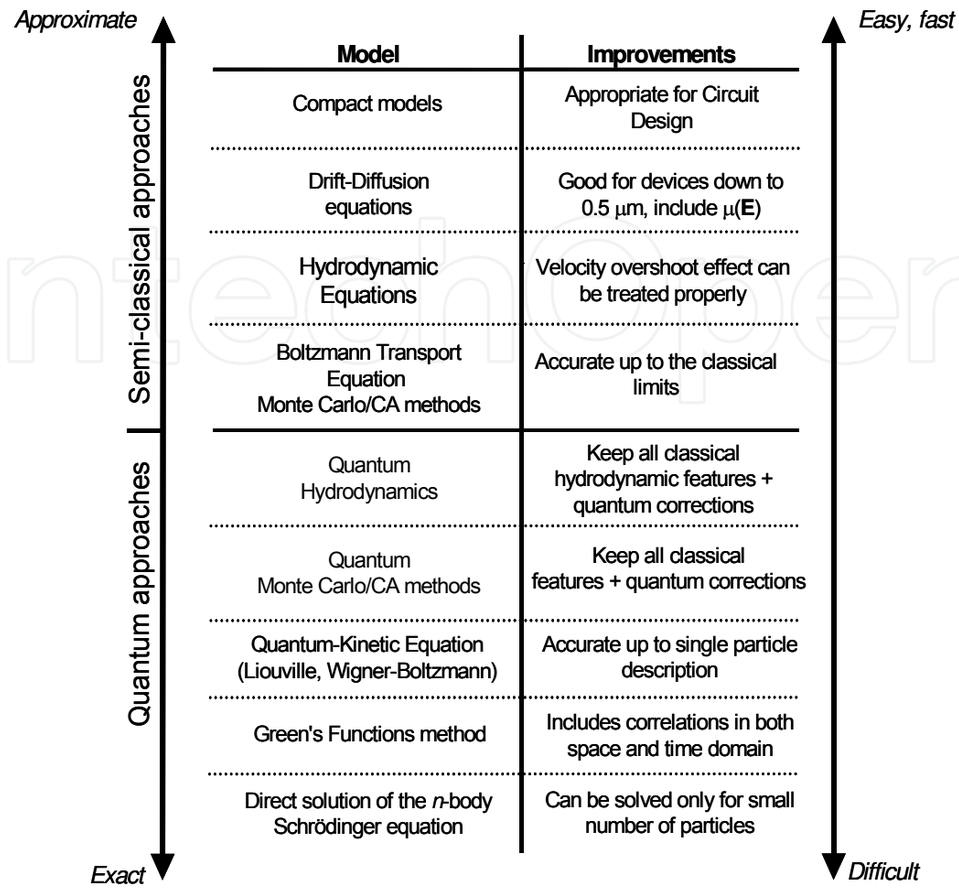


Fig. 3. Illustration of the hierarchy of transport models.

However, 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 [9]. Furthermore, an additional driving term is added to the current expression which is proportional to the gradient of the carrier temperature. However, a vast number of these models exist, and there is a considerable amount of confusion as to their relation to each other. It is now a common practice in industry to use standard hydrodynamic models in trying to understand the operation of as-fabricated devices, by adjusting any number of phenomenological parameters (e.g. mobility, impact ionization coefficient, etc.). However, such tools do not have predictive capability for ultra-small structures, for which it is

necessary to relax some of the approximations in the Boltzmann transport equation [10]. Therefore, one needs to move downward to the quantum transport area in the hierarchical map of transport models shown in Figure 3, where, at the very bottom we have the Green's function approach [11,12,13]. The latter is the most exact, but at the same time the most difficult of all. In contrast to, for example, the Wigner function approach (which is Markovian in time), 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 difficulties in understanding the various terms in the resultant equations and the enormous computational burden needed for its actual 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 [14], which is effectively 1D and is primarily applicable to resonant tunneling diodes.

From the discussion above it follows that, contrary to the recent technological advances, the present state of the art in device simulation is currently lacking in the ability to treat these new challenges in scaling of device dimensions from conventional down to quantum scale devices. For silicon devices with active regions below 0.2 microns in diameter, macroscopic transport descriptions based on drift-diffusion models are clearly inadequate. As already noted, even standard hydrodynamic models do not usually provide a sufficiently accurate description since they neglect significant contributions from the tail of the phase space distribution function in the channel regions [15,16]. 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.

2.2 Drift-Diffusion and hydrodynamic models

In Section 1.1 above, we discussed the various levels of approximations that are employed in the modeling of semiconductor devices. The direct solution of the full BTE is challenging computationally, particularly when combined with field solvers for device simulation. Therefore, for traditional semiconductor device modeling, the predominant model corresponds to solutions of the so-called drift-diffusion equations, which are 'local' in terms of the driving forces (electric fields and spatial gradients in the carrier density), i.e. the current at a particular point in space only depends on the instantaneous electric fields and concentration gradient at that point. The complete drift-diffusion model is based on the following set of equations:

Current equations:

$$\begin{aligned} J_n &= qn(x)\mu_n E(x) + qD_n \frac{dn}{dx} \\ J_p &= qp(x)\mu_p E(x) - qD_p \frac{dn}{dx} \end{aligned} \tag{1}$$

Continuity equations:

$$\begin{aligned}\frac{\partial n}{\partial t} &= \frac{1}{q}\nabla\cdot\mathbf{J}_n + U_n \\ \frac{\partial p}{\partial t} &= -\frac{1}{q}\nabla\cdot\mathbf{J}_p + U_p\end{aligned} \tag{2}$$

Poisson's equation:

$$\nabla\cdot(\varepsilon\nabla V) = -\left(p - n + N_D^+ - N_A^-\right), \tag{3}$$

where U_n and U_p are the net generation-recombination rates .

The continuity equations are the conservation laws for the carriers. A numerical scheme which solves the continuity equations should

- Conserve the total number of particles inside the device being simulated.
- Respect local positive definite nature of carrier density. Negative density is unphysical.
- Respect monotonicity of the solution (i.e. it should not introduce spurious space oscillations).

Conservative schemes are usually achieved by subdivision of the computational domain into patches (boxes) surrounding the mesh points. The currents are then defined on the boundaries of these elements, thus enforcing conservation (the current exiting one element side is exactly equal to the current entering the neighboring element through the side in common). In the absence of generation-recombination terms, the only contributions to the overall device current arise from the contacts. Remember that, since electrons have negative charge, the particle flux is opposite to the current flux. When the equations are discretized, using finite differences for instance, there are limitations on the choice of mesh size and time step [17]:

- The mesh size Δx is limited by the Debye length.
- The time step is limited by the dielectric relaxation time.

A mesh size must be smaller than the Debye length where one has to resolve charge variations in space. A simple example is the carrier redistribution at an interface between two regions with different doping levels. Carriers diffuse into the lower doped region creating excess carrier distribution which at equilibrium decays in space down to the bulk concentration with approximately exponential behavior. The spatial decay constant is the Debye length

$$L_D = \sqrt{\frac{\varepsilon k_B T}{q^2 N}} \tag{4}$$

where N is the doping density, ε is the dielectric constant, k_B is the Boltzmann constant, T is the lattice temperature and q is the elementary charge. In GaAs and Si, at room temperature the Debye length is approximately 400 Å when $N \approx 10^{16}\,cm^{-3}$ and decreases to about only 50 Å when $N \approx 10^{18}\,cm^{-3}$.

The dielectric relaxation time, on the other hand, is the characteristic time for charge fluctuations to decay under the influence of the field that they produce. The dielectric relaxation time may be estimated using

$$t_{dr} = \frac{\varepsilon}{qN\mu} \tag{5}$$

where μ is the carrier mobility.
The drift-diffusion semiconductor equations constitute a coupled nonlinear set. It is not possible, in general, to obtain a solution directly in one step, but a nonlinear iteration method is required. The two most popular methods for solving the discretized equations are the Gummel's iteration method [18] and the Newton's method [19]. It is very difficult to determine an optimum strategy for the solution, since this will depend on a number of details related to the particular device under study.
Finally, the discretization of the continuity equations in conservation form requires the determination of the currents on the mid-points of mesh lines connecting neighboring grid nodes. Since the solutions are accessible only on the grid nodes, interpolation schemes are needed to determine the currents. The approach by Scharfetter and Gummel [8] has provided an optimal solution to this problem, although the mathematical properties of the proposed scheme have been fully recognized much later.
In the computational electronics community, the necessity for the hydrodynamic (HD) transport model is normally checked by comparison of simulation results for HD and DD simulations. Despite the obvious fact that, depending on the equation set, different principal physical effects are taken into account, the influence on the models for the physical parameters is more subtle. The main reason for this is that in the case of the HD model, information about average carrier energy is available in form of carrier temperature. Many parameters depend on this average carrier energy, e.g., the mobilities and the energy relaxation times. In the case of the DD model, the carrier temperatures are assumed to be in equilibrium with the lattice temperature, that is $T_C = T_L$, hence, all energy dependent parameters have to be modeled in a different way.

2.2.1 Extensions of the Drift-Diffusion model

In the DD approach, the electron gas is assumed to be in thermal equilibrium with the lattice temperature $\left(T_n = T_L\right)$. However, in the presence of a strong electric field, electrons gain energy from the field and the temperature T_n of the electron gas is elevated. Since the pressure of the electron gas is proportional to nk_BT_n, the driving force now becomes the pressure gradient rather then merely the density gradient. This introduces an additional driving force, namely, the temperature gradient besides the electric field and the density gradient. Phenomenologically, one can write the electron current density equation as

$$\mathrm{J} = q\left(n\mu_n \mathrm{E} + D_n \nabla n + nD_T \nabla T_n\right) \tag{6}$$

where D_T is the thermal diffusivity and $\mathrm{D_n}$ is the diffusion constant.

2.2.2 Stratton's approach

One of the first derivations of extended transport equations was performed by Stratton [20]. First the distribution function is split into the even and odd parts

$$f(\mathrm{k},\mathrm{r}) = f_0(\mathrm{k},\mathrm{r}) + f_1(\mathrm{k},\mathrm{r})\,. \tag{7}$$

From $f_1(-\mathrm{k},\mathrm{r}) = -f_1(\mathrm{k},\mathrm{r})$, it follows that $\langle f_1 \rangle = 0$. Assuming that the collision operator C in the Boltzmann transport equation is linear and invoking the microscopic relaxation time approximation for the collision operator

$$C[f] = -\frac{f - f_{eq}}{\tau(\varepsilon,\mathrm{r})} \tag{8}$$

the BTE can be split into two coupled equations. In particular f_1 is related to f_0 via

$$f_1 = -\tau(\varepsilon,\mathrm{r})\left(\mathrm{v}\cdot\nabla_{\mathrm{r}} f_0 - \frac{q}{\hbar}\mathrm{E}\cdot\nabla_{\mathrm{k}} f_0\right). \tag{9}$$

The microscopic relaxation time is then expressed by a power law

$$\tau(\varepsilon) = \tau_0\left(\frac{\varepsilon}{k_B T_L}\right)^{-p}. \tag{10}$$

When f_0 is assumed to be heated Maxwellian distribution, the following equation system is obtained

$$\nabla\cdot\mathrm{J} = q\frac{\partial n}{\partial t}$$

$$\mathrm{J} = qn\mu\mathrm{E} + k_B\nabla\left(n\mu T_n\right)$$

$$\nabla\cdot(n\mathrm{S}) = -\frac{3}{2}k_B\partial\left(nT_n\right) + \mathrm{E}\cdot\mathrm{J} - \frac{3}{2}k_B n\frac{T_n - T_L}{\tau_\varepsilon} \tag{11}$$

$$n\mathrm{S} = -\left(\frac{5}{2} - p\right)\left(\mu n k_B T_n\mathrm{E} + \frac{k_B^2}{q}\nabla\left(n\mu T_n\right)\right)$$

Equation for the current density can be rewritten as:

$$J = q\mu\left(nE + \frac{k_B}{q}T_n\nabla n + \frac{k_B}{q}n\left(1+\nu_n\right)\nabla T_n\right), \tag{12a}$$

with

$$\nu_n = \frac{T_n}{\mu}\frac{\partial\mu}{\partial T_n} = \frac{\partial\ln\mu}{\partial\ln T_n} \tag{12b}$$

which is commonly used as a fit parameter with values in the range [-0.5,-1.0]. For ν_n =-1.0, the thermal distribution term disappears. The problem with Eq. (10) for τ is that p must be approximated by an average value to cover the relevant processes. In the particular case of impurity scattering, p can be in the range [-1.5,0.5], depending on charge screening. Therefore, this average depends on the doping profile and the applied field; thus, no unique

value for p can be given. Note also that the temperature T_n is a parameter of the heated Maxwellian distribution, which has been assumed in the derivation. Only for parabolic bands and a Maxwellian distribution, this parameter is equivalent to the normalized second-order moment.

2.2.3 Balance equations model

The first three balance equations, derived by taking moments of Boltzmann Transport Equation (BTE), take the form:

$$\begin{aligned}
\frac{\partial n}{\partial t} &= \frac{1}{e}\nabla\cdot \mathrm{J}_n + S_n \\
\frac{\partial J_z}{\partial t} &= \frac{2e}{m^*}\sum_i \frac{\partial W_{iz}}{\partial x_i} + \frac{ne^2}{m^*}E_z - \left\langle\left\langle \frac{1}{\tau_m}\right\rangle\right\rangle J_z \\
\frac{\partial W}{\partial t} &= -\nabla\cdot \mathrm{F}_W + \mathrm{E}\cdot\mathrm{J} - \left\langle\left\langle \frac{1}{\tau_E}\right\rangle\right\rangle (W - W_0)
\end{aligned} \tag{13}$$

The balance equation for the carrier density introduces the carrier current density, which balance equation introduces the kinetic energy density. The balance equation for the kinetic energy density, on the other hand, introduces the energy flux. Therefore, a new variable appears in the hierarchy of balance equations and the set of infinite balance equations is actually the solution of the BTE. The momentum and energy relaxation rates, that appear in Eq. (13) are ensemble averaged quantities. For simple scattering mechanisms one can utilize the drifted-Maxwellian form of the distribution function, but for cases where several scattering mechanisms are important, one must use bulk Monte Carlo simulations to calculate these quantities.

One can express the energy flux that appears in Eq. (13) in terms of the temperature tensor. The energy flux, is calculated using

$$\mathbf{F}_W = \frac{1}{V}\sum_{\mathbf{p}} \mathbf{v}E(\mathbf{p})f(\mathbf{r},\mathbf{p},t), \tag{14}$$

which means that the *i*-th component of this vector equals to

$$F_{Wi} = v_{di}W + nk_B\sum_j T_{ij}v_{dj} + Q_i \tag{15}$$

where Q_i is the component of the heat flux vector which describes loss of energy due to flow of heat out of the volume. To summarize, the kinetic energy flux equals the sum of the kinetic energy density times velocity plus the velocity times the pressure, which actually represents the work to push the volume plus the loss of energy due to flow of heat out. In mathematical terms this is expressed as

$$\mathbf{F}_W = \mathbf{v}W + nk_B\vec{\vec{T}}\cdot\mathbf{v} + \mathbf{Q}\ . \tag{16}$$

With the above considerations, the momentum and the energy balance equations reduce to

$$\frac{\partial J_z}{\partial t} = \frac{2e}{m^*}\sum_i \frac{\partial}{\partial x_i}\left(K_{iz} + \frac{1}{2}nk_B T_{iz}\right) + \frac{ne^2}{m^*}E_z - \left\langle\left\langle \frac{1}{\tau_m}\right\rangle\right\rangle J_z$$
$$\frac{\partial W}{\partial t} = -\nabla\cdot\left(\mathbf{v}W + \mathbf{Q} + nk_B\vec{\vec{T}}\cdot\mathbf{v}\right) + \mathbf{E}\cdot\mathbf{J}_n - \left\langle\left\langle \frac{1}{\tau_E}\right\rangle\right\rangle (W - W_0) \tag{17}$$

For displaced-Maxwellian approximation for the distribution function, the heat flux **Q** = 0. However, Blotekjaer [21] has pointed out that this term must be significant for non-Maxwellian distributions, so that a phenomenological description for the heat flux, of the form described by Franz-Wiedermann law, which states that

$$\mathbf{Q} = -\kappa\nabla T_c \tag{18}$$

is used, where κ is the thermal or heat conductivity. In silicon, the experimental value of κ is 142.3 W/mK. The above description for **Q** actually leads to a closed set of equations in which the energy balance equation is of the form

$$\frac{\partial W}{\partial t} = -\nabla\cdot\left(\mathbf{v}W - \kappa\nabla T_c + nk_B T_c\mathbf{v}\right) + \mathbf{E}\cdot\mathbf{J}_n - \left\langle\left\langle \frac{1}{\tau_E}\right\rangle\right\rangle (W - W_0) \tag{19}$$

