IL NUOVO CIMENTO Vol. 33 C, N. 1 Gennaio-Febbraio 2010
DOI 10.1393/ncc/i2010-10582-4

Colloquia: ICTT2009

A statistical enhancement method for Direct Simulation Monte Carlo in semiconductor devices

O. Muscato(*) and V. Di Stefano(**)

Dipartimento di Matematica e Informatica, Università di Catania
Viale A. Doria 6, 95125 Catania, Italy

(ricevuto il 5 Ottobre 2009; approvato il 29 Gennaio 2010; pubblicato online il 16 Marzo 2010)

Summary. — The Multicomb variance reduction technique has been introduced in the Direct Simulation Monte Carlo for submicrometric semiconductors. We have implemented the method in a silicon diode $n^+ - n - n^+$ and demonstrated its effectiveness. The steady-state statistical error and the figures of merit are obtained. The results of the simulations indicate that the method can enhance the high-energy distribution tail with a good accuracy.

PACS 72.20.-i – Conductivity phenomena in semiconductors and insulators.
PACS 05.10.Ln – Monte Carlo methods.
PACS 85.30.-z – Semiconductor devices.

1. – Introduction

Enhanced functional integration in modern electron devices requires an accurate modeling of energy transport in semiconductors in order to simulate the effects of high-field phenomena. The semiclassical Boltzmann Transport Equation (hereafter BTE), coupled with the Poisson equation, provides a general theoretical framework for transport phenomena in semiconductor devices. A statistical solution of the BTE can be obtained with the Direct Simulation Monte Carlo (DSMC) in which the distribution function is replaced with a representative set of particles. This solution is useful because it can provide information about regions of phase space which are too sparsely populated for fluid approximations to describe. Unfortunately, computational limits on the number of particles in a Monte Carlo simulation often lose this information in statistical noise. In order to make statistically significant statements about those regions of phase space which are of interest, "variance reduction" techniques which enhance statistical tallies (the particles) while maintaining physical distributions (through weighting) are often used.

(*) E-mail: muscato@dmi.unict.it
(**) E-mail: vdistefano@dmi.unict.it

In this paper we have studied the Multicomb variance reduction method, which assigns to the particles a statistical weight proportional to the number of physical charged carriers that a single computational particle represents [1, 2]. The plan of the paper is the following: in sect. **2** we introduce the Direct Simulation Monte Carlo. In sect. **3** the Multicomb algorithm is formulated as well as the definition of the Figure of Merit and the statistical error. Simulation results are shown for a silicon diode $n^+ - n - n^+$ in sect. **4**. Finally, conclusions are drawn in sect. **5**.

2. – The Direct Simulation Monte Carlo

Let be $f(\mathsf{t}, \mathsf{x}, \mathsf{k})$ the probability density to find an electron at time t, position x, with wave vector k and energy $\varepsilon(\mathsf{k})$, the Boltzmann transport equation describing transport fenomena in semiconductors writes [3]

$$\left[\frac{\partial}{\partial \mathsf{t}} + \mathsf{v}(\mathsf{k}) \cdot \nabla_{\mathsf{x}} + \frac{q}{\hbar} \nabla_{\mathsf{x}} \Phi \cdot \nabla_{\mathsf{k}}\right] f(\mathsf{t}, \mathsf{x}, \mathsf{k}) = \mathcal{Q}(f), \tag{1}$$

where $\mathcal{Q}(f)$ is the collisional operator, $\mathsf{v}(\mathsf{k})$ the electron (group) velocity, $\hbar$ denotes Planck's constant divided by 2π, q the absolute value of the electric charge, Φ the electric potential satisfying the Poisson equation

$$\epsilon \, \Delta_{\mathsf{x}} \Phi(\mathsf{t}, \mathsf{x}) = q \left[n(\mathsf{t}, \mathsf{x}) - N_D(\mathsf{x}) \right], \tag{2}$$

where N_D denotes the donor density, ϵ the permittivity, and n the electron density.

The Monte Carlo method for evolving a solution of the Boltzmann transport equation consists in recreating the history evolution of electrons in time and space inside the crystal, subject to the action of external and self-consistent electric field and of the given scattering mechanisms [3]. The simulation starts with one or more electrons in given initial conditions with momentum $\hbar\mathsf{k}$ and position x. During the free flight (*i.e.* the time between two collisions) the external forces are made to act according to Newton's equations of motion in the crystal, coupled with the Poisson equation for the electric field. The free-flight algorithm chosen is the self-scattering mechanism. The equations of motion are solved with a stable numerical scheme by using an appropriate time step $\Delta\mathsf{t}$ [4]. Then a scattering mechanism is chosen randomly as responsible for the end of the free flight, according to the relative probabilities of all possible scattering mechanisms. From the differential cross-section of this mechanism a new k state after scattering is randomly chosen as initial state of the new free flight.