It has been recognized in recent years that this approach is not correct for semiconductors in the junction regions, where high and unphysical velocity peaks are established by the Franz-Wiedemann law. To avoid this problem, Stettler, Alam and Lundstrom [22] have suggested a new form of closure

$$\mathbf{Q} = -\kappa\nabla T_c + \frac{5}{2}(1-r)\frac{k_B T_L}{e}\mathbf{J} \tag{20}$$

where **J** is the current density and *r* is a tunable parameter less than unity. Now using

$$\frac{\partial}{\partial x}\left(2K_{iz}\right) = \frac{\partial}{\partial x_i}\left(nm^* v_{di}v_{dz}\right) = nm^*\frac{\partial}{\partial x}\left(v_{di}v_{dz}\right) = nm^*\left[\frac{\partial v_{di}}{\partial x_i}v_{dz} + v_{dz}\frac{\partial v_{dz}}{\partial x_z}\right] \tag{21}$$

and assuming that the spatial variations are confined along the z-direction, we have

$$\frac{\partial}{\partial x_z}\left(2K_{iz}\right) = \frac{\partial}{\partial x_z}\left(nm^* v_{dz}^2\right). \tag{22}$$

To summarize, the balance equations for the drifted-Maxwellian distribution function simplify to

$$\frac{\partial n}{\partial t} = \frac{1}{e}\nabla \cdot J_n + S_n$$
$$\frac{\partial J_z}{\partial t} = \frac{e}{m^*}\frac{\partial}{\partial x_z}\left(nm^* v_{dz}^2 + nk_B T_c\right) + \frac{ne^2}{m^*}E_z - \left\langle\left\langle \frac{1}{\tau_m} \right\rangle\right\rangle J_z \quad (23)$$
$$\frac{\partial W}{\partial t} = -\frac{\partial}{\partial x_z}\left[\left(W + nk_B T_c\right)v_{dz} - \kappa\frac{\partial T_c}{\partial x_z}\right] + J_z E_z - \left\langle\left\langle \frac{1}{\tau_E} \right\rangle\right\rangle \left(W - W_0\right)$$

where

$$J_z = -env_{dz} = -\frac{e}{m^*}P_z$$
$$W = \frac{1}{2}nm^* v_{dz}^2 + \frac{3}{2}nk_B T_c \quad (24)$$

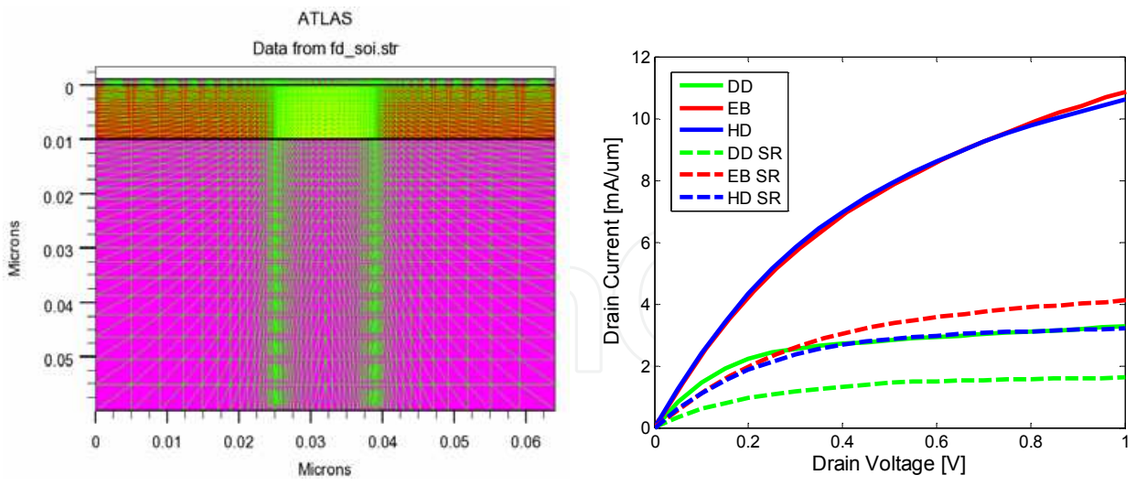
2.3 Failure of the Drift-Diffusion and hydrodynamic models

To understand the advantages and the limitations of the drift-diffusion and of the hydrodynamic model, let us consider the following examples: Fully Depleted (FD) SOI devices with channel lengths 25, 45 and 90 nm. The oxide thickness and the doping of the channel of the three devices considered are summarized in Table 1. In Figures 4 and 5 (a-c) we compare the output characteristics of the three devices when using the drift-diffusion and the hydrodynamic model.

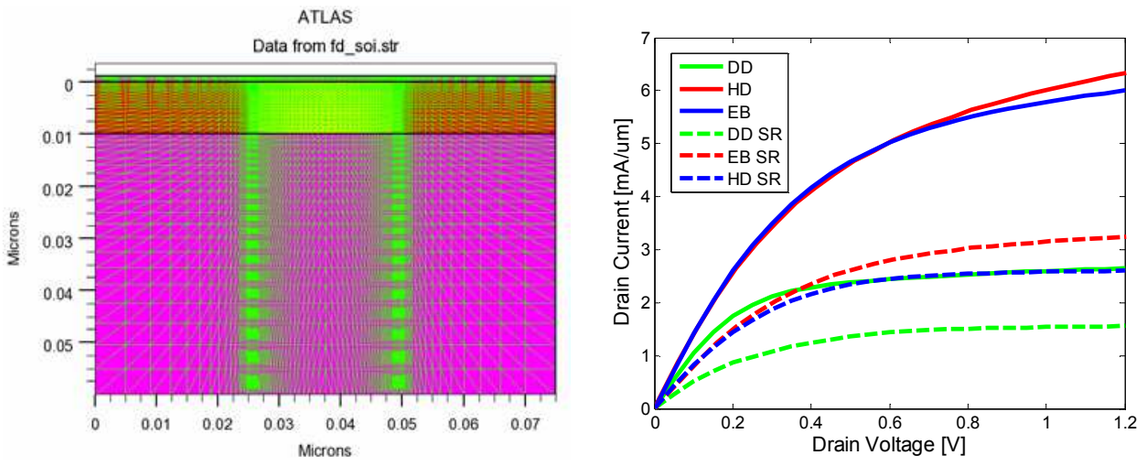
Feature (channel length)	14 nm	25 nm	90 nm
Tox	1 nm	1.2 nm	1.5 nm
V_{DS}	1V	1.2 V	1.4 V
Overshoot EB/HD without series resistance	233% / 224%	139% / 126%	31% /21%
Overshoot EB/DD with series resistance	153%/96%	108%/67%	39%/26%

Source/drain doping = 10^{20} cm^{-3} and 10^{19} cm^{-3} (series resistance (SR) case)
Channel doping = 10^{18} cm^{-3}
Overshoot= (ID_{HD}-ID_{DD})/ID_{DD} (%) ; ID is the on-state current

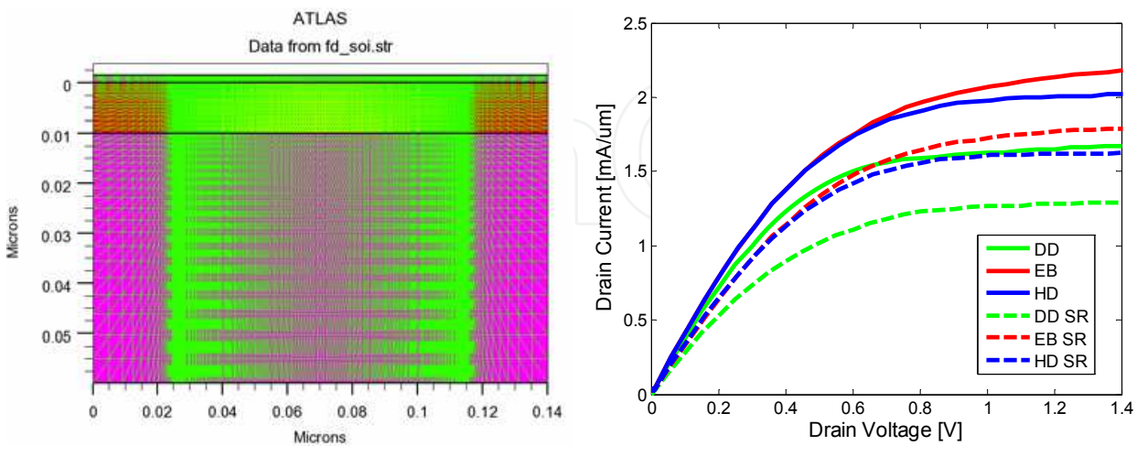
Table 1. Geometrical dimensions and applied biases of the fully-depleted SOI nMOSFETs simulated here.



a) channel length = 14 nm. V_G=1 V.

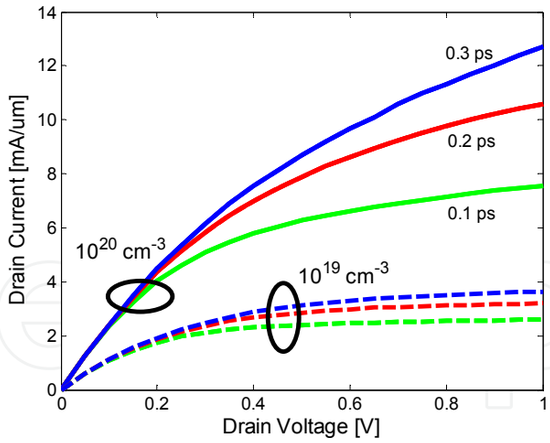


b) channel length = 25 nm. V_G=1.2 V.

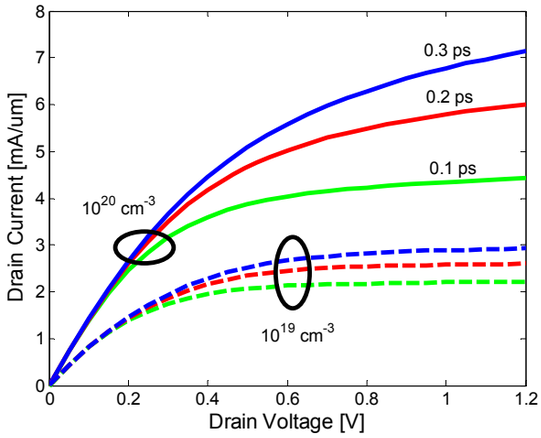


c) channel length = 90 nm. V_G=1.4 V.

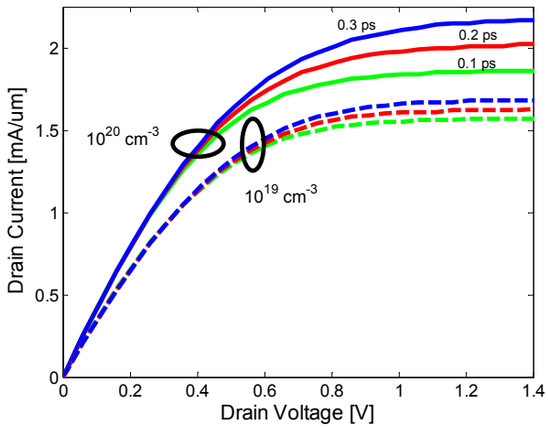
Fig. 4. Mesh and output characteristics of 14, 25 and 90 nm channel length FD SOI devices in the on-state when using drift-diffusion, energy balance, and hydrodynamic models. SR stands for series resistance.



a) Simulated characteristics for different energy relaxation times for two different source/drain doping for a channel length of 14 nm, V_G=1 V.



b) Simulated characteristics for different energy relaxation times for two different source/drain doping, for a channel length of 25 nm, V_G=1.2 V.



c) Simulated characteristics for different energy relaxation times for two different source/drain doping, for a channel length of 90 nm, V_G=1.4 V.

Fig. 5. Dependence of the on-state current upon the choice of the energy relaxation time for three different channel length FD SOI devices.

Here we use the commercial Silvaco Atlas (PISCEC) simulation package [23] that includes hydrodynamic modeling with momentum and energy relaxation times of 0.2 ps, Auger generation/recombination (important for the proper modeling of the heavily doped source and drain contacts), and the Schockley-Read-Hall (SRH) generation-recombination mechanism are included here for completeness, although the latter is not really important for this device structure. Impact ionization is not included in these simulations. In the hydrodynamic calculation, it is important that one uses the NEWTON method for solving the coupled set of equations, otherwise the simulation will not converge due to the strong coupling of the equations at high drain biases. We consider both the simplified energy balance (EB) model and the complete hydrodynamic model (HD). We present simulation results for the following two cases:

1. Source and drain doping of 10^{20} and 10^{19} cm^{-3} to examine series resistance effects. This is very important to know as in prototypical Monte Carlo device simulations source and drain regions are usually doped up to 10^{19} cm^{-3} to reduce the computational cost (total number of particles simulated). In these simulations we assume that the energy relaxation time is 0.2 ps, which is a typical value used for the silicon material system. The results from these simulations are presented in Figure 4 for the 14 nm, 25 nm and 90 nm channel length devices. On the left panel, we show the meshing used in these simulations and on the right panel we show the output characteristics for the appropriate on-state gate bias and drift-diffusion and hydrodynamic transport models.
2. In this second case we perform only hydrodynamic simulations to investigate the sensitivity of the hydrodynamic model to variations in the energy relaxation time which, in principle, is a material and device geometry dependent parameter which makes it almost impossible to determine analytically. This variation for the three technology nodes of devices is shown in Figure 5.

The results first show that the source/drain doping plays an important role in terms of the drive current, which is primarily effect of series resistance. From the results presented it is evident that non-stationary transport plays smaller role in 90 nm gate-length FD SOI devices, whereas the importance of non-stationary transport and the velocity overshoot associated with it increases drastically for 14 nm gate length FD SOI device. These results suggest that one must include energy balance equation if proper modeling of nano-scale devices with gate lengths less than 100 nm is to be achieved.

Yet another issue that deserves further attention is the dependence of the simulation results upon the choice of the energy relaxation time. In Figure 5 we plot the output characteristics of 14, 25 and 90 nm gate length FD SOI devices in which parameter is the energy relaxation time. We see strong dependence of the on-current upon the choice of the energy relaxation time for the smallest structure being simulated which suggests that proper determination of the energy relaxation time is needed. The energy relaxation time, in turn, is bias and geometry dependent parameter and its exact determination is impossible. The inability to properly determine the energy relaxation time in hydrodynamic/energy balance models has been the main motivation for the development of particle-based simulators discussed in Chapters 3 and 4.

3. Bulk Monte Carlo method

In the previous section we have considered continuum methods of describing transport in semiconductors, specifically the drift-diffusion and hydrodynamic models, which are derived from moments of the semi-classical Boltzmann Transport Equation (BTE). As approximations to the BTE, it was shown that in the case of small devices (see Section 2.3 above), such approaches become inaccurate, or fail completely. Indeed, one can envision that, as physical dimensions are reduced, at some level a continuum description of current breaks down, and the granular nature of the individual charge particles constituting the charge density in the active device region becomes important.

The microscopic simulation of the motion of individual particles in the presence of the forces acting on them due to external fields as well as the internal fields of the crystal lattice and other charges in the system has long been popular in the chemistry community, where *molecular dynamics* simulation of atoms and molecules have long been used to investigate the thermodynamic properties of liquids and gases. In solids, such as semiconductors and

metals, transport is known to be dominated by random scattering events due to impurities, lattice vibrations, etc., which randomize the momentum and energy of charge particles in time. Hence, stochastic techniques to model these random scattering events are particularly useful in describing transport in semiconductors, in particular the *Monte Carlo* method.
The Ensemble Monte Carlo techniques have been used for well over 30 years as a numerical method to simulate nonequilibrium transport in semiconductor materials and devices and has been the subject of numerous books and reviews [24,25,26]. In application to transport problems, a random walk is generated using the random number generating algorithms common to modern computers, to simulate the stochastic motion of particles subject to collision processes. This process of random walk generation is part of a very general technique used to evaluate integral equations and is connected to the general random sampling technique used in the evaluation of multi-dimensional integrals [27].
The basic technique as applied to transport problems is to simulate the free particle motion (referred to as the *free-flight*) terminated by instantaneous random *scattering events*. The Monte Carlo algorithm consists of generating random free flight times for each particle, choosing the type of scattering occurring at the end of the free flight, changing the final energy and momentum of the particle after scattering, and then repeating the procedure for the next free flight. Sampling the particle motion at various times throughout the simulation allows for the statistical estimation of physically interesting quantities such as the single particle distribution function, the average drift velocity in the presence of an applied electric field, the average energy of the particles, *etc*. By simulating an *ensemble* of particles, representative of the physical system of interest, the non-stationary time-dependent evolution of the electron and hole distributions under the influence of a time-dependent driving force may be simulated.
This particle-based picture, in which the particle motion is decomposed into free flights terminated by instantaneous collisions, is basically the same approximate picture underlying the derivation of the semi-classical Boltzmann Transport Equation (BTE). In fact, it may be shown that the one-particle distribution function obtained from the random walk Monte Carlo technique satisfies the BTE for a homogeneous system in the long-time limit [28]. This semi-classical picture breaks down when quantum mechanical effects become pronounced, and one cannot unambiguously describe the instantaneous position and momentum of a particle. In the following, we first describe the derivation of the free-flight scatter sequence using the path-integral method (section 3.1) and then we describe the standard Monte Carlo algorithm used to simulate charge transport in semiconductors (section 3.2). We then discuss how this basic model for charge transport within the BTE is self-consistently solved with the appropriate field equations to perform particle based device simulation (section 4).

3.1 Monte Carlo and path-integral methods

The path-integral method for solving the BTE is a rather a useful and an intuitive procedure for describing the Monte Carlo method. In its general form the BTE is:

$$\begin{aligned} &\frac{\partial f}{\partial t} + \vec{v}\cdot\nabla_r f + (-e)\vec{\varepsilon}\cdot\nabla_p f = \frac{\partial f}{\partial t}\Big|_{coll} = \\ &= \sum_{i=1}^{N}\sum_{\vec{p}'}[S_i(\vec{p}',\vec{p})f(\vec{r},\vec{p}',t) - S_i(\vec{p},\vec{p}')f(\vec{r},\vec{p},t)] \end{aligned} \quad . \tag{25}$$

The first term on the right-hand side (RHS) of Eq. (25) gives the scattering into state $\vec{p}$, while the second term on the RHS of Eq. (25) is the scattering out of state $\vec{p}$. This form of the BTE is valid for non-degenerate semiconductors. The collision integral on the RHS can also be expressed as:

$$RHS = \sum_{\vec{p}'} \sum_{i=1}^{N} [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t)] - f(\vec{r}, \vec{p}, t) \sum_{\vec{p}'} \sum_{i=1}^{N} S_i(\vec{p}, \vec{p}'), \tag{26}$$

where, $\frac{1}{\tau(\vec{p})} = \sum S_i(\vec{p}, \vec{p}')$ is the total scattering rate out of state $\vec{p}$.

Hence:

$$\begin{aligned} RHS &= \sum_{\vec{p}'} \sum_{i=1}^{N} [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t)] - f(\vec{r}, \vec{p}, t) \left[\frac{1}{\tau(\vec{p})} + \Omega(\vec{p}) \right] + f(\vec{r}, \vec{p}, t) \Omega(\vec{p}) = \\ &= \sum_{\vec{p}'} \sum_{i=1}^{N} [S_i(\vec{p}', \vec{p}) f(\vec{r}, \vec{p}', t)] - \Gamma(\vec{p}) f(\vec{r}, \vec{p}, t) + \sum_{\vec{p}'} f(\vec{r}, \vec{p}', t) \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'} \end{aligned}. \tag{27}$$

In this last expression we have added a term $\Omega(\vec{p})$ such that the total scattering rate out of a state p is constant. To understand the meaning of this term we need to go backwards, i.e. write $\Gamma(\vec{p})$ as:

$$\begin{aligned} \Gamma(\vec{p}) = \frac{1}{\tau(\vec{p})} + \Omega(\vec{p}) &= \sum_{\vec{p}'} \sum_{i=1}^{N} S_i(\vec{p}, \vec{p}') + \sum_{\vec{p}'} \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'} = \\ &= \sum_{\vec{p}'} \left[\sum_{i=1}^{N} S_i(\vec{p}, \vec{p}') + \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'} \right] \end{aligned}. \tag{28}$$

We now define an effective transition rate:

$$S_{eff}(\vec{p}, \vec{p}') = \sum_{i=1}^{N} S_i(\vec{p}, \vec{p}') + \Omega(\vec{p}) \delta_{\vec{p}, \vec{p}'}, \tag{29}$$

which consists of the sum of the N physical transition rates plus a term that has a momentum conserving δ-function. This second term has no effect on the carrier momentum and energy and it is a fictitious scattering process which is called self-scattering. The self-scattering can be calculated from:

$$\Omega(\vec{p}) = \Gamma(\vec{p}) - \frac{1}{\tau(\vec{p})}. \tag{30}$$

With the above definition for the self-scattering term, the BTE becomes:

$$\frac{\partial f}{\partial t} + \vec{v} \nabla_r f + (-e) \vec{\varepsilon} \nabla_p f + \Gamma(\vec{p}) f(\vec{r}, \vec{p}, t) = \sum_{\vec{p}'} \left[\sum_{i=1}^{N} S_i(\vec{p}', \vec{p}) + \left(\Gamma(\vec{p}) - \frac{1}{\tau(\vec{p})} \right) \delta_{\vec{p}, \vec{p}'} \right] f(\vec{r}, \vec{p}', t). \tag{31}$$

For homogenous samples, the BTE reduces to:

$$\begin{aligned}&\frac{\partial f}{\partial t}+\vec{v}\nabla_r f+(-e)\vec{\varepsilon}\nabla_p f+\Gamma(\vec{p})f(\vec{r},\vec{p},t)=\tilde{I}(\vec{p},t)=\\&=\sum_{\vec{p}'}\left[\sum_{i=1}^{N}S_i(\vec{p}',\vec{p})+\left(\Gamma(\vec{p})-\frac{1}{\tau(\vec{p})}\right)\delta_{\vec{p},\vec{p}'}\right]f(\vec{p}',t)\end{aligned}\quad . \tag{32}$$

In the last formulation of the BTE, the coordinate space (phase space) is fixed and the electrons move along given trajectory in response to the applied forces. With the introduction of variables:

$$\begin{cases}\tilde{t}=t\\ \tilde{p}=\vec{p}+e\vec{\varepsilon}t\end{cases}, \tag{33}$$

we go to a description in which electrons are frozen in their positions and the coordinate system is moving. Then,

$$\frac{\partial f}{\partial t}=\frac{\partial f}{\partial \tilde{t}}\cdot\frac{\partial \tilde{t}}{\partial t}+\frac{\partial f}{\partial \tilde{p}}\cdot\frac{\partial \tilde{p}}{\partial t}=\frac{\partial f}{\partial \tilde{t}}+e\vec{\varepsilon}\cdot\frac{\partial f}{\partial \tilde{p}}, \tag{34a}$$

$$\frac{\partial f}{\partial \vec{p}}=\frac{\partial f}{\partial \tilde{t}}\cdot\frac{\partial \tilde{t}}{\partial \vec{p}}+\frac{\partial f}{\partial \tilde{p}}\frac{\partial \tilde{p}}{\partial \vec{p}}=\frac{\partial f}{\partial \tilde{p}}. \tag{34b}$$

In this notation, the BTE becomes:

$$\frac{\partial f}{\partial \tilde{t}}+\Gamma\cdot f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=\tilde{I}(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t}). \tag{35}$$

The solution of the homogenous equation of the form:

$$\frac{\partial f}{\partial \tilde{t}}+\Gamma\cdot f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=0, \tag{36}$$

can be found using a separation of variables method, to be:

$$\ln f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})\Big|_0^{\tilde{t}}=-\Gamma\int_0^{\tilde{t}}d\tilde{t}, \tag{37a}$$

or:

$$\ln\left[f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})/f(\tilde{p},0)\right]=-\Gamma\tilde{t}, \tag{37b}$$

to get:

$$f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=f(\tilde{p},0)e^{-\Gamma\tilde{t}}. \tag{38}$$

Going back to the original coordinate system gives:

$$f(\vec{p},t)=f(\vec{p}+e\vec{\varepsilon}t,0)e^{-\Gamma t}. \tag{39}$$

This term is the transient term. It states that an electron initially in a state $(\vec{p}+e\vec{\varepsilon}t)$ at time *t=0,* has arrived in a state $\vec{p}$ at time *t* without scattering. This event occurs with a transition probability:

$$P(\vec{p},t,0)=e^{-\Gamma t}\text{ , for }\Gamma(\vec{p})=\Gamma=const. \tag{40}$$

For general case, when $\Gamma(\vec{p})$ is not a constant, one would have had:

$$P(\vec{p},t,0)=\exp\left[-\int_0^t\Gamma(\vec{p})dt'\right]. \tag{41}$$

Also, if the initial momentum of the electron is $(\vec{p}+e\vec{\varepsilon}t)$, because all of the drift motion and the acceleration by the electric field, the final electron momentum at time *t* equals to: $\vec{p}'=\vec{p}+e\vec{\varepsilon}t-e\vec{\varepsilon}t=\vec{p}$.

The next task is to find a solution of the BTE for homogenous systems. The homogenous solution suggests that the general solution will also involve an exponentials. For this purpose we define a function:

$$f_1(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})e^{\Gamma\tilde{t}}, \tag{42a}$$

which leads to:

$$f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=f_1(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})e^{-\Gamma\tilde{t}}. \tag{42b}$$

Then:

$$\frac{\partial f}{\partial\tilde{t}}=\frac{\partial f_1}{\partial\tilde{t}}e^{-\Gamma\tilde{t}}+\Gamma\cdot f_1(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})e^{-\Gamma\tilde{t}}. \tag{43}$$

Substituting this result back into the BTE gives:

$$\frac{\partial f_1}{\partial\tilde{t}}=e^{\Gamma\tilde{t}}\tilde{I}(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t}). \tag{44}$$

Solving the last equation for f_1 finally leads to:

$$f_1(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})\Big|_0^{\tilde{t}}=\int_0^{\tilde{t}}d\tilde{t}_1\cdot e^{\Gamma\tilde{t}_1}\tilde{I}(\tilde{p}-e\vec{\varepsilon}\tilde{t}_1,\tilde{t}_1), \tag{45a}$$

or:

$$f_1(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=f_1(\tilde{p},0)+\int_0^{\tilde{t}}d\tilde{t}_1\cdot e^{\Gamma\tilde{t}_1}\tilde{I}(\tilde{p}-e\vec{\varepsilon}\tilde{t}_1,\tilde{t}_1). \tag{45b}$$

Multiplying the last equation by $e^{-\Gamma\cdot\tilde{t}}$ one gets:

$$f(\tilde{p}-e\vec{\varepsilon}\tilde{t},\tilde{t})=f(\tilde{p},0)\cdot e^{-\Gamma\tilde{t}}+\int_0^{\tilde{t}}d\tilde{t}_1\cdot e^{-\Gamma(\tilde{t}-\tilde{t}_1)}\tilde{I}(\tilde{p}-e\vec{\varepsilon}\tilde{t}_1,\tilde{t}_1)\cdot \tag{46}$$

Returning to the original coordinate system gives:

$$f(\vec{p},t) = f(\vec{p} + e\vec{\varepsilon}t, 0)e^{-\Gamma t} + \int_0^t dt_1 \cdot e^{-\Gamma(t-t_1)}\tilde{I}(\vec{p},t_1)\,, \tag{47}$$

where $\tilde{I}(\vec{p},t) = \sum_{\vec{p}'} S_{eff}(\vec{p}',\vec{p}) f(\vec{p}',t)$. Substituting this back gives:

$$f(\vec{p},t) = f(\vec{p} + e\vec{\varepsilon}t, 0)e^{-\Gamma t} + \int_0^t dt_1 \sum_{\vec{p}'} f(\vec{p}',t_1) S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}(t-t_1))e^{-\Gamma(t-t_1)}\,, \tag{48}$$

where:

- $f(\vec{p} + e\vec{\varepsilon}t, 0)e^{-\Gamma t}$ is the transient term;
- $f(\vec{p}',t_1)$ is the probability that at time t_1 a state $\vec{p}'$ is occupied by an electron;
- $S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}(t-t_1))$ is the transition rate (probability) from state $\vec{p}'$ to state $\vec{p} + e\vec{\varepsilon}(t-t_1)$.
- $e^{-\Gamma(t-t_1)}$ is the probability that an electron will not undergo collision event in interval $(t\text{-}t_1)$.

This last expression is known as Chambers-Rees path integral. Rees [29] innovation is the introduction of the fictitious scattering term. Ignoring the transient term, one can find the solution of the distribution function using the following iterative procedure that is obtained by time discretization, i.e. using $t=N\cdot\Delta t$ and $t_n=n\cdot\Delta t$. Then,

$$f_N(\vec{p}) = \Delta t \sum_{m=0}^{N-1} \sum_{\vec{p}'} f_m(\vec{p}') S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}(N-m)\Delta t)e^{-\Gamma(N-m)\Delta t}\,. \tag{49}$$

The two step procedure is then found by using $N=1$, which means that $t=\Delta t$, i.e.,

$$f_1(\vec{p}) = \Delta t \sum_{\vec{p}'} f_0(\vec{p}') S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}\Delta t)e^{-\Gamma\Delta t}\,, \tag{50}$$

where,

- $g_0(\vec{p} + e\vec{\varepsilon}\Delta t) = f_0(\vec{p}') \cdot S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}\Delta t)$ is the intermediate function that describes the occupancy of a state $\vec{p} + e\vec{\varepsilon}\Delta t$ at time $t=0$, which can be changed due to in-scattering events;
- $e^{-\Gamma\Delta t}$ is the probability that no scattering occurred within time integral Δt (free-flight).

Now assume that $t=2\Delta t$. This then gives:

$$\begin{aligned} f_2(\vec{p}) &= \Delta t \sum_{m=0}^{1} \sum_{\vec{p}'} f_m(\vec{p}') S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}(2-m)\Delta t)e^{-\Gamma(2-m)\Delta t} = \\ &= \Delta t \left\{ \sum_{\vec{p}'} f_0(\vec{p}') S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}(2\Delta t))e^{-\Gamma(2\Delta t)} + \right. \\ &\quad \left. + \sum_{\vec{p}'} f_1(\vec{p}') S_{eff}(\vec{p}',\vec{p} + e\vec{\varepsilon}\Delta t)e^{-\Gamma\Delta t} \right\} \end{aligned} \tag{51}$$

These examples suggest that the evaluation of $f_{n+1}(\vec{p})$ involves integration over trajectories and the exponential factors just give the probability that no scattering has occurred.

3.2 Bulk Monte Carlo method

According to the description provided in Section 3.1 above, in the bulk Monte Carlo method, particle motion is assumed to consist of free flights terminated by instantaneous scattering events, which change the momentum and energy of the particle after scattering. So, the first task is to generate free flights of random time duration for each particle. To simulate this process, the probability density, $P(t)$, is required, in which $P(t)dt$ is the joint probability that a particle will arrive at time t without scattering after a previous collision occurring at time $t = 0$, and then suffer a collision in a time interval dt around time t. The probability of scattering in the time interval dt around t may be written as $\Gamma[\mathbf{k}(t)]dt$, where $\Gamma[\mathbf{k}(t)]$ is the scattering rate of an electron or hole of wavevector $\mathbf{k}$. The scattering rate, $\Gamma[\mathbf{k}(t)]$, represents the sum of the contributions from each individual scattering mechanism, which are usually calculated quantum mechanically using perturbation theory, as described later. The implicit dependence of $\Gamma[\mathbf{k}(t)]$ on time reflects the change in $\mathbf{k}$ due to acceleration by internal and external fields. For electrons subject to time independent electric and magnetic fields, the time evolution of $\mathbf{k}$ between collisions is represented as

$$\mathbf{k}(t) = \mathbf{k}(0) - \frac{e(\mathbf{E} + \mathbf{v} \times \mathbf{B})t}{\hbar}, \tag{52}$$

where $\mathbf{E}$ is the electric field, $\mathbf{v}$ is the electron velocity and $\mathbf{B}$ is the magnetic flux density. In terms of the scattering rate, $\Gamma[\mathbf{k}(t)]$, the probability that a particle has not suffered a collision after a time t is given by $\exp\left(-\int_0^t \Gamma\left[\mathbf{k}\left(t'\right)\right]dt'\right)$. Thus, the probability of scattering in the time interval dt after a free flight of time t may be written as the joint probability

$$P(t)\,dt = \Gamma[\mathbf{k}(t)]\exp\left[-\int_0^t \Gamma\left[\mathbf{k}\left(t'\right)\right]dt'\right]dt\,. \tag{53}$$

Random flight times may be generated according to the probability density $P(t)$ above using, for example, the pseudo-random number generator implicit on most modern computers, which generate uniformly distributed random numbers in the range [0,1]. Using a direct method (see, for example [24]), random flight times sampled from $P(t)$ may be generated according to

$$r = \int_0^{t_r} P(t)dt\,, \tag{54}$$

where r is a uniformly distributed random number and t_r is the desired free flight time. Integrating Eq. (54) with $P(t)$ given by Eq. (53) above yields

$$r = 1 - \exp\left[-\int_0^{t_r} \Gamma\left[\mathbf{k}\left(t'\right)\right]dt'\right]. \tag{55}$$

Since 1-*r* is statistically the same as *r*, Eq. (55) may be simplified to

$$-\ln r = \int_0^{t_r} \Gamma\left[\mathbf{k}\left(t'\right)\right] dt' \,. \tag{56}$$

Eq. (56) is the fundamental equation used to generate the random free flight time after each scattering event, resulting in a random walk process related to the underlying particle distribution function. If there is no external driving field leading to a change of **k** between scattering events (for example in ultrafast photo-excitation experiments with no applied bias), the time dependence vanishes, and the integral is trivially evaluated. As noted in the previous section, in the general case where this simplification is not possible, it is expedient to introduce the so called self-scattering method [29], in which one introduces fictitious scattering mechanism whose scattering rate always adjusts itself in such a way that the total (self-scattering plus real scattering) rate is a constant in time

$$\Gamma = \Gamma\left[\mathbf{k}\left(t'\right)\right] + \Gamma_{self}\left[\mathbf{k}\left(t'\right)\right], \tag{57}$$

where $\Gamma_{self}[\mathbf{k}(t')]$ is the self-scattering rate (see the discussion in Section 3.1 above). The self-scattering mechanism itself is defined such that the final state before and after scattering is identical. Hence, it has no effect on the free flight trajectory of a particle when selected as the terminating scattering mechanism, yet results in the simplification of Eq. (56) such that the free flight is given by

$$t_r = -\frac{1}{\Gamma}\ln r \,. \tag{58}$$

The constant total rate (including self-scattering) Γ, must be chosen at the start of the simulation interval (there may be multiple such intervals throughout an entire simulation) so that it is larger than the maximum scattering encountered during the same time interval. In the simplest case, a single value is chosen at the beginning of the entire simulation (constant gamma method), checking to ensure that the real rate never exceeds this value during the simulation. Other schemes may be chosen that are more computationally efficient, and which modify the choice of Γ at fixed time increments [30].