The electrons can scatter by themselves, with the impurities and the lattice. The electron-electron interaction is considered in the framework of the mean-field approximation through the Poisson equation. This is reasonable since we consider the case of low doping and therefore we can neglect the short-range collisions between electrons. The scattering between electron and the lattice is quantized, and is modeled as a scattering with a particle called *phonon*. In our code we have considered the scattering with the acoustic (in the elastic approximation) and non-polar optical phonons, which are, at room temperature, the main scattering mechanisms in silicon [3].

3. – The algorithm

One of the most important classes of variance reduction techniques is the population control, where the particles distribution and the weights are adjusted to populate

interesting regions of the phase space. In this class we have used the so-called Multicomb algorithm which has been introduced, firstly, in the field of neutral particle transport [5]. This algorithm works under the hypothesis that the particles are independent.

In a semiconductor device the particles are not independent because the self-consistent electrostatic field correlates all particles through the Poisson equation. To overcome this problem, usually one runs the simulation with sufficient particles and for a sufficient time ($\simeq 5\,\mathrm{ps}$) to reach the steady-state regime, in which the fields fluctuate around their average values. In the case of a steady-state simulation with fixed electric field the particles are independent, and we can freely use variance reduction techniques developed for independent particles.

Let us suppose to have N initial particles with momentum $\hbar\mathsf{k}_i$, energy ε_i and weight w_i. At time zero we choose $w_i = 1/N$. Then the phase space is partitioned into K non-empty different regions (*e.g.*, the "stat-boxes"). For each j-th stat-box we can count:

– the number of particles N_j which belongs to the j-th stat-box, such that

$$N = \sum_{j=1}^{K} N_j; \tag{3}$$

– the total weight W_j, as the sum of the weights of the particles which belong to the j-th stat-box

$$W_j = \sum_i w_i, \quad j = 1, \ldots, K. \tag{4}$$

Then we fix the desired number of particles in the regions M_j ($j = 1, \ldots, K-1$) and consequently

$$M_K = N - \sum_{j=1}^{K-1} M_j.$$

The enhancement algorithm applied to each stat-box is called simple comb. In the simple comb, from an initial distribution of N_j particles a new distribution is produced formed by M_j particles with equal weights that preserves the expectation values of the original distribution. Let us consider the j-th stat-box

1) we construct a comb of length W_j with M_j equally spaced teeth. The position of the m-th tooth is given by

$$t_m = (u + m - 1)\frac{W_j}{M_j}, \quad m = 1, \ldots, M_j \tag{5}$$

where u is a uniform random number in $[0, 1)$.

2) Place the weights w_i of the particles which belong to the j-th stat-box consecutively on a line segment, obtaining N_j bins

$$[0, w_1], \left]w_1, \sum_{i=1}^{2} w_2\right], \left]\sum_{i=1}^{2} w_i, \sum_{i=1}^{3} w_i\right], \ldots, \left]\sum_{i=1}^{N_j-1} w_i, \sum_{i=1}^{N_j} w_i\right]. \tag{6}$$

3) Now we "comb" this line segment with the previous one, obtained in 1).

do $m = 1, \ldots, M_j$

if $t_m \in \left]\sum_{k=1}^{i-1} w_k, \sum_{k=1}^{i} w_k\right]$, for some $i = 1, \ldots, N_j$

a copy of the i-th particle from the old distribution is added to the new distribution, with assigned weight

$$w_i' = \frac{W_j}{M_j} \tag{7}$$

else no copy.

enddo

In this way the algorithm maintains constant, in the j-th stat-box, the total weight W_j and the desired particle number M_j. It is possible to prove that [1]

- the *simple comb* preserves on average the individual weights of the pre-combed particles;
- the distribution of particles with identical weights produced by the comb gives the smaller variance than any distribution with unequal weights.

The application of different *simple comb* to all stat-boxes is called *Multicomb* method. Consequently, during the simulation, the *Multicomb* maintains constant (by default) the total particles number N, and the sum of the overall weights

$$W = \sum_{j=1}^{K} W_j. \tag{8}$$

During the simulation in steady-state regime, we run the algorithm each "enhancement time step" Δt_{enh}, which usually is taken to be a multiple of the simulation time step Δt.

This algorithm produces three different effects: i) the improvement of the tail of the energy electron distribution function (hereafter EED); ii) a more CPU time consumption; iii) the introduction of some error in the fields.