The algorithm described above determines the random free flight times during which the particle dynamics is treated semi-classically. For the scattering process itself, we need the type of scattering (i.e. impurity, acoustic phonon, photon emission, etc.) which terminates the free flight, and the final energy and momentum of the particle(s) after scattering. The type of scattering which terminates the free flight is chosen using a uniform random number between 0 and Γ, and using this pointer to select among the relative total scattering rates of all processes including self-scattering at the final energy and momentum of the particle

$$\Gamma = \Gamma_{self}\left[n,\mathbf{k}\right] + \Gamma_1\left[n,\mathbf{k}\right] + \Gamma_2\left[n,\mathbf{k}\right] + \ldots \Gamma_N\left[n,\mathbf{k}\right], \tag{59}$$

with *n* the band index of the particle (or subband in the case of reduced-dimensionality systems), and **k** the wavevector at the end of the free-flight. This process is illustrated schematically in Figure 6.

Once the type of scattering terminating the free flight is selected, the final energy and momentum (as well as band or subband) of the particle due to this type of scattering must

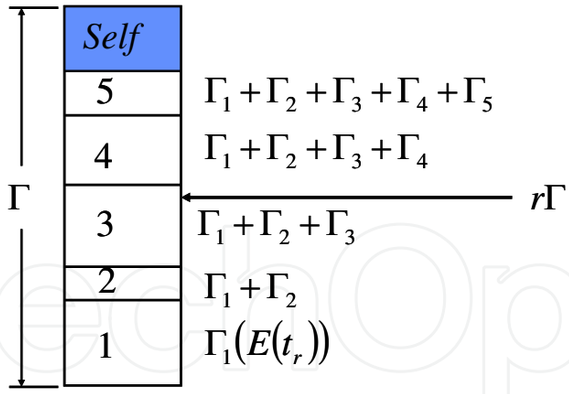


Fig. 6. Selection of the type of scattering terminating a free flight in the Monte Carlo algorithm.

be selected. For elastic scattering processes such as ionized impurity scattering, the energy before and after scattering is the same. For the interaction between electrons and the vibrational modes of the lattice described as quasi-particles known as phonons, electrons exchange finite amounts of energy with the lattice in terms of emission and absorption of phonons. For determining the final momentum after scattering, the scattering rate, $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$ of the jth scattering mechanism is needed, where n and m are the initial and final band indices, and $\mathbf{k}$ and $\mathbf{k}'$ are the particle wavevectors before and after scattering. Defining a spherical coordinate system around the initial wavevector $\mathbf{k}$, the final wavevector $\mathbf{k}'$ is specified by $|\mathbf{k}'|$ (which depends on conservation of energy) as well as the azimuthal and polar angles, φ and θ around $\mathbf{k}$. Typically, the scattering rate, $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$, only depends on the angle θ between $\mathbf{k}$ and $\mathbf{k}'$. Therefore, φ may be chosen using a uniform random number

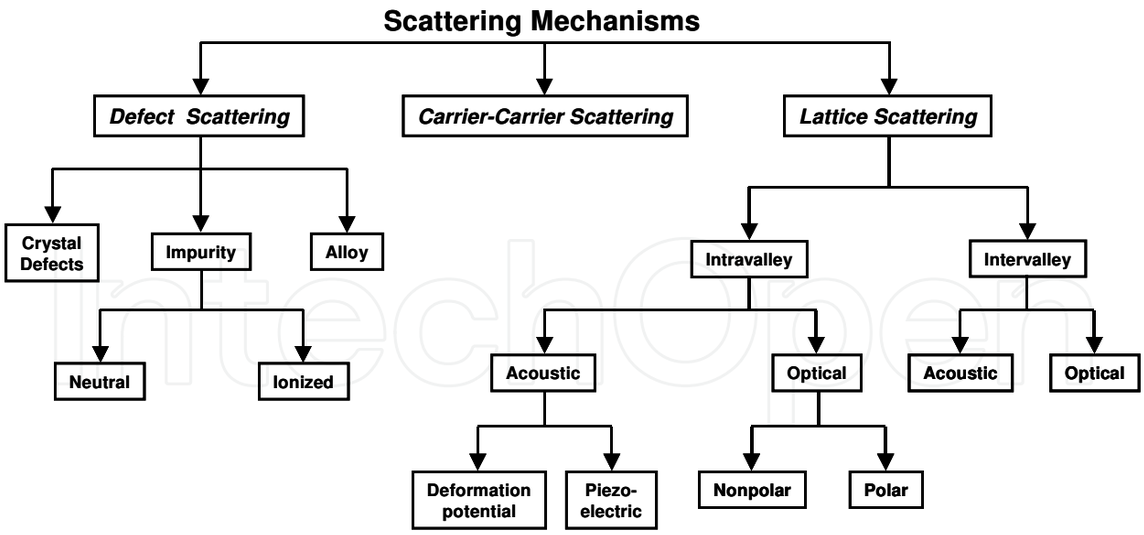


Fig. 7. Scattering mechanisms in a typical semiconductor.

between 0 and 2π (i.e. $2\pi r$), while θ is chosen according to the angular dependence for scattering arising from $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$. If the probability for scattering into a certain angle $P(\theta)d\theta$ is integrable, then random angles satisfying this probability density may be generated from a uniform distribution between 0 and 1 through inversion of Eq. (54).

Otherwise, a rejection technique (see, for example, [24,25]) may be used to select random angles according to $P(\theta)$. Scattering mechanisms that contribute to transport are summarized in Figure 7. The corresponding scattering rates for general non-parabolic bands are summarized in Table 2.

A general Monte Carlo code is developed as follows. First a subroutine is typically called that contains all material and scattering rates parameters for the scattering mechanisms included in the theoretical model. After the material and run parameters are read in, in the first step of the Monte Carlo simulation procedure it is necessary to construct scattering tables for the Γ, L and X valleys (for GaAs as a prototypical example) that initializes a series of events that are summarized in Figure 8. At each energy, the cumulative scattering rates for each valley are stored in separate look-up tables, and renormalized according to the maximum scattering rate (including self-scattering) that occurs over the range of energies stored. The structure of these subroutines is such that adding additional scattering event has to be trivial.

1. Acoustic Phonon Scattering

$$W(E)=\left(\frac{2\pi D_{ac}^2 K_B T_L}{\hbar C_l}\right)*\left(\frac{(2m_d)^{\frac{3}{2}}\sqrt{E(1+\alpha E)}}{4\pi^2\hbar^3}\right)*(1+2\alpha E)$$

2. Intervalley Phonon Scattering

$$W(E)=\left(\frac{\pi D_{ij}^2 Z_j}{\rho W_{ij}}\right)*\left(n(W_{ij})+\frac{1}{2}\mp\frac{1}{2}\right)*\left(\frac{(2m_d)^{\frac{3}{2}}\sqrt{E_f(1+\alpha E_f)}}{4\pi^2\hbar^3}\right)*(1+2\alpha E_f)$$

$$E_f = E\pm\hbar W_{ij}-\Delta E_{ij}$$

3. Ionized Impurity Scattering

$$W(E)=\left(\frac{\sqrt{2}e^4 N_I m_d^{\frac{3}{2}}}{\pi\varepsilon_s^2\hbar^4}\right)*\left(\sqrt{E(1+\alpha E)}*(1+2\alpha E)\right)*\left(\frac{1}{q_D^2\left(q_D^2+\frac{8m_d E(1+\alpha E)}{\hbar^2}\right)}\right)$$

$$q_D=\sqrt{\frac{e^2 N_I}{\varepsilon K_B T_L}}$$

4. Polar Optical Phonon Scattering

$$W(E)=\left(\frac{\sqrt{m_d}e^2 W_{LO}}{4\sqrt{2}\pi\hbar\varepsilon_p}\right)*\left(N_o+\frac{1}{2}\mp\frac{1}{2}\right)*\left(\frac{1+2\alpha E_k'}{\gamma_k}\right)*F(E_k,E_k')$$

$$N_o=\frac{1}{e^{\frac{\hbar W_{LO}}{K_B T_L}}-1}\qquad \varepsilon_p=\frac{1}{\frac{1}{\varepsilon_{high}}-\frac{1}{\varepsilon_{low}}}\qquad F(E_k,E_k')=\ln\left(mod\left(\frac{\sqrt{\gamma_k}+\sqrt{\gamma_{k'}}}{\sqrt{\gamma_k}-\sqrt{\gamma_{k'}}}\right)\right)$$

$$\gamma_k = E_k(1+\alpha E_k)$$

$$E_k' = E_k\pm\hbar W_{LO}$$

5. Piezoelectric Scattering
$$W(E)=\left(\frac{m_d^{\frac{1}{2}}K_BT_L}{4\sqrt{2}\pi\rho v_s^2\hbar^2}\right)*\left(\frac{1+2\alpha E}{\sqrt{E(1+\alpha E)}}\right)*\left(\frac{ee_{pz}}{\varepsilon_\infty}\right)^2*\ln\left(1+\frac{8m_dE(1+\alpha E)}{\hbar^2q_D^2}\right)$$ $$q_D=\sqrt{\frac{e^2N_l}{\varepsilon K_BT_L}}$$
6. Dislocation Scattering (e.g. GaN)
$$W(E)=\left(\frac{N_{dis}m_de^4}{4\hbar^3\varepsilon^2c^2}\right)*\left(\frac{\lambda^4}{\left(1+\frac{8\lambda^2m_dE(1+\alpha E)}{\hbar^2}\right)^{\frac{3}{2}}}\right)*\left(1+\frac{4\lambda^2m_dE(1+\alpha E)}{\hbar^2}\right)*(1+2\alpha E)$$ $$\lambda=\sqrt{\frac{\varepsilon K_BT_L}{e^2n'}}$$ where n' is the effective screening concentration N_{dis} is the Line dislocation density
7. Alloy Desorder Scattering ($Al_xGa_{1-x}As$)
$$W(E)=\left(\frac{x(1-x)a^3}{\pi}\right)*\left(\frac{D_{alloy}^2d}{\hbar^4}\right)*m_d\sqrt{2m_dE(1+\alpha E)}*(1+2\alpha E)$$ Where: d is the lattice disorder ($0\le d\le 1$) D_{alloy} is the alloy disorder scattering potential

Table 2. Scattering rates expressions for non-parabolic bands.

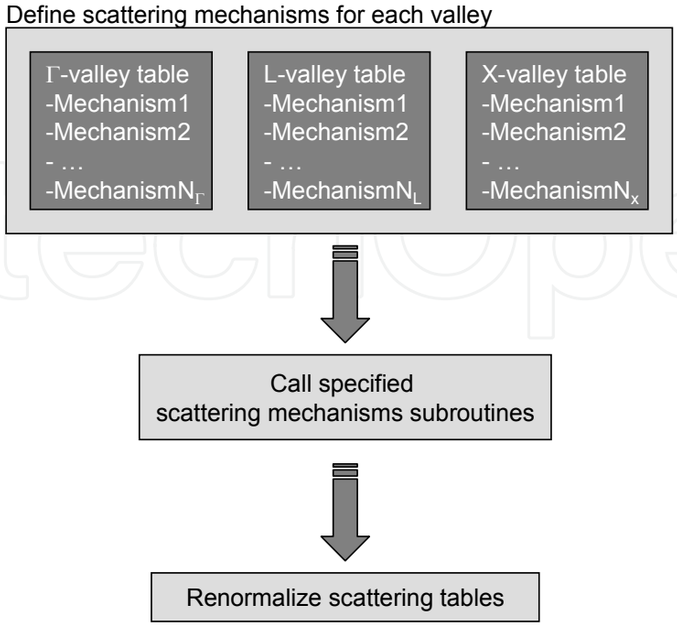


Fig. 8. Procedure for the creation of the scattering tables.

Having constructed the scattering table and after renormalizing the table, examples of which are given in Figure 9 and Figure 10 for the Γ, L, and X valley, the next step is to initialize carriers wavevector and energy and the initial free-flight time. This is accomplished by calling the initialization subroutine. Energy and wavevector histograms of the initial carrier energy and the components of the wave-vector along the x-, y-, and z-axes are shown in Figure 11. For good statistics, the number of particles simulated is 10000, and one can see the statistical fluctuation of these average quantities associated with the finite number of particles. Notice that the initial y-component for the wavevector is symmetric around the y-axis which means that the average wavevector along the y-axis is zero, which should be expected since the electric field along the y-component is zero at t=0. Identical distributions have been obtained for the x- and for the z-components of the wavevector. Also note that the energy distribution has the Maxwell-Boltzmann form as it should be expected. One can also estimate from this graph that the average energy of the carriers is on the order of $(3/2)k_{\mathrm{B}}T$.

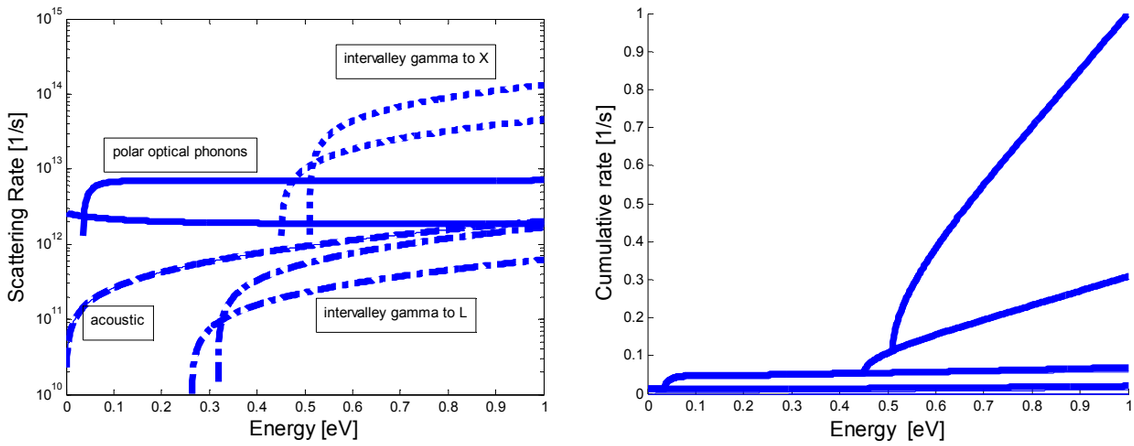


Fig. 9. Left panel: scattering rates for the Γ-valley. For simplicity we have omitted Coulomb scattering in these calculations. In the left figure, the dashed line corresponds to the acoustic phonon scattering rate, solid lines correspond to polar optical phonon scattering (absorption and emission), and the dashed-dotted line corresponds to intervalley scattering from Γ-valley to L-valley. Since the L-valley is along the [111] direction, there are 8 equivalent directions and since these valleys are shared there are a total of 4 equivalent L valleys. The dotted line corresponds to scattering from the Γ-valley to X-valleys. The X-valleys are at the [100] direction and since there are 6 equivalent [100] directions and the valleys are shared between Brillouin zones, there are 3 equivalent X valleys. Right panel: normalized cumulative scattering table for the Γ-valley. Everything above the top line up to Γ=1 is self-scattering so it is advisable when checking the scattering mechanisms to first check whether the scattering mechanism chosen is self-scattering or not. This is in particular important for energies below 0.5 eV for this particular scattering table when the Γ to X intervalley scattering (absorption and emission) takes over.

When the initialization process is finished, the main free-flight-scatter procedure takes place until the completion of the simulation time. There are two components in this routine; first the carriers accelerate freely due to the electric field, accomplished by calling the **drift()** subroutine, and then their free-flights are interrupted by random scattering events that are

managed by the **scatter_carrier()** subroutine. The flow-chart for performing the free-flight-scatter process within one time step Δt is shown diagrammatically in Figure 12.

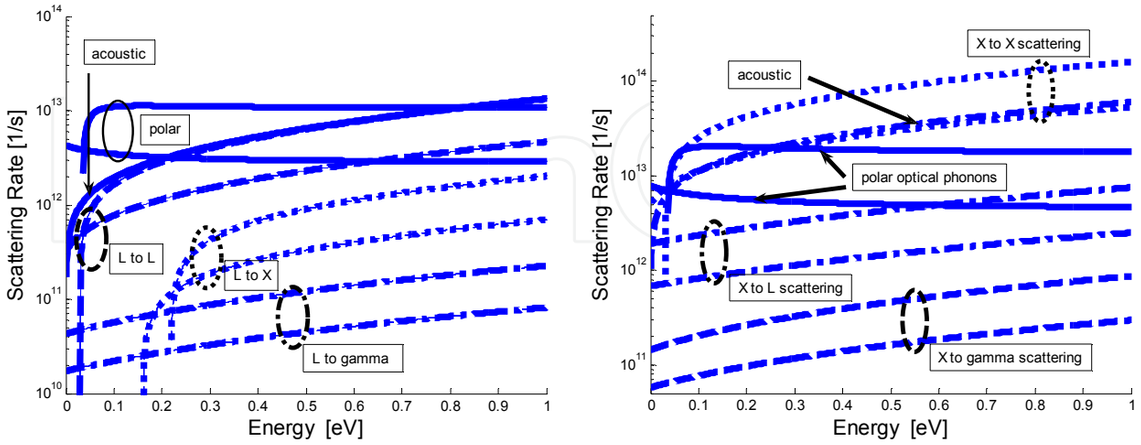


Fig. 10. Scattering rates for the L (left panel) and X (right panel) valleys used to create the corresponding normalized scattering tables (not shown here).