To quantify the efficiency of the method we have introduced the *Figure of Merit*, that takes into account the relative error (RE) and the CPU time. To determine the EED function we count the sum of the weights of the particles $\xi(\varepsilon)$ which are in the shell $[\varepsilon, \varepsilon + \Delta\varepsilon]$. We define the Figure of Merit (FoM), for the EED, as

$$\text{FoM} = \frac{1}{(\text{RE})^2\, T_{\text{CPU}}}\,, \qquad \text{RE} = \frac{S_\xi}{\bar{\xi}} \tag{9}$$

with

$$\bar{\xi} = \frac{1}{N_r}\sum_{j=1}^{N_r} \xi_j, \qquad S_\xi = 3\sqrt{\frac{1}{N_r}\left(\frac{1}{N_r}\sum_{j=1}^{N_r}\xi_j^2 - \left[\frac{1}{N_r}\sum_{j=1}^{N_r}\xi_j\right]^2\right)}, \tag{10}$$

where the factor 3 corresponds to a 99.7% confidence level, N_r is the number of independent runs performed and ξ_j ($j = 1, \ldots, Nr$) are the quantities obtained in the runs. T_{CPU} is the total CPU time. Since the Relative Error is proportional to $1/\sqrt{Nr}$ and the time T_{CPU}, that takes to run Nr runs, is proportional to Nr, then the FoM is independent of Nr. Because the FoM is inversely proportional to the total CPU time, a method which approaches a given level of error faster will have a higher figure of merit. A large figure of merit indicates that a highly accurate energy distribution estimate can be calculated in a short time; a small figure of merit indicates that a longer simulation time is needed for an accurate estimate.

In order to study the error in the fields, we have introduced a measure called *Average Absolute Error* (AAE) as follows. Let consider some functionals (velocity, energy, density, ...) $F(x_j)$, where x_j, $j = 1, \ldots, C$, indicates the spatial cell and a reference solution $F^{\mathrm{ref}}(x_j)$, which represents the best available dataset obtained using a huge number of particles, and gathering statistics for a long time. For each cell x_j, the absolute error between the functional F_i and the reference solution $F^{\mathrm{ref}}(x_j)$ is measured. Then the average is taken over the cells, *i.e.*

$$\mathrm{AAE} = \frac{1}{C} \sum_{j=1}^{C} \left| F(x_j) - F^{\mathrm{ref}}(x_j) \right|. \tag{11}$$

For such functional, performing several independent runs, one can evaluate the confidence intervals.

4. – Results

We have tested the *Multicomb* algorithm in a silicon diode $n^+ - n - n^+$, 550 nm long, source and drain regions doped to a density of $10^{19}\,\mathrm{cm}^{-3}$, channel 250 nm long doped to a density of $10^{18}\,\mathrm{cm}^{-3}$.

In order to apply the Multicomb method and to determine the EED function, we divided the particles in four sets ($K = 4$) which form a partition of the phase space, as follows (see fig. 1):

- S_1: particles under the source contact ($0 \leq x < 150\,\mathrm{nm}$),
- S_2: particles under the drain contact ($400\,\mathrm{nm} \leq x < 550\,\mathrm{nm}$),
- S_3: the low-energy particles in the extended channel ($\varepsilon \leq 0.5\,\mathrm{eV}$, $150\,\mathrm{nm} \leq x < 400\,\mathrm{nm}$),
- S_4: the high-energy particles in the extended channel ($\varepsilon > 0.5\,\mathrm{eV}$, $150\,\mathrm{nm} \leq x < 400\,\mathrm{nm}$),

and in those sets, we fix the particle number (in percentage) $M_1 = 42\%$, $M_2 = 37\%$, $M_3 = 10\%$ and $M_4 = 11\%$.

For our simulations we have chosen four different points of the device, which are A = 270 nm, B = 330 nm, C = 390 nm, D = 415 nm, which belong to the different regions S_1, S_2, S_3 and S_4, as one can see in fig. 1. In this figure (left) we plot the average energy

$$\langle \varepsilon \rangle = \frac{1}{n} \int \varepsilon f(\mathsf{t}, \mathsf{x}, \mathsf{k}) \mathrm{d}\mathsf{k} \tag{12}$$

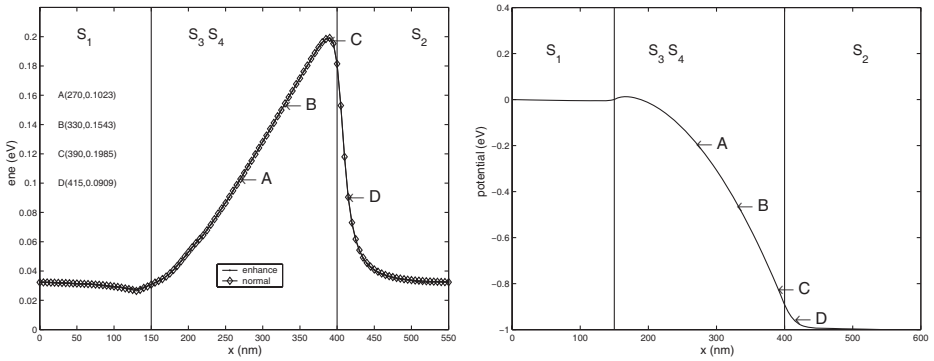

Fig. 1. – The average energy (12) in the device for the unenhanced and multicomb methods (left). The potential energy in the device (right).

for the enhancement method and for the normal (or unenhanced) one, revealing a perfect agreement. On the right of the same figure the potential energy ($-q\Phi$) is plotted along the device. In figs. 2–5 we plot the energy distribution function (on the left) and the FoM (on the right) obtained with the Multicomb algorithm and with the *normal* one for the four points. Figures 2, 3, 4 show that, with the Multicomb technique, the tails are more populated. Moreover, in the corresponding FoMs an additional cost of the Multicomb algorithm for low energies is evident, due to an additional run time; however, at higher energies the Multicomb is better than the normal code.

In the point D (see fig. 5) the method only deteriorates the FoM of the low-energy part without providing any gain in the high-energy tailing. We guess that this behaviour is related to the fact that, in this point, the potential is almost constant (as shown in the right fig. 1), and the electrons do not gain appreciable energy from the electric field. Consequently the EED tail is almost constant, and the Multicomb cannot improve the tallying.

The Multicomb introduces an error in the density, measured by eq. (11), as shown in fig. 6. The corresponding maximum percentage error is about 3%. Whereas the error in the velocity and energy is negligible.

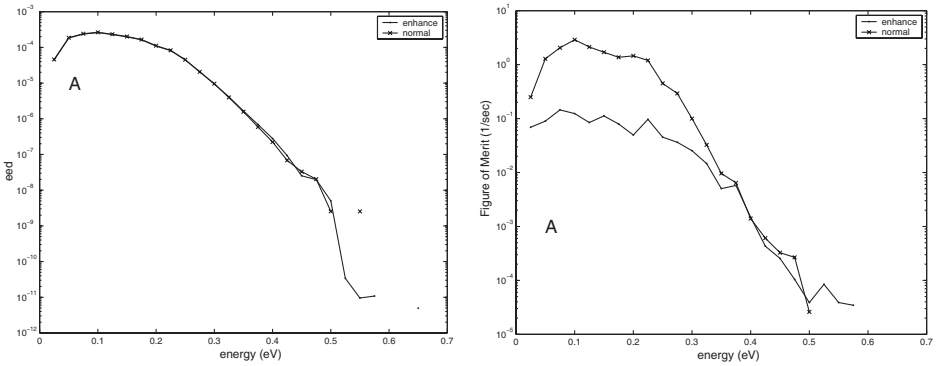

Fig. 2. – The electron energy distribution function (on the left) and the figure of merit (on the right) *vs.* the energy in the point A.

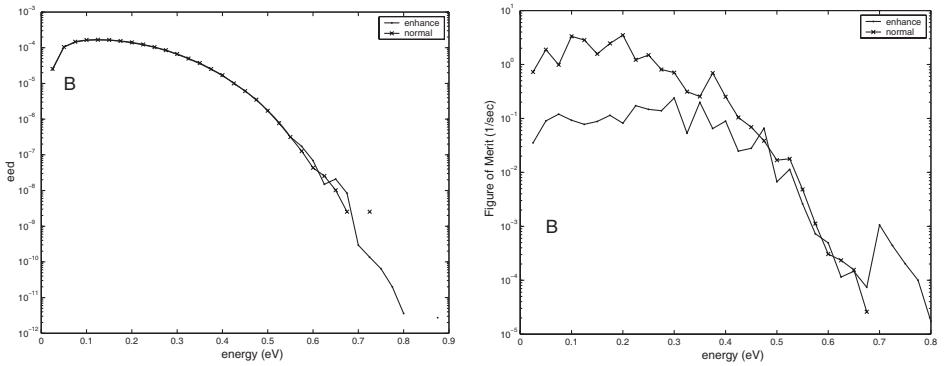

Fig. 3. – The electron energy distribution function (on the left) and the figure of merit (on the right) *vs.* the energy in the point B.