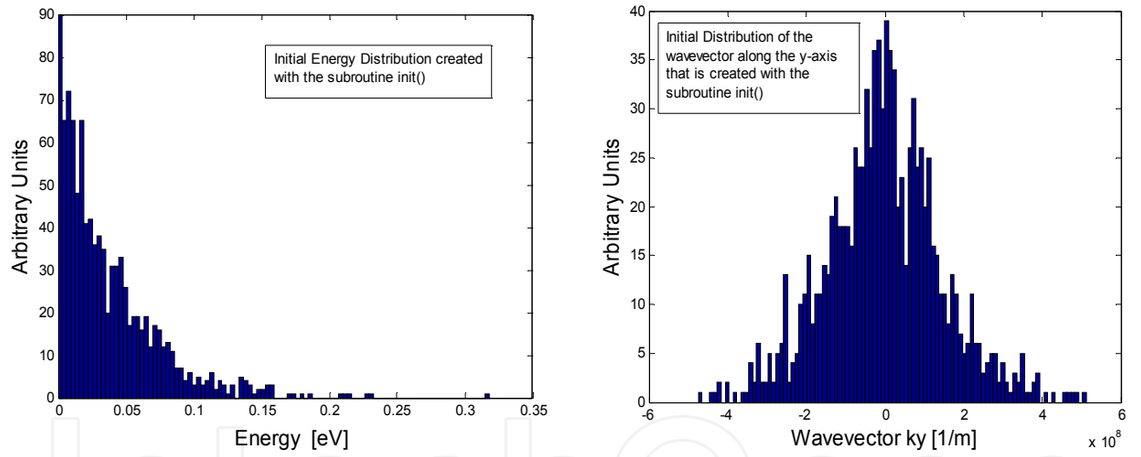


Fig. 11. Initial carrier distribution for an ensemble of 10000 Particles. Left panel: distribution of wavevector k_y. Right panel: energy distribution.

In the **scatter_carrier()** subroutine, first the scattering mechanism terminating the free flight is chosen, to which certain attributes are associated such as the change in energy after scattering. For inelastic scattering processes, we have the change in energy due to emission or absorption of phonons, for example. Also, the nature of the scattering process is identified: isotropic or anisotropic. Note that when performing acoustic phonon and intervalley scattering for GaAs, both of which are isotropic scattering processes, no coordinate system transformation is needed to determine the final wavevector after scattering. Because polar optical phonon and Coulomb scattering mechanisms are anisotropic, it is necessary to do a rotation of the coordinate system, scatter the carrier in the rotated system and then perform inverse coordinate transformation. This procedure is needed because it is much easier to determine final carrier momentum in the rotated

coordinate system in which the initial wavevector k is aligned with the z-axis. For this case, one can calculate that the final polar angle for scattering with polar optical phonons for parabolic bands in the rotated coordinate system is

$$\cos\theta = \frac{(1+\xi)-(1+2\xi)^{r}}{\xi}, \quad \xi = \frac{2\sqrt{E_k\left(E_k \pm \hbar\omega_0\right)}}{\left(\sqrt{E_k}-\sqrt{E_k \pm \hbar\omega_0}\right)^2} \tag{60}$$

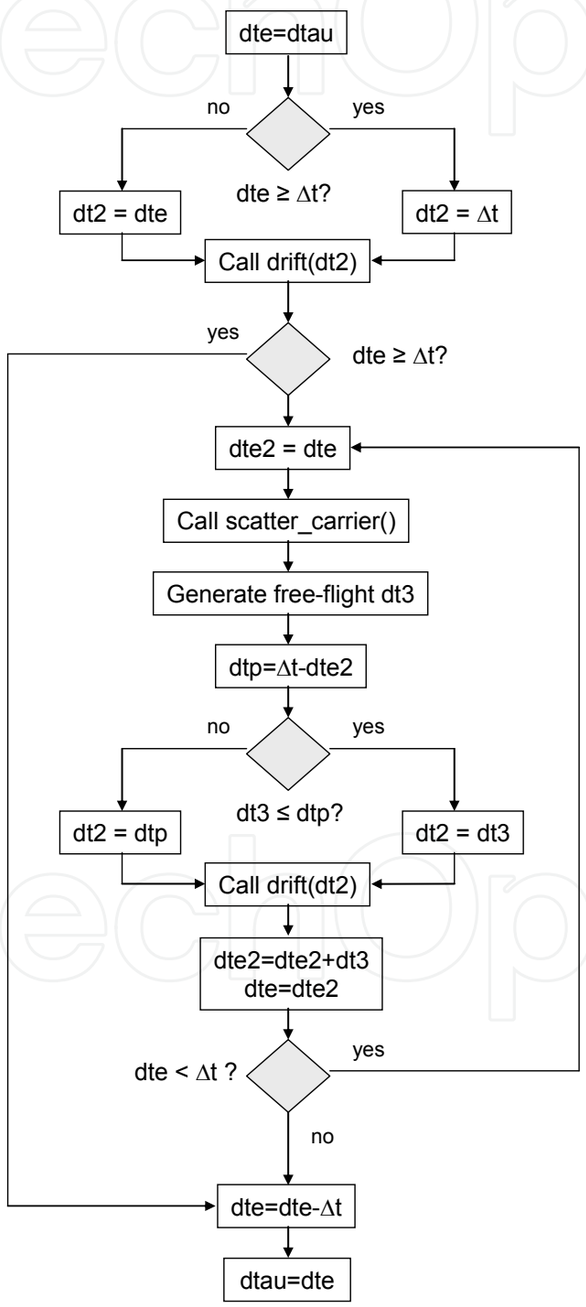


Fig. 12. Free-flight-scatter procedure within one time step.

where E_k is the carrier energy, $\hbar\omega_0$ is the polar optical phonon energy and r is a random number uniformly distributed between 0 and 1. The final angle for scattering with ionized impurities (Coulomb scattering) and for parabolic bands is

$$\cos\theta = 1 - \frac{2r}{1 + 4k^2 L_D^2 (1-r)} \tag{61}$$

where **k** is the carrier wavevector, and L_D is the Debye screening length. The azimuthal angle for both scattering processes is simply calculated using $\varphi = 2\pi r$. The importance of properly calculating the angle θ after scattering to describe small angle deflections in the case of Coulomb or polar optical phonon scattering is illustrated in Figure 13 (from 0 to π=3.141592654) where we plot the histogram of the polar angle after scattering for electron-polar optical phonon scattering, where we can clearly see the preference for small angle deflections that are characteristic for any Coulomb type interaction (polar optical phonon is in fact electron-dipole interaction). Graphical representation of the determination of the final angle after scattering for both isotropic and anisotropic scattering processes is given in Figure 14.

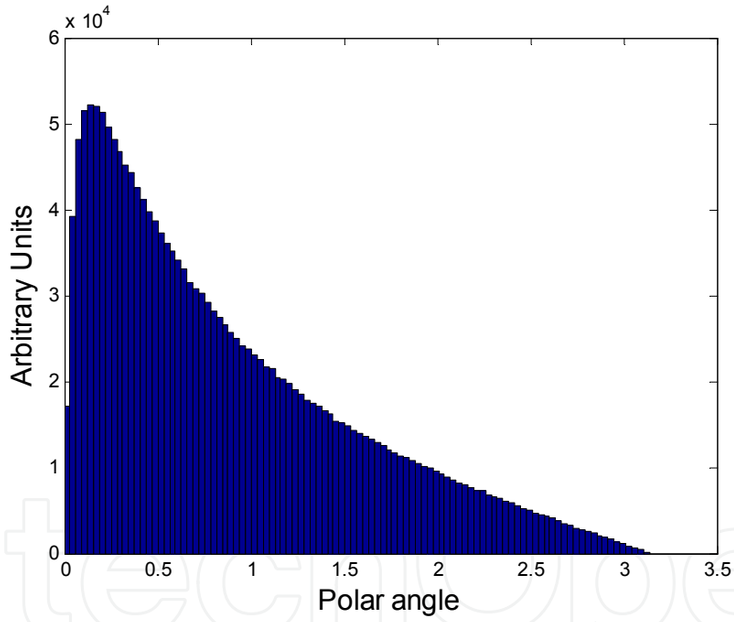


Fig. 13. Histogram of the polar angle for electron – polar optical phonon scattering.

The direct technique described above can be applied when the integrals describing cosθ can be analytically calculated. For most cases of interest, the integral cannot be easily inverted. In these cases a rejection technique may be employed. The procedure of the rejection technique goes as follows:

- Choose a maximum value C, such that $C > f(x)$ for all x in the interval (a,b).
- Choose pairs of random numbers, one between a and b ($x_1 = a + r_1(b-a)$) and another $f_1 = r_1' C$ between 0 and C, where r_1 and r_1' are random numbers uniformly distributed between zero and 1.
- If $f_1 \le f(x_1)$, then the number x_1 is accepted as a suitable value, otherwise it is rejected.

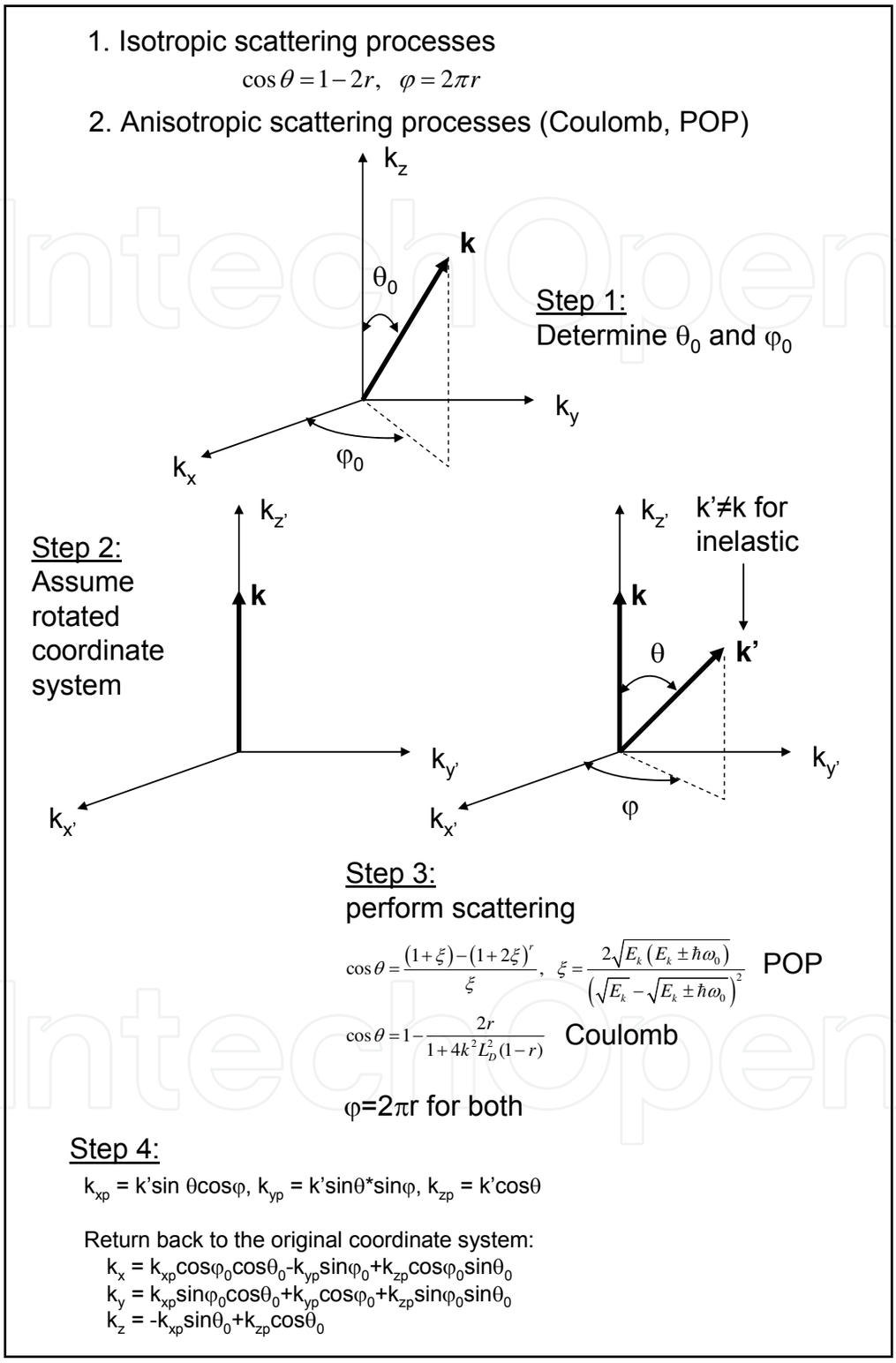


Fig. 14. Description of final angle selection for isotropic and anisotropic scattering processes using the direct technique.

The three steps described above are schematically shown in the figure below (Figure 15). For $x = x_1$, r_1C is larger than $f(x_1)$ and in this case if this represents the final polar angle for scattering, this angle is rejected and a new sequence of two random numbers is generated to determine x_2 and r_2C. In this second case, $f(x_2)>r_2C$ and the polar angle $\theta=x_2$ is selected (for polar angle selection $a = 0$ and $b=\pi$).

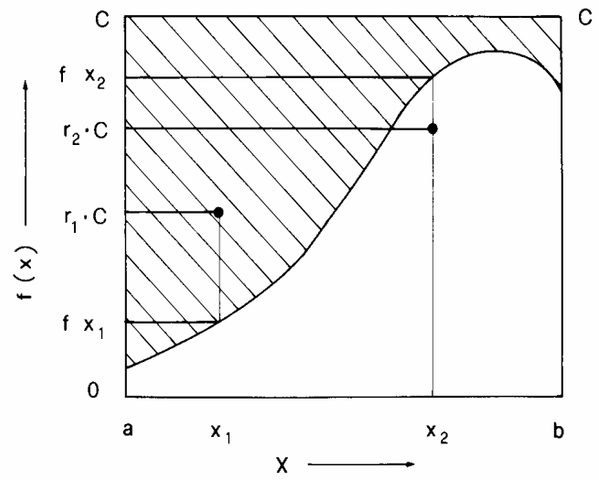


Fig. 15. Schematic description of the rejection technique.

After the simulation is completed, typical results to check are the velocity-time, the energy-time and the valley occupation versus time characteristics, such as those shown in Figure 16, where the velocity time characteristics for applied electric fields ranging from 0.5 to 7 kV/cm, with an electric field increment of 0.5 kV/cm, are shown. These clearly demonstrate that after a transient phase, the system reaches a stationary steady state, after which time we can start taking averages for calculating steady-state quantities.

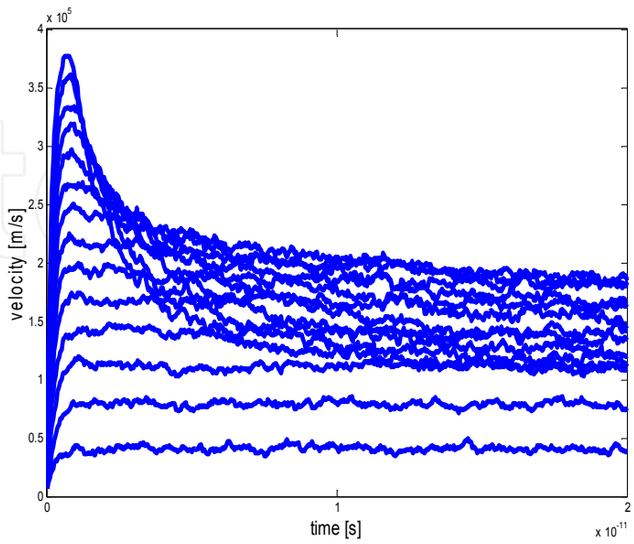


Fig. 16. Time evolution of the drift velocity for electric field strengths ranging between 0.5 and 7 kV/cm, in 0.5 kV/cm increments.

From the results shown in Figure 16, one can see that steady-state is achieved for larger time intervals when the electric field value is increased and the carriers are still sitting in the Γ-valley. Afterwards the time needed to get to steady-state decreases. This trend is related to the valley repopulation and movement of the carriers from the Γ, into the X and finally into the L valley. The steady-state velocity-field and valley population versus electric field characteristics are shown in Figure 17 and Figure 18, respectively. One can clearly see on the velocity-field characteristics that a low-field mobility of about 8000 $cm^2/V\text{-}s$ is correctly reproduced for GaAs without the use of any adjustable parameters.