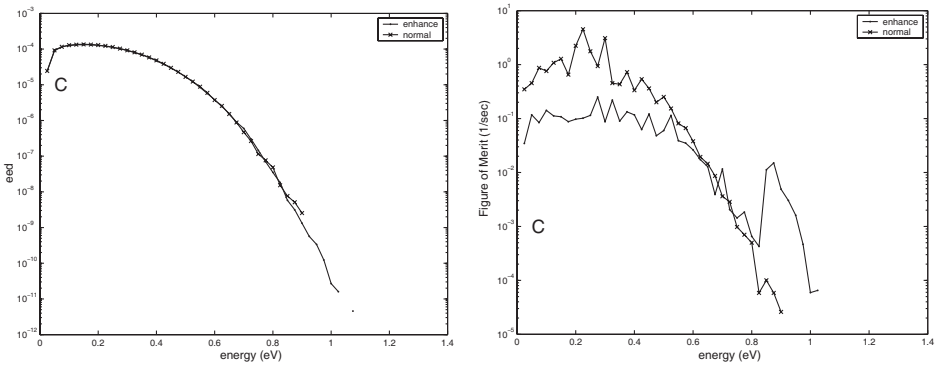

Fig. 4. – The electron energy distribution function (on the left) and the figure of merit (on the right) *vs.* the energy in the point C.

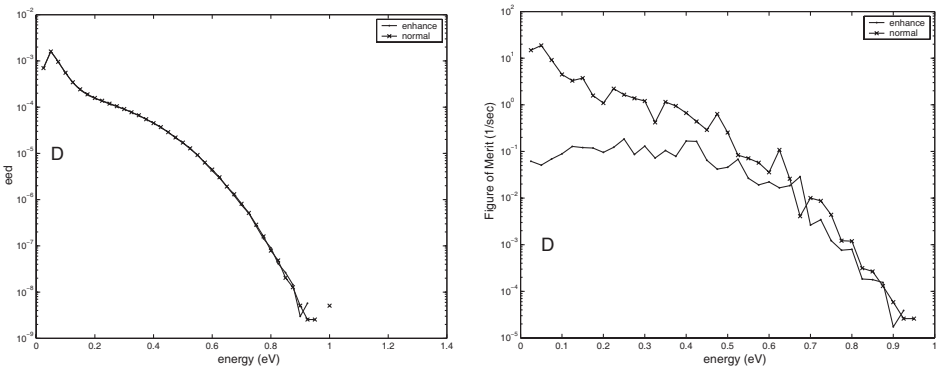

Fig. 5. – The electron energy distribution function (on the left) and the figure of merit (on the right) *vs.* the energy in the point D.

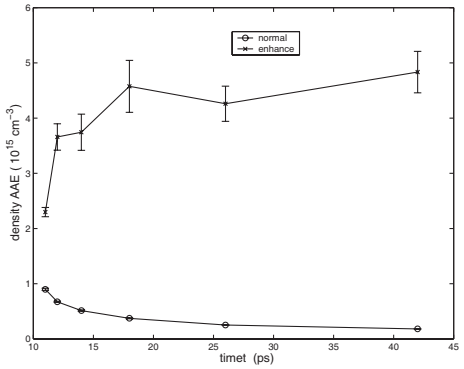

Fig. 6. – The Average Absolute Error given by eq. (11) for the density, as a function of the simulation time, for the enhanced (Multicomb) and normal codes.

The CPU consumed during the run with the Multicomb method was 11227 seconds, whereas that with the normal code was 2832 seconds. The particle number simulated was 61000.

5. – Conclusions

Monte Carlo semiconductor simulations need some form of variance reduction to minimize the noise in the data. We have implemented the *Multicomb* algorithm, in the case of $n^+ - n - n^+$ silicon diode, with the aim to quantify the efficiency of the method. Our simulations show that the method is able to populate the tail of the distribution function with an additional computational cost and by introducing a reasonable error in the density.

* * *

This work has been supported by "Progetto Giovani Ricercatori GNFM 2009" and "Progetti di ricerca di Ateneo", Università degli Studi di Catania.

REFERENCES

[1] Gray M. G., Booth T. E., Kwan T. J. T. and Snell C. M., *IEEE Trans. Electron. Dev.*, **45** (1998) 918.
[2] Wordelman C. J., Booth T. E. and Snell C. M., *IEEE Trans. Comput. Aided Design*, **17** (1998) 1230.
[3] Jacoboni C. and Reggiani L., *Rev. Mod. Phys.*, **55** (1983) 645.
[4] Muscato O. and Wagner W., *COMPEL*, **24** (2005) 1351.
[5] Lux I. and Koblinger L., *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations* (CRC, Boca Raton, FL) 1991.