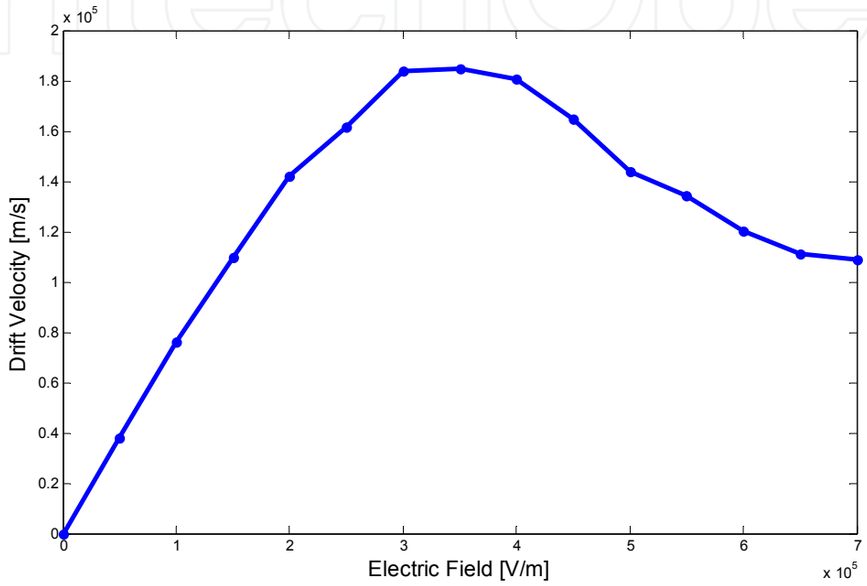


Fig. 17. Steady state drift velocity vs. electric field.

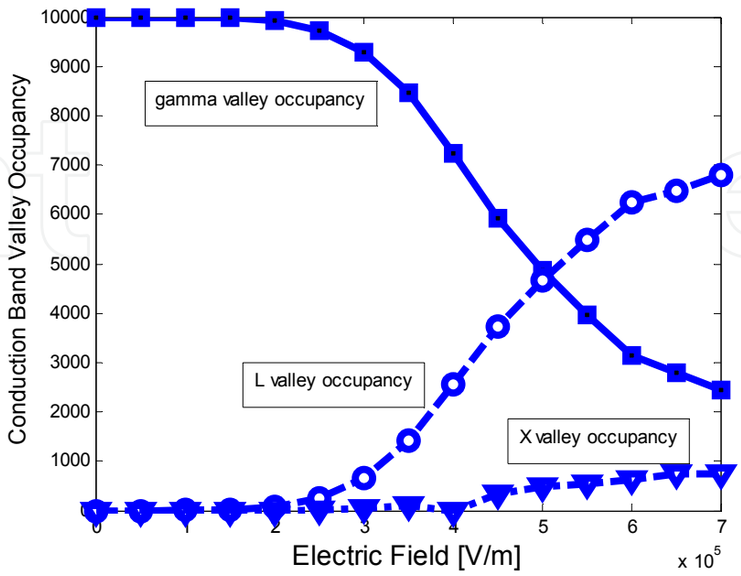


Fig. 18. Different valley occupancy vs. electric field.

At this point, it is advisable to check the energy and wavevector histograms (Figure 19) to ensure that the energy range chosen in the scattering tables is correct or not for the particular maximum electric field strength being considered, which gives the worst case scenario. Since, as already noted, we apply the electric field in the y-direction, for comparative purposes we plot the histograms of the x-component of the wavevector, y-component of the wavevector, and the histogram of the final carrier energy distribution for which a drifted Maxwellian form is evident. Since there is no field applied in the x-direction, we see that the average wavevector in the x-direction is 0. Due to the application of the field in the y-direction, there is a finite positive shift in the y-component of the velocity, which is yet another signature for the displaced Maxwellian form of the energy distribution in the bottom histogram.

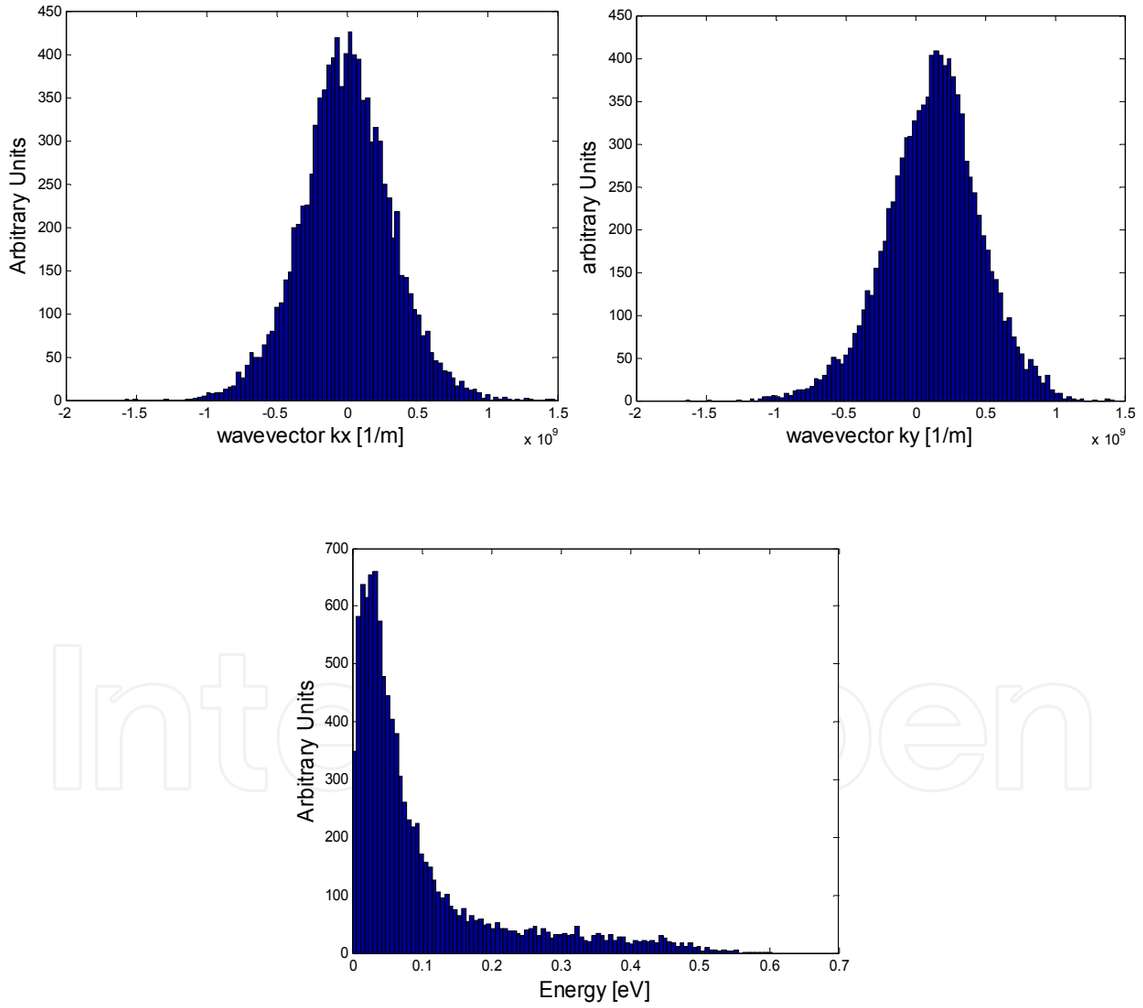


Fig. 19. Top left panel: histogram of the x-component of the wavevector. Top right panel: Histogram of the y-component of the wavevector. Bottom panel: histogram of the carrier energy. Applied electric field is 7kV/cm.

4. Particle-based device simulation

In Section 3.2, we introduced the numerical solution of the BTE using Monte Carlo method. Within a device, both the transport kernel and the field solver are coupled to each other (see Figure 2). The field associated with the potential coming from Poisson's equation is the driving force accelerating particles in the Monte Carlo phase, for example, while the distribution of mobile (both electrons and holes) and fixed charges (e.g. donors and acceptors) provides the source of the electric field in Poisson's equation. Below we give an extensive description of the Monte Carlo particle-based device simulators with emphasis on the particle-mesh coupling.

Within the particle-based EMC method with its time-marching algorithm, Poisson's equation may be decoupled from the BTE over a suitably small time step (typically less than the inverse plasma frequency corresponding to the highest carrier density in the device). Over this time interval, carriers accelerate according to the frozen field profile from the previous time-step solution of Poisson's equation, and then Poisson's equation is solved at the end of the time interval with the frozen configuration of charges arising from the Monte Carlo phase (see discussion in Ref. [45]). Note that Poisson's equation is solved on a mesh, whereas the solution of charge motion using EMC occurs over a continuous range of coordinate space in terms of the particle position. Therefore, a particle-mesh (PM) coupling is needed for both the charge assignment and the force interpolation. The PM coupling is broken into four steps: (1) assign particle charge to the mesh; (2) solve the Poisson equation on the mesh; (3) calculate the mesh-defined forces; and (4) interpolate to find forces on the particle. There are a variety of schemes that can be used for the PM coupling and these are discussed in Section 4.4.

The motion in real space of particles under the influence of electric fields is somewhat more complicated due to the band structure. The velocity of a particle in real space is related to the *E*-**k** dispersion relation defining the bandstructure as

$$\begin{aligned} \mathbf{v}(t) &= \frac{d\mathbf{r}}{dt} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E\left(\mathbf{k}(t)\right) \\ \frac{d\mathbf{k}}{dt} &= \frac{q\mathbf{E}(\mathbf{r})}{\hbar} \end{aligned} \tag{62}$$

where the rate of change of the crystal momentum is related to the local electric field acting on the particle through the acceleration theorem expressed by the second equation. In turn, the change in crystal momentum, **k**(*t*), is related to the velocity through the gradient of *E* with respect to **k.** If one has to use the full band-structure of the semiconductor, then integration of these equations to find **r**(*t*) is only possible numerically, using for example a Runge-Kutta algorithm. If a three valley model with parabolic bands is used, then the expression is integrable.

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{\hbar\mathbf{k}}{m^*}; \quad \frac{d\mathbf{k}}{dt} = \frac{q\mathbf{E}(\mathbf{r})}{\hbar} \tag{63}$$

Therefore, for a constant electric field in the *x* direction, the change in distance along the *x* direction is found by integrating twice and is given by equation

$$x(t) = x(0) + v_x(0)\,t + \frac{qE_x^0 t^2}{2m^*} \tag{64}$$

To simulate the steady-state behavior of a device, the system must be initialized in some initial condition, with the desired potentials applied to the contacts, and then the simulation proceeds in a time stepping manner until steady-state is reached. This process may take several picoseconds of simulation time, and consequently several thousand time-steps based on the usual time increments required for stability. Clearly, the closer the initial state of the system is to the steady state solution, the quicker the convergence. If one is, for example, simulating the first bias point for a transistor simulation, and has no a priori knowledge of the solution, a common starting point for the initial guess is to start out with charge neutrality, i.e. to assign particles randomly according to the doping profile in the device and based on the super-particle charge assignment of the particles, so that initially the system is charge neutral on the average. For two-dimensional device simulation, one should keep in mind that each particle actually represents a rod of charge into the third dimension. Subsequent simulations at the same device at different bias conditions can use the steady state solution at the previous bias point as a good initial guess. After assigning charges randomly in the device structure, charge is then assigned to each mesh point using the NGP or CIC or NEC particle-mesh methods, and Poisson's equation solved. The forces are then interpolated on the grid, and particles are accelerated over the next time step. A flow-chart of a typical Monte Carlo device simulation is shown in Figure 20.

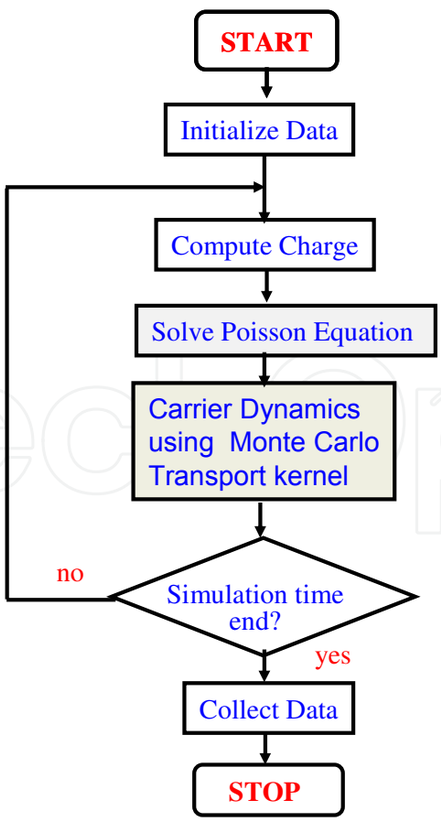


Fig. 20. Flow-chart of a typical particle based device simulation.

As the simulation evolves, charge will flow in and out of the contacts, and depletion regions internal to the device will form until steady state is reached. The charge passing through the contacts at each time step can be tabulated, and a plot of the cumulative charge as a function of time gives the steady-state current. Figure 21 shows the particle distribution in 3D of a MESFET, where the dots indicate the individual simulated particles for two different gate biases. Here, the heavily doped MESFET region (shown by the inner box) is surrounded by semi-insulating GaAs forming the rest of the simulation domain. The upper curve corresponds to no net gate bias (i.e. the gate is positively biased to overcome the built-in potential of the Schottky contact), while the lower curve corresponds to a net negative bias applied to the gate, such that the channel is close to pinch-off. One can see the evident depletion of carriers under the gate under the latter conditions.

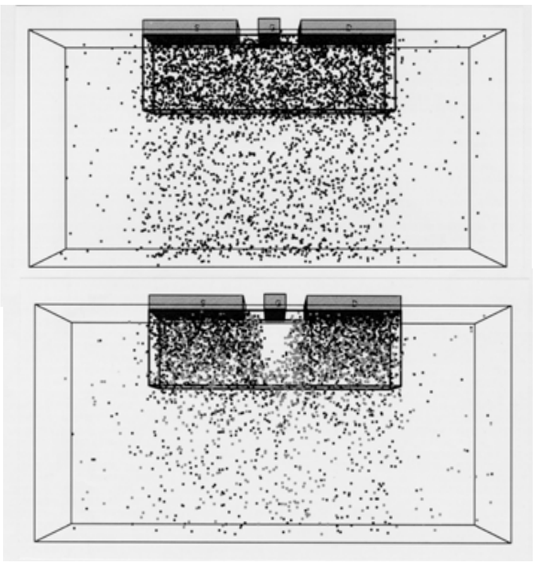


Fig. 21. Example of the particle distribution in a MESFET structure simulated in 3D using an EMC approach. The upper plot is the device with zero gate voltage applied, while the lower is with a negative gate voltage applied, close to pinch-off.

4.1 Calculation of the current

The device output current can be determined using two different yet consistent methods. *First*, by keeping track of the charges entering and exiting each terminal/contact, the net number of charges over a period of the simulation can be used to calculate the terminal current. The net charge crossing a terminal boundary is determined by

$$\boldsymbol{Q}(t) = \boldsymbol{e}\left(\boldsymbol{n}_{abs}(t) - \boldsymbol{n}_{injec}(t)\right) + \varepsilon \int \boldsymbol{E}_y(\boldsymbol{x},t)\boldsymbol{dy}, \tag{65}$$

where n_{abs} is the number of particles that are absorbed by the contact (exit), n_{injec} is the number of particles that have been injected at the contact, E_y is the vertical field at the contact. The second term in Eq. (65) on the right-hand-side is used to account for the displacement current due to the changing field at the contact. Eq. (65) assumes the contact is at the top of the device and that the fields in the x and z direction are negligible. The charge e in Eq. (65) should be multiplied by the particle charge if it is not unity. The slope of

$Q(t)$ versus time gives a measure of the terminal current. In steady state, the current can be found by

$$I = \frac{dQ(t)}{dt} = \frac{e(n_{net})}{\Delta t}, \tag{66}$$

where n_{net} is the net number of particles exiting the contact over a fixed period of time Δt. The method is quite noisy, due to the discrete nature of the electrons. An example of calculation of the current and keeping the ohmic contacts charge neutral is given Figure 22.

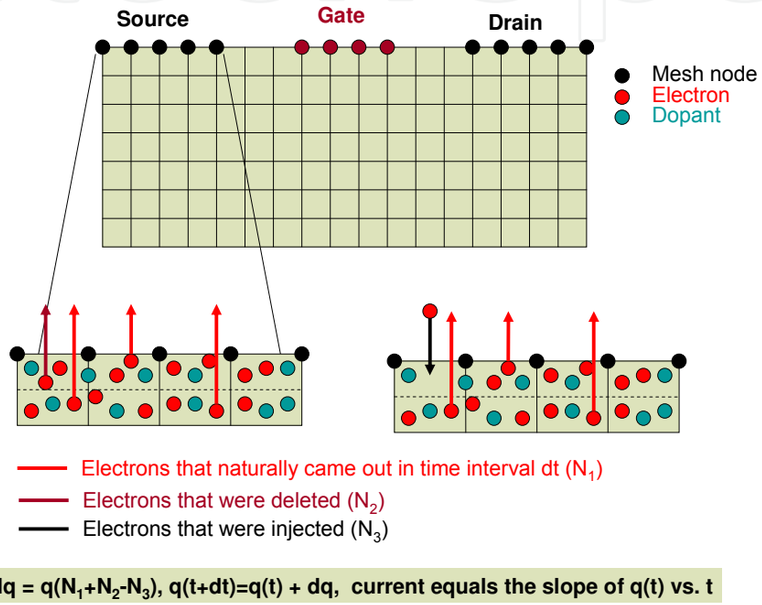


Fig. 22. Keeping charge neutrality at the ohmic contacts and contributions of various terms to the current.

In a *second* method, the sum of the electron velocities in a portion of the channel region of the device is used to calculate the current. The electron current density through a cross-section of the device is given by

$$J = env_d, \tag{67}$$

where v_d is the average electron drift velocity and n is the carrier concentration. If there are a total of N particles in a differential volume, $dV = dL \cdot dA$, the current found by integrating Eq. (67) over the cross-sectional area, dA, is

$$I = \frac{eNv_d}{dL}, \text{ or } I = \frac{e}{dL}\sum_{i=1}^{N} v_x(i), \tag{68}$$

where $v_x(i)$ is the velocity along the channel of the i^{th} electron. The device is divided into several sections along the x-axis, and the number of electrons and their corresponding velocity is added for each section after each free-flight. The total x-velocity in each section is

then averaged over several timesteps to determine the current for that section. Total device current can be determined from the average of several sections, which gives a much smoother result compared to counting terminal charges. By breaking the device into sections, individual currents can be compared to verify that there is conservation of particles (constant current) throughout the device. In addition, sections near the source and drain regions may have a high *y*-component in their velocity and should be excluded from the current calculations. Finally, by using several sections in the channel, the average energy and velocity of electrons along the channel can be observed to ensure the proper physical characteristics. The two methods for the calculation of the current are illustrated in Figure 23 on the example of a 50 nm channel length MOSFET device.

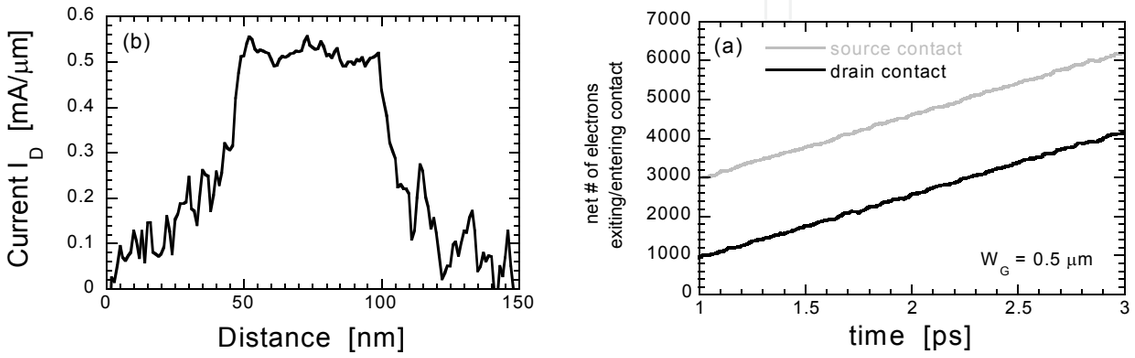


Fig. 23. (Left panel) Net charge entering/exiting the source/drain contact. (Right panel) Average current along the channel. The gate-length of the device being modeled equals 50 nm. We use V_G = 1.4 V and V_D = 1 V in these simulations.

Extrapolating the slope of the curve shown in Figure 23 (left panel), that represents the cumulative electron charge that enters/exits the source/drain contact, leads to source/drain current of 0.5205/0.5193 mA/μm. When compared with the results shown in Figure 23 (right panel), it is evident that both the current measurement techniques discussed in this section give current values with relative error less than 2 %.

4.2 Ohmic contacts

Another issue that has to be addressed in particle-based simulations is the real space boundary conditions for the particle part of the simulation. Reflecting boundary conditions are usually imposed at the artificial boundaries. As far as the ohmic contacts are concerned, they require more careful consideration because electrons crossing the source and drain contact regions contribute to the corresponding terminal current. Commonly employed models for the contacts include [31]:

- Electrons are injected at the opposite contact with the same energy and wavevector **k**. If the source and drain contacts are in the same plane, as in the case of MOSFET simulations, the sign of **k**, normal to the contact will change. This is an unphysical model, however [32].
- Electrons are injected at the opposite contact with a wavevector randomly selected based upon a thermal distribution. This is also an unphysical model.
- Contact regions are considered to be in thermal equilibrium. The total number of electrons in a small region near the contact are kept constant, with the number of

electrons equal to the number of dopant ions in the region. This is a very good model most commonly employed in actual device simulations.

- Another method uses 'reservoirs' of electrons adjacent to the contacts. Electrons naturally diffuse into the contacts from the reservoirs, which are not treated as part of the device during the solution of Poisson's equation. This approach gives results similar to the velocity weighted Maxwellian [31], but at the expense of increased computational time due to the extra electrons simulated. It is an excellent model employed in few most sophisticated particle-based simulators.

There are also several possibilities for the choice of the distribution function — Maxwellian, displaced Maxwellian, and velocity-weighted Maxwellian [33].

4.3 Time step

As in the case of solving the Drift-Diffusion, Hydrodynamic or full Maxwell's equations, for a stable Monte Carlo device simulation, one has to choose the appropriate time step, Δt, and the spatial mesh size (Δx, Δy, and/or Δz). The time step and the mesh size may correlate to each other in connection with the numerical stability. For example, as discussed in the context of solving Drift-Diffusion simulations, the time step Δt must be related to the plasma frequency

$$\omega_p = \sqrt{\frac{e^2 n}{\varepsilon_s m^*}}, \tag{69}$$

where n is the carrier density. From the viewpoint of the stability criterion, Δt must be much smaller than the inverse plasma frequency. The highest carrier density specified in the device model is used to estimate Δt. If the material is a multi-valley semiconductor, the smallest effective mass to be experienced by the carriers must be used in Eq. (69) as well. In the case of GaAs, with the doping of 5×10^{17} cm^{-3}, $\omega_p \cong 5\times10^{13}$; hence, Δt must be smaller than 0.02 ps.

The mesh size for the spatial resolution of the potential is dictated by the charge variations. Hence, one has to choose the mesh size to be smaller than the smallest wavelength of the charge variations. The smallest wavelength is approximately equal to the Debye length, given as

$$\lambda_D = \sqrt{\frac{\varepsilon_s k_B T}{e^2 n}}. \tag{70}$$

The highest carrier density specified in the model should be used to estimate λ_D from the stability criterion. The mesh size must be chosen to be smaller than the value given by Eq. (70). In the case of GaAs, with the doping density of 5×10^{17} cm^{-3}, $\lambda_D \cong 6$ nm.

Based on the discussion above, the time step (Δt), and the mesh size (Δx, Δy, and/or Δz) can be specified separately. However, the Δt chosen must be checked again by calculating the distance l_{max}, defined as

$$l_{\max} = \mathbf{v}_{\max} \times \Delta t, \tag{71}$$

where $\mathbf{v}_{max}$ is the maximum carrier velocity that can be approximated by the maximum group velocity of the electrons in the semiconductor (on the order of 10^8 cm/s). Therefore, the distance l_{max} is regarded as the maximum distance the carriers can propagate during Δt. The time step chosen must be small enough so that l_{max} is smaller than the spatial mesh size chosen using Eq. (71). This is because large Δt chosen may cause substantial change in the charge distribution, while the field distribution in the simulation is only updated every Δt.

4.4 Particle-mesh (PM) coupling

As mentioned earlier, the position of charge as described by the EMC algorithm is continuous, whereas Poisson's equation is solved on a mesh, hence the charge associated with the individual particles must be mapped onto the field mesh in some fashion. The charge assignment and force interpolation schemes usually employed in self-consistent Monte Carlo device simulations are the nearest-grid-point (NGP) and the cloud-in-cell (CIC) schemes [35]. In the NGP scheme, the particle position is mapped into the charge density at the closest grid point to a given particle. This has the advantage of simplicity, but leads to a noisy charge distribution, which may exacerbate numerical instability. Alternately, within the CIC scheme a finite volume is associated with each particle spanning several cells in the mesh, and a fractional portion of the charge per particle is assigned to grid points according to the relative volume of the 'cloud' occupying the cell corresponding to the grid point. This method has the advantage of smoothing the charge distribution due to the discrete charges of the particle based method, but may result in an artificial 'self-force' acting on the particle, particularly if an inhomogeneous mesh is used. The particle-mesh coupling sequence is presented in Figure 24.

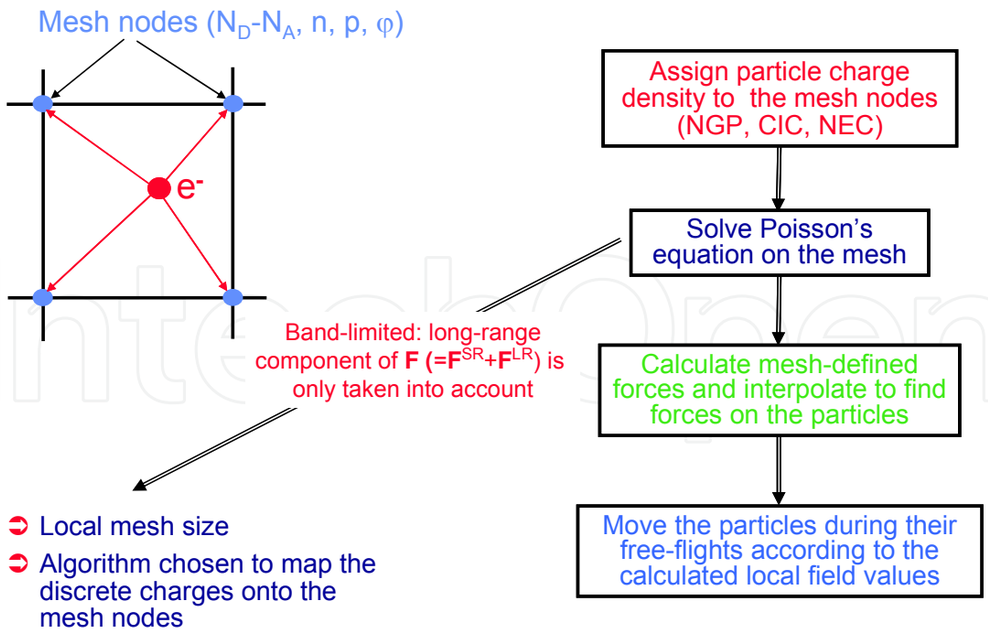


Fig. 24. Particle-mesh coupling sequence.

To better understand the NGP and the CIC scheme, consider a tensor product mesh with mesh lines $x_i, i = 1,\ldots,N_x$ and $y_j, j = 1,\ldots,N_y$. If the mesh is uniformly spaced in each axis

direction, then $(x_{l+1} - x_l) = (x_{l+2} - x_{l+1})$. The permittivities are considered constant within each mesh element and are denoted by ε_{kl}, $k = 1,\ldots,N_x - 1$ and $l = 1,\ldots,N_y - 1$. Define centered finite-differences of the potential ψ in the *x*- and *y*-axis at the midpoints of element edges as follows:

$$\begin{cases} \Delta^x_{k+\frac{1}{2},l} = -\dfrac{\psi_{k+1,l} - \psi_{k,l}}{x_{k+1} - x_k}, \\ \Delta^y_{k,l+\frac{1}{2}} = -\dfrac{\psi_{k,l+1} - \psi_{k,l}}{y_{l+1} - y_l}, \end{cases} \tag{72}$$

where the minus sign is included for convenience because the electric field is negative of the gradient of the potential. Consider now a point charge in 2-D located at (x, y) within an element $\langle i, j \rangle$. If the restrictions for the permittivity (P) and the tensor-product meshes with uniform spacing in each direction (M) apply, the standard NGP/CIC schemes in two dimensions can be summarized by the following four steps:

Charge assignment to the mesh: The portion of the charge ρ_L assigned to the element nodes (*k,l*) is $w_{kl}\rho_L$, *k*=*i*, *i*+1 and *l*=*j*, *j*+1, where w_{kl} are the four charge weights which sum to unity by charge conservation. For the NGP scheme, the node closest to (*x,y*) receives a weight w_{kl}=1, with the remaining three weights set to zero. For the CIC scheme, the weights are $w_{ij} = w_x w_y$, $w_{i+1,j} = (1 - w_x) w_y$, $w_{i,j+1} = w_x (1 - w_y)$, and $w_{i+1,j+1} = (1 - w_x)(1 - w_y)$, $w_x = (x_{i+1} - x)/(x_{i+1} - x_i)$ and $w_y = (y_{j+1} - y)/(y_{j+1} - y_j)$.

Solve the Poisson equation: The Poisson equation is solved by some of the numerical techniques discussed in Ref. [34].

Compute forces on the mesh: The electric field at mesh nodes (*k,l*) is computed as: $E^x_{kl} = \left(\Delta^x_{k-\frac{1}{2},l} + \Delta^x_{k+\frac{1}{2},l} \right)/2$ and $E^y_{kl} = \left(\Delta^y_{k,l-\frac{1}{2}} + \Delta^y_{k,l+\frac{1}{2}} \right)/2$, for k = *i*, *i*+1 and l=*j*, *j*+1.

Interpolate to find forces on the charge: Interpolate the field to position (*x,y*) according to $E^x = \sum_{kl} w_{kl} E^x_{kl}$ and $E^y = \sum_{kl} w_{kl} E^y_{kl}$, where *k* = *i*, *i*+1, *l* =*j*, *j*+1 and the w_{ij} are the NGP or CIC weights from step 1.

The requirements (P) and (M) severely limit the scope of devices that may be considered in device simulations using the NGP and the CIC schemes. Laux [35] proposed a new particle-mesh coupling scheme, namely, the nearest-element-center (NEC) scheme, which relaxes the restrictions (P) and (M). The NEC charge assignment/force interpolation scheme attempts to reduce the self-forces and increase the spatial accuracy in the presence of nonuniformly spaced tensor-product meshes and/or spatially-dependent permittivity. In addition, the NEC scheme can be utilized in one axis direction (where local mesh spacing is nonuniform) and the CIC scheme can be utilized in the other (where local mesh spacing is uniform). Such hybrid schemes offer smoother assignment/interpolation on the mesh compared to the pure NEC. The new steps of the pure NEC PM scheme are:

1. *Charge assignment to the mesh:* Divide the line charge ρ_L equally to the four mesh points of the element $\langle i, j \rangle$.
2. *Solve the Poisson equation.*

Compute forces on the mesh: Calculate the fields $\Delta^x_{i+\frac{1}{2},l}$, l=*j*, *j*+1, and $\Delta^y_{k,j+\frac{1}{2}}$, k=*i*, *i*+1.

Interpolate to find force on the charge: Interpolate the field according to the following $E^x = \left(\Delta^x_{i+\frac{1}{2},j} + \Delta^x_{i+\frac{1}{2},j+1}\right)/2$ and $E^y = \left(\Delta^x_{i,j+\frac{1}{2}} + \Delta^x_{i+1,j+\frac{1}{2}}\right)/2$.

The NEC designation derives from the appearance, in step (1') of moving the charge to the center of its element and applying a CIC-like assignment scheme. The NEC scheme involves only one mesh element and its four nodal values of potential. This locality makes the method well-suited to non-uniform mesh spacing and spatially-varying permittivity. The interpolation and error properties of the NEC scheme are similar to the NGP scheme.

4.5 Higher order effects

Multi-particle effects relate to the interaction between particles in the system, which is a nonlinear effect when viewed in the context of the BTE, due to the dependence of such effects on the single particle distribution function itself. Most algorithms developed to deal with such effects essentially linearize the BTE by using the previous value of the distribution function to determine the time evolution of a particle over the successive time-step. Multi-carrier effects may range from simple consideration of the Pauli exclusion principle (which depends on the exact occupancy of states in the system), to single particle and collective excitations in the system. Inclusion of carrier-carrier interactions in Monte Carlo simulation has been an active area of research for quite some time and is briefly discussed below. Another carrier-carrier effect, that is of considerable importance when estimating leakage currents in MOSFET devices, is impact ionization, which is a pure generation process involving three particles (two electrons and a hole or two holes and an electron). The latter is also discussed below.

4.5.1 Pauli exclusion principle

The Pauli exclusion principle requires that the bare scattering rate be modified by a factor $1 - f_m(\mathbf{k}')$ in the collision integral of the BTE, where $f_m(\mathbf{k}')$ is the one-particle distribution function for the state $\mathbf{k}'$ in band (subband) *m* after scattering. Since the net scattering rate including the Pauli exclusion principle is always less than the bare scattering rate, a self-scattering rejection technique may be used in the Monte Carlo simulation as proposed by Bosi and Jacoboni [36] for one particle simulation and extended by Lugli and Ferry [37] for EMC. In the self-scattering rejection algorithm, an additional random number *r* is generated (between 0 and 1), and this number is compared to $f_m(\mathbf{k}')$, the occupancy of the final state (which is also between 0 and 1 when properly normalized for the numerical **k**-space discretization). If *r* is greater than $f_m(\mathbf{k}')$, the scattering is accepted and the particle's momentum and energy are changed. If this condition is not satisfied, the scattering is rejected, and the process is treated as a self-scattering event with no change of energy or momentum after scattering. Through this algorithm, it is clear that no scattering occurs if the final state is completely full.

4.5.2 Carrier-carrier interactions

Carrier-carrier interactions, apart from degeneracy effects, may be treated as a scattering process within the Monte Carlo algorithm on the same footing as other mechanisms. In the simplest case of bulk electrons in a single parabolic conduction band, the process may be

treated as a binary collision where the scattering rate for a particle of wavevector $\mathbf{k}_0$ due to all the other particles in the ensemble is given by [38]

$$\Gamma_{ee}(\mathbf{k}_0) = \frac{nm_n e^4}{4\pi\hbar^3\varepsilon^2\beta^2}\int d\mathbf{k} f(\mathbf{k}) \frac{|\mathbf{k}-\mathbf{k}_0|}{\left(|\mathbf{k}-\mathbf{k}_0|^2+\beta^2\right)}, \tag{73}$$

where $f(\mathbf{k})$ is the one-particle distribution function (normalized to unity), ε is the permittivity, n is the electron density, and β is the screening constant. In deriving Eq. (73), one assumes that the two particles interact through a statically screened Coulomb interaction, which ignores the energy exchange between particles in the screening which in itself is a dynamic, frequency-dependent effect. Similar forms have been derived for electrons in 2D [39,40] and 1D [41], where carrier-carrier scattering leads to inter-subband as well as intra-subband transitions. Since the scattering rate in Eq. (73) depends on the distribution function of all the other particles in the system, this process represents a nonlinear term as discussed earlier. One method is to tabulate $f(\mathbf{k})$ on a discrete grid, as is done for the Pauli principle, and then numerically integrate Eq. (73) at each time step. An alternate method is to use a self-scattering rejection technique [42], where the integrand excluding $f(\mathbf{k})$ is replaced by its maximum value and taken outside the integral over $\mathbf{k}$. The integral over $f(\mathbf{k})$ is just unity, giving an analytic form used to generate the free flight. Then, the self-scattering rejection technique is used when the final state is chosen to correct for the exact scattering rate compared to this artificial maximum rate, similar to the algorithm used for the Pauli principle.

The treatment of intercarrier interactions as binary collisions above neglects scattering by collective excitations such as plasmons or coupled plasmon-phonon modes. These effects may have a strong influence on carrier relaxation, particularly at high carrier density. One approach is to make a separation of the collective and single particle spectrum of the interacting many-body Hamiltonian, and treat them separately, i.e. as binary collisions for the single particle excitations, and as electron-plasmon scattering for the collective modes [43]. Another approach is to calculate the dielectric response within the random phase approximation, and associate the damping given by the imaginary part of the inverse dielectric function with the electron lifetime [44].

A semiclassical approach to carrier-carrier interaction, which is fully compatible with the Monte Carlo algorithm, is the use of Molecular Dynamics [45], in which carrier-carrier interaction is treated continuously in real space during the free-flight phase through the Coulomb force of all the particles. A very small time step is required when using Molecular Dynamics to account for the dynamic distribution of the system. A time step on the order of 0.5 *fs* is often sufficiently small for this purpose. The small time step assures that the forces acting on the particles during the time of flight are essentially constant, that is $f(t) \cong f(t+\Delta t)$, where $f(t)$ is the single particle distribution function.

Using Newtonian kinematics, we can write the real space trajectories of each particle as

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}\Delta t + \frac{1}{2}\frac{\mathbf{F}(t)}{m}\Delta t^2, \tag{74}$$

and

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{m}\Delta t \cdot \tag{75}$$

Here, $\mathbf{F}(t)$ is the force arising from the applied field as well as that of the Coulomb interactions. We can write $\mathbf{F}(t)$ as

$$\mathbf{F}(t) = q\left[\mathbf{E} - \sum_i \nabla\varphi(\mathbf{r}_i(t))\right], \tag{76}$$

where $q\mathbf{E}$ is the force due to the applied field and the summation is the interactive force due to all particles separated by distance $\mathbf{r}_i$, with $\varphi(\mathbf{r}_i)$ the electrostatic potential. As in Monte Carlo simulation, one has to simulate a finite number of particles due to practical computational limitations on execution time. In real space, this finite number of particles corresponds to a particular simulation volume given a certain density of carriers, $V = N/n$, where n is the density. Since the carriers can move in and out of this volume, and since the Coulomb interaction is a long-range force, one must account for the region outside V by periodically replicating the simulated system. The contributions due to the periodic replication of the particles inside V in cells outside has a closed form solution in the form of an Ewald sum [46], which gives a linear as well as $1/r^2$ contribution to the force. The equation for the total force in the Molecular Dynamics technique then becomes

$$\mathbf{F} = \frac{-e^2}{4\pi\varepsilon}\sum_i^N\left(\frac{1}{\mathbf{r}_i^{\,2}}\mathbf{a}_i + \frac{2\pi}{3V}\mathbf{r}_i\right). \tag{77}$$

The above equation is easily incorporated in the standard Monte Carlo simulation discussed up to this point. At every time step the forces on each particle due to all the other particles in the system are calculated from Eq. (77). From the forces, an interactive electric field is obtained which is added to the external electric field of the system to couple the Molecular Dynamics to the Monte Carlo.

The inclusion of the carrier-carrier interactions in the context of particle-based device simulations is discussed in Ref. [47]. The main difficulty in treating this interaction term in device simulations arises from the fact that the long-range portion of the carrier-carrier interaction is included via the numerical solution of the quasi-static Poisson equation. Under these circumstances, special care has to be taken when incorporating the short-range portion of this interaction term to prevent double counting of the force.

4.5.3 Band to Band impact ionization

Another carrier-carrier scattering process is that of impact ionization, in which an energetic electron (or hole) has sufficient kinetic energy to create an electron-hole pair. Impact ionization therefore leads to the process of carrier multiplication. This process is critical for example in the avalanche breakdown of semiconductor junctions, and is a detrimental effect in short channel MOS devices in terms of excess substrate current and decreased reliability.

The ionization rate of valence electrons by energetic conduction band electrons is usually described by Fermi's rule in which a screened Coulomb interaction is assumed between the two particles, where screening is described by an appropriate dielectric function such as that proposed by Levine and Louie [48]. In general, the impact ionization rate should be a function of the wavevector of the incident electron, hence of the direction of an electric field in the crystal, although there is still some debate as to the experimental and theoretical evidence. More simply, the energy dependent rate (averaged over all wavevectors on a constant energy shell) may be expressed analytically in the power law form

$$\Gamma_{ii}(E) = P\left[E - E_{th}\right]^a, \tag{78}$$

where E_{th} is the threshold energy for the process to occur, which is determined by momentum and energy conservation considerations, but minimally is the bandgap of the material itself. P and a are parameters which may be fit to more sophisticated models. The Keldysh formula [49] is derived by expanding the matrix element for scattering close to threshold, which gives a=2, and the constant $P=C/E_{th}^2$, with C=1.19×10^{14}/s and assuming a parabolic band approximation,

$$E_{th} = \frac{3 - 2m_v / m_c}{1 - m_v / m_c} E_g, \tag{79}$$

where m_v and m_c are the effective masses of the valence and conduction band respectively, and E_g is the bandgap. More complete full-bandstructure calculations of the impact ionization rate have been reported for Si [50,51], GaAs [51,52] and wide bandgap materials [53], which are fairly well fit using using power law model.

Within the ensemble Monte Carlo method, the scattering rate given by Eq. (78) is used to generate the free flight time. The state after scattering of the initial electron plus the additional electron and hole must satisfy both energy and momentum conservation within the Fermi rule model, which is somewhat complicated unless simple parabolic band approximations are made.

5. What is the future for Monte Carlo device simulations?

The drift-diffusion solvers are applicable in situations in which the bias conditions and the device geometry are such that electric fields are relatively low and velocity saturation model is applicable. Situations in which drift-diffusion models are applicable are the silicon based power MOSFET devices, bipolar junction transistors, light emitting diodes (that are used more and more in solid state lightning) and crystalline solar cells, just to name a few. In some of these devices, such as power transistors and LEDs used for solid state lighting, it is of paramount importance to incorporate self-heating models within the drift-diffusion framework. The accuracy of simple heating models in conjunction with drift-diffusion models is to some degree questionable so that it is in many circumstances justifiable to use particle-based device simulators with more exact self-heating models.

On the other hand, the hydrodynamic models do not suffer from the limitations of the drift-diffusion approaches and the incorporation of the additional energy balance equation allows one to include velocity overshoot in the model. Velocity overshoot and non-stationary transport are key features of conventional MOSFET devices with gate lengths of 200 nm and below. However, as it was explained in Section 1.3 of this book chapter, the magnitude of the velocity overshoot observed via simulations depends strongly upon the choice of the energy relaxation time, mostly in sub-100 nm channel length devices. This, in turn, affects the magnitude of the drain current. The reason for such drastic differences in the results when different energy relaxation times are used is the fact that the energy relaxation time is material as well as device geometry dependent parameter. So, to calculate better estimates for the velocity overshoot, higher moments of the Boltzmann transport equation are needed. These, in turn involve parameters that are more and more ambiguous on the expense of increased computational cost and when the computational cost of hydrodynamic models exceeds the one of particle-based device simulators, there is no point in using moment

methods. In these circumstances the direct solution of the Boltzmann Transport Equation via the Monte Carlo method becomes a method of choice. Thus, we might conclude that it is advisable to use particle-based device simulators when nano-scale devices are concerned.
But how far down in the scaling can we go? Particle-based device simulators capture on an equal footing ballistic and diffusive transport, so if the ballisticity factor in the device increases, there is no problem that ballistic transport is effectively captured with particle-based device simulators. Quantum mechanical size quantization effects can also be captured by solving in slices the corresponding 1D or 2D Schrödinger equation if one is concerned with conventional or fully-depleted SOI MOSFETs, or nanowire transistors, respectively. What can not be captured with particle-based device simulators is if there are local strains and stresses in the ultra-nano-scale devices, but that can also be cured via coupling of Monte-Carlo device simulators with atomistic models for band-structure calculation.
In summary, Monte Carlo device simulators are a powerful tool for modeling devices ranging from the nano-scale regime to the microscale regime. What can not be modeled with particle-based device simulators are resonant tunneling diodes in which quantum interference effects dominate the device behavior. Efforts have been made along this direction as well, but the inclusion of the quantum-mechanical phase alongside with well defined particle trajectory still remains open field of research.

6. References

[1] David K. Ferry and Stephen M. Goodnick, *Transport in Nanostructures* (Cambridge Studies in Semiconductor Physics and Microelectronic Engineering, 1997).
[2] Vasileska and S. M. Goodnick, *Materials Science and Engineering, Reports: A Review: Journal*, R38, no. 5, 181 (2002).
[3] S. M. Goodnick and D. Vasileska, *Encyclopedia of Materials: Science and Technology*, Vol. 2, Ed. By K. H. J. Buschow, R. W. Cahn, M. C. Flemings, E. J. Kramer and S. Mahajan, Elsevier, New York, 1456, (2001).
[4] D. Vasileska and S. M. Goodnick, *Computational Electronics* (Morgan and Claypool, 2006).
[5] A. Schütz, S. Selberherr, H. Pötzl, *Solid-State Electronics*, Vol. 25, 177 (1982).
[6] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE* (McGraw-Hill, New York, 1988).
[7] M. Shur, *Physics of Semiconductor Devices* (Prentice Hall Series in Solid State Physical Electronics).
[8] D. L. Scharfetter and D. L. Gummel, *IEEE Transaction on Electron Devices*, Vol. ED-16, 64 (1969).
[9] K. Bløtekjær, *IEEE Trans. Electron Dev.*, Vol. 17, 38 (1970).c
[10] M. V. Fischetti and S. E. Laux, "Monte Carlo Simulation of Submicron Si MOSFETs", *Simulation of Semiconductor Devices and Processes*, vol. 3, G. Baccarani and M. Rudan Eds. (Technoprint, Bologna, 1988), 349.
[11] L. V. Keldysh, *Sov. Phys. – JETP*, Vol. 20, 1018 (1965).
[12] A. L. Fetter, J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill 1971).
[13] G. D. Mahan, *Many-Particle Physics* (Kluwer Academic/Plenum Publishers, New York, 2000).
[14] R. Lake, G. Klimeck, R.C. Bowen, and D. Jovanovic, *J. Appl. Phys.*, Vol. 81, 7845 (1997)
[15] G. Baccarani, M. Wordeman, *Solid State Electron.*, Vol. 28 , 407 (1985).
[16] S. Cordier, *Math. Mod. Meth. Appl. Sci.*, Vol. 4, 625 (1994).
[17] K. Tomizawa, *Numerical Simulation of Submicron Semiconductor Devices* (The Artech House Materials Science Library).

[18] H. K. Gummel, *IEEE Transactions on Electron Devices*, Vol. 11, 455 (1964).
[19] T. M. Apostol, *Calculus, Vol. II, Multi-Variable Calculus and Linear Algebra* (Blaisdell, Waltham, MA, 1969) ch. 1.
[20] R. Straton, *Phys. Rev.*, Vol. 126, 2002 (1962).
[21] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, *Proceedings of the IEEE*, Vol. 91, 251 (2003).
[22] M.A. Stettler, M.A. Alam, and M.S. Lundstrom, *Proceedings of the NUPAD Conference*, 97 (1992).
[23] www.silvaco.com
[24] C. Jacoboni and L. Reggiani, *Rev. Mod. Phys*., Vol. 55, 645 (1983).
[25] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag, Vienna (1989).
[26] K. Hess, *Monte Carlo Device Simulation: Full Band and Beyond*, (Kluwer Academic Publishing, Boston , 1991).
[27] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, (Wiley, New York, 1986).
[28] D. K. Ferry, *Semiconductors*, (Macmillan, New York, 1991).
[29] H. D. Rees, *J. Phys. Chem. Solids*, Vol. 30, 643 (1969).
[30] R. M. Yorston, *J. Comp. Phys*., Vol. 64, 177 (1986).
[31] T. Gonzalez and D. Pardo, *Solid State Electron*., 39 (1996) 555.
[32] P. A. Blakey, S. S. Cherensky and P. Sumer, *Physics of Submicron Structures*, Plenum Press, New York, (1984).
[33] T. Gonzalez and D. Pardo, *Solid-State Electron*., 39, 555 (1996).
[34] D. Vasileska and S.M. Goodnick, "Computational Electronics", *Morgan & Claypool*, 2006
[35] S. E. Laux, *IEEE Trans. Comp.-Aided Des. Int. Circ. Sys*., 15, 1266 (1996).
[36] S. Bosi S and C. Jacoboni, *J. Phys. C*, 9, 315 (1976).
[37] P. Lugli and D. K. Ferry, *IEEE Trans. Elec. Dev*., 32, 2431 (1985).
[38] N. Takenaka, M. Inoue and Y. Inuishi, *J. Phys. Soc. Jap*., 47, 861 (1979).
[39] S. M. Goodnick and P. Lugli, *Phys. Rev. B*, 37 (1988) 2578.
[40] M. Moško, A. Mošková and V. Cambel, *Phys. Rev. B*, 51, 16860 (1995).
[41] L. Rota, F. Rossi, S. M. Goodnick, P. Lugli, E. Molinari and W. Porod, *Phys. Rev. B*, 47,1632 (1993).
[42] R. Brunetti, C. Jacoboni, A. Matulionis and V. Dienys, *Physica B&C*, 134, 369 (1985).
[43] P. Lugli and D. K. Ferry, *Phys. Rev. Lett*., 56, 1295 (1986).
[44] J. F. Young and P. J. Kelly, *Phys. Rev. B*, 47, 6316 (1993).
[45] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, Institute of Physics Publishing, Bristol, (1988).
[46] D. J. Adams and G. S. Dubey, *J. Comp. Phys*., 72, 156 (1987).
[47] D. Vasileska, H.R. Khan, S.S. Ahmed, "Modeling Coulomb effects in nanoscale devices", Journal of Computational and Theoretical Nanoscience, Volume 5, Number 9, September 2008, pp. 1793-1827(35).
[48] Z. H. Levine, and S. G. Louie, *Phys. Rev. B*, 25, 6310 (1982).
[49] L. V. Keldysh, Zh. Eksp. *Teor. Fiz*., 37, 713 (1959).
[50] N. Sano and A. Yoshii, *Phys. Rev. B*, 45, 4171 (1992).
[51] M. Stobbe, R. Redmer and W. Schattke, *Phys. Rev. B*, 47, 4494 (1994).
[52] Y. Wang and K. Brennan, *J. Appl. Phys*., 71, 2736 (1992).
[53] M. Reigrotzki, R. Redmer, N. Fitzer, S. M. Goodnick, M. Dür, and W. Schattke, *J. Appl. Phys*., 86, 4458, (1999).

Applications of Monte Carlo Method in Science and Engineering
Edited by Prof. Shaul Mordechai

ISBN 978-953-307-691-1
Hard cover, 950 pages
Publisher InTech
Published online 28, February, 2011
Published in print edition February, 2011

In this book, Applications of Monte Carlo Method in Science and Engineering, we further expose the broad range of applications of Monte Carlo simulation in the fields of Quantum Physics, Statistical Physics, Reliability, Medical Physics, Polycrystalline Materials, Ising Model, Chemistry, Agriculture, Food Processing, X-ray Imaging, Electron Dynamics in Doped Semiconductors, Metallurgy, Remote Sensing and much more diverse topics. The book chapters included in this volume clearly reflect the current scientific importance of Monte Carlo techniques in various fields of research.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Dragica Vasileska, Katerina Raleva and Stephen M. Goodnick (2011). Monte Carlo Device Simulations, Applications of Monte Carlo Method in Science and Engineering, Prof. Shaul Mordechai (Ed.), ISBN: 978-953-307-691-1, InTech, Available from: http://www.intechopen.com/books/applications-of-monte-carlo-method-in-science-and-engineering/monte-carlo-device-simulations

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen