Communications to SIMAI Congress,
ISSN 1827-9015, Vol. 3 (2009) 223 (12pp)
DOI: 10.1685/CSC09223

Free-flight time generation in Direct Simulation Monte Carlo for carrier transport in semiconductors

Vincenza Di Stefano

Dipartimento di Matematica e Informatica, Facoltà di Scienze MM.FF.NN.
Università degli Studi di Catania, Italy
vdistefano@dmi.unict.it

Abstract

This paper deals with the Direct Simulation Monte Carlo for carrier transport in submicron semiconductor devices. Different methods for the generation of the scattering times, called Constant-Time technique (CTT) and Self-Scattering technique (SST), are presented and compared, in order to analyze their efficiency and precision. One dimensional steady-state simulations of a $n^+ - n - n^+$ silicon diode have been carried out. For this particular device, SST seems to be more efficient than CTT.

Keywords: `Submicron Silicon semiconductor devices, Direct Monte Carlo Simulation, free flight.`

1. Introduction.

The continued miniaturization of integrated circuits and the current trend toward nanoscale electronics have led to tremendous integration levels, with hundred million transistors assembled on a chip area no larger than a few square centimeters. As a result, large electric fields and field-gradients generate hot or energetic electrons, and a very large quantity of heat being generated per unit volume.

The drift-diffusion equations, widely used in TCAD tools, are not able to describe accurately these regimes, and for this reason other transport models are needed. The natural framework for describing these regimes is the Boltzmann Transport Equation (hereafter BTE), coupled with the Poisson equation [6]. To solve the BTE is a hard task because it is an integro-differential equation with six dimensions in the phase space and one in time. A stochastic solution of the BTE can be obtained by the Direct Simulation Monte Carlo (DSMC) method, which replaces the distribution function with a representative set of particles. Direct Simulation Monte Carlo (DSMC) provides an accurate description of carrier transport phenomena because

Received 28/01/2009, in final form 18/06/2009
Published 31/07/2009

the various scattering mechanisms and band structure models are taken into account. The alternative is to use hydrodynamic models obtained from the BTE, which involve the solution of a coupled system of partial differential equations [1]. A larger difference between the two approaches is the way in which the semiconductor physics is introduced into the formulation. In the deterministic approaches, the physics is lumped into the parameterized mobilities, diffusion constants and the analytic band approximation is deal with only. In the DSMC technique, the physics enter in a more detailed way by: (i) using the full semiconductor band-structure obtained from the empirical pseudopotentials in order to study the hot electron transport beyond the analytic band approximation; (ii) employing the carrier-lattice, carrier-impurity and short-range Coulomb intercarrier collisions; (iii) treating the Coulomb long-range interactions and degenerate statistics in heavily doped regions. For this reason, access to the physics is more straightforward in the Monte Carlo (MC) approach. One of the basic recipe of the DSMC is the algorithm of generation of the free flight duration, i.e. the time between two collisions.
The plan of the paper is the following: in section II the physics and numerics used in the DSMC method are introduced, and in section III the free flight generation mechanism is recalled. In section IV and V Self Scattering technique and Constant Time technique respectively are explained. Finally, in section VI these methods are compared by using a $n^+ - n - n^+$ silicon diode, the results are shown and conclusions are drawn.

2. Physics and Numerics.

The BTE is an integro-differential equation which describes the time evolution and the variation in the phase space of the unknown distribution function $f(t, \mathsf{x}, \mathsf{k})$,

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

where q is the absolute value of the electron charge, $\hbar$ the Planck constant divided by 2π, v the electron group velocity, E the electric field, and $Q(f)$ the collisional operator. The linear scattering collision operator has the form

$$Q(f)(t, \mathsf{x}, \mathsf{k}) = \int_{\Omega} w(\mathsf{k}', \mathsf{k}) f(t, \mathsf{x}, \mathsf{k}') d\mathsf{k}' - w(\mathsf{k}) f(t, \mathsf{x}, \mathsf{k}) \tag{2}$$

where

$$w(\mathsf{k}) = \int_{\Omega} w(\mathsf{k}, \mathsf{k}') d\mathsf{k}' = \frac{1}{\tau(\mathsf{k})} \tag{3}$$

DOI: 10.1685/CSC09223

is the total scattering rate.
Ω is called first Brillouin zone, which is a characteristic of each material. In silicon this zone is formed by six equivalent ellipsoidal valleys located along the axis of the frame of reference at about 0.85 G from the center zone [3], where $|\mathsf{G}|= 2\pi/a$ and a is the lattice constant.
The main scattering mechanisms in silicon, at room temperature, are due to electron-phonon interactions (acoustic and optical phonons).
Their transition probability, per unit time, from a state k to a state k' can be modeled as:

$$w(\mathsf{k},\mathsf{k}^{'}) = K_0(\mathsf{k},\mathsf{k}')\delta(\varepsilon(\mathsf{k}')-\varepsilon(\mathsf{k})) + \sum_{i=1}^{6} K_i(\mathsf{k},\mathsf{k}') \times \\ \left[\delta(\varepsilon(\mathsf{k}')-\varepsilon(\mathsf{k})+\hbar\omega_i)(\mathsf{n}_{q_i}+1) + \delta(\varepsilon(\mathsf{k}')-\varepsilon(\mathsf{k})-\hbar\omega_i)\mathsf{n}_{q_i}\right]$$

where $\hbar\omega_i$ is a phonon energy and n_{q_i} the phonon equilibrium distribution function. According to the Bose-Einstein statistics, n_{q_i} is given by

$$\mathsf{n}_{q_i} = \frac{1}{\exp(\hbar\omega_i/k_B T_L)-1} \quad ,$$

where T_L is the lattice temperature.
The function K_0 represents intravalley elastic scattering transition probability and it reads

$$K_0(\mathsf{k},\mathsf{k}') = \frac{k_B T_L \Xi_d^2}{4\,\pi^2\hbar\,\rho\,v_\rho^2} \quad ,$$

where Ξ_d is the acoustic-phonon deformation potential, ρ the silicon mass density, v_s the sound velocity of the longitudinal acoustic mode.
K_i represents the inelastic scattering probability and is given by

$$K_i(\mathsf{k},\mathsf{k}') = \frac{Z_f (D_t K_i)^2}{8\,\pi^2\rho\ \ \omega_i} \quad ,$$

where $D_t K_i$ is the deformation potential for the i-th optical phonon, and Z_f is the number of final equivalent valleys for the considered inter-valley scattering. In the **quasi parabolic** approximation, considered in this paper, the kinetic energy $\varepsilon(\mathsf{k})$ of an electron satisfies the relation:

$$\varepsilon(\mathsf{k})[1+\alpha\varepsilon(\mathsf{k})] = \frac{\hbar^2\mathsf{k}^2}{2m^\star}, \quad \mathsf{k}\in\Omega,$$

where $m^\star$ is the effective electron mass (which is $0.32m_e$ in silicon) and α is the nonparabolicity factor. The electron group velocity $\mathsf{v}\equiv(v^1,v^2,v^3)$ is

given by

$$\mathsf{v}(\mathsf{k}) = \frac{1}{\hbar}\nabla_{\mathsf{k}}\varepsilon = \frac{\hbar\mathsf{k}}{m^{\star}[1+2\alpha\varepsilon(\mathsf{k})]} \quad .$$

During the free flight, particles move according to Newton's equations of motion

$$\frac{d\mathsf{x}}{dt} = \frac{1}{\hbar}\nabla_{\mathsf{k}}\varepsilon(\mathsf{k})$$
$$\hbar\frac{d\mathsf{k}}{dt} = -q\mathsf{E}(t,\mathsf{x})$$

where $\varepsilon(\mathsf{k})$ is the kinetic energy of the considered crystal conduction band structure measured from the band minimum. The electric field $\mathsf{E}(t,\mathsf{x})$ satisfies the Poisson equation

$$\Delta(\epsilon\phi) = q\Big[n(t,\mathsf{x}) - N_D(\mathsf{x}) + N_A(\mathsf{x})\Big],$$

$$\mathsf{E} = -\nabla_{\mathsf{x}}\phi, \qquad n(t,\mathsf{x}) = \int_{\Omega} f(t,\mathsf{x},\mathsf{k})d\mathsf{k},$$

where $\phi(t,\mathsf{x})$ is the electric potential, N_D and N_A are respectively the donor and acceptor densities (which are positive functions).
The system formed by the Newton's equations and the Poisson equation is solved with a numerical scheme, e.g. Runge-Kutta scheme [4], up to the next scattering time, or to a fixed time step Δt if no scattering occurs.

Remark 2.1. Regarding the choice of the time step Δt, during which the equations of motion are integrated, in order to avoid plasma oscillations [5], it is necessary to choose

$$\Delta t \ll \frac{1}{\omega_p} \simeq 76fs$$

where ω_p is the plasma frequency. The number in the previous formula, has been evaluated for the high doping region in the device.

3. Free flight generation.

Usually the scattering process is considered **markovian**. Consequently, the probability that a particle does not suffer a collision in the time interval $[0,t]$ is :

$$P(t) = \exp\left[-\int_0^t w[\mathsf{k}(t')]dt'\right]$$

DOI: 10.1685/CSC09223

where $w(\mathsf{k})$ is the total transition probability given in (3). In order to generate a stochastic free flight, one takes an uniform distribution of random numbers r and tries to solve the problem:

$$P(t) = r$$

or by taking the logarithm,

$$-\log r = \int_0^t w[\mathsf{k}(t')]dt'. \tag{4}$$

The time t is the free flight duration.
Since the solution of the eq. (4) is of paramount importance to obtain an efficient implementation of the algorithm, some methods have been introduced [2], [9]. These methods are:

- Self Scattering Technique (SST)
- Constant Time Technique (CTT)

In both methods the DSMC introduces an error, called **splitting error**, due to the fact that the equations of motion are solved during the time step Δt with a frozen electric field, evaluated at the previous step.

4. Self Scattering technique.

In SST a common approach is the introduction of fictitious scattering events.
If the self scattering is selected, nothing happens to the particle which maintains, after the scattering, the same energy and the same momentum it had before.
Self scattering does not alter statistical distribution of the real scattering events, but the total scattering rate changes.
The algorithms implemented, in the class of the SST, are the following:

- Constant-Gamma scheme [9]
- Piecewise-Constant-Gamma scheme [3]
- Individual-Gamma scheme [8]

4.1. *Constant Gamma scheme (CG).*

Let Γ be a number greater than the largest scattering rate possible in the simulation and let introduce a fictitious scattering probability $w_{ss}(\mathsf{k})$ so that

$$w(\mathsf{k}) + w_{ss}(\mathsf{k}) = \Gamma = \text{const} \quad .$$

Then the equation (4) gives

$$t = -\frac{\log r}{\Gamma} \quad .$$

The main advantage of this technique is that the programming is very simple; one disadvantage is that a lot of computer time, about the 90%, is spent performing computations related to this scattering.
In order to reduce the number of self scatterings, some improvements to this algorithm have been proposed [10].

4.2. *Piecewise-Constant-Gamma scheme (PWCG).*

Sometimes the total scattering probability has a large variation around some threshold value due to strong scattering mechanism with a given activation energy. In this case a single value of Γ may result in a large number of self-scattering events at low electron energy (see Fig. (1)). It is then possible to introduce a step-shaped scattering rate, as follows.
Let t_0 be the current simulation time, $\varepsilon_0, \mathsf{k}_0$ the corresponding particle energy and wave vector and ε_1 a suitable threshold energy. Let divide the particles into two subsets: S_1 whose particle energy is $\varepsilon \leq \varepsilon_1$ (with inverse scattering rate Γ_1) and S_2 whose particle energy is $\varepsilon > \varepsilon_1$ (with inverse scattering rate Γ_2). The quantity Γ_2 is chosen to be the maximum value during the simulations, which corresponds to a maximum energy $\varepsilon_m = 1$ eV. Moreover, let r be a random uniform number and t_r the free flight duration.
The implemented PWCG algorithm is the following:

1. generate r
2. if $\varepsilon_0 > \varepsilon_1$ goto 3 else goto 4
3. $t_r = t_0 - \frac{1}{\Gamma_2} \log r \rightarrow$ return
4. $t_r = t_0 - \frac{1}{\Gamma_1} \log r$
5. move the particle in free flight in the time interval $[t_0\ ,\ t_r\ ]$, and evaluate its final energy ε_r
6. if $\varepsilon_r < \varepsilon_1$ return else goto 7
7. since the particle energy is above ε_1, t_r cannot be retained. The new time is $t_r = -\frac{1}{\Gamma_2} \log r + \tilde{t}\left(1 - \frac{\Gamma_1}{\Gamma_2}\right) + t_0$, where $\tilde{t}$ is the time necessary for the electron to reach the energy ε_1
8. goto 5

Through this algorithm, it is possible to obtain a gain in the number of scattering events. The gain is given by the region of Fig.(1) characterized

DOI: 10.1685/CSC09223

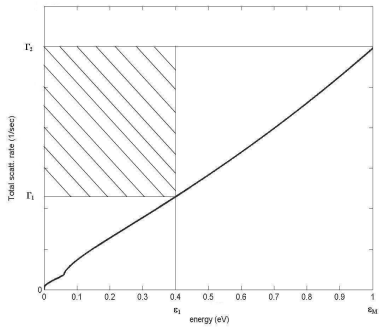

Fig. 1. Sketch of a two-level step-shaped scattering rate, including self-scattering.

by slanting lines. For the calculation of $\tilde{t}$ a linear interpolation is performed, which introduces a further error in the algorithm.

4.3. *Individual-Gamma scheme (IG).*

In order to further reduce the number of self scatterings, it was implemented an algorithm in which a scattering rate Γ is generated for each particle.
Let t_0 be the current simulation time and ε_0 the corresponding particle energy.
The implemented algorithm is:

1. fix the i-th particle
2. let move the particle in $[t_0, t_0 + \Delta t]$, and evaluate its final energy $\varepsilon_1 = \varepsilon(t_0 + \Delta t)$
3. calculate $\varepsilon_M = \max(\varepsilon_0,\varepsilon_1)$ and evaluate $\Gamma = \Gamma(\varepsilon_M)$
4. generate a random uniform number r
5. $t_r = t_0 - \frac{1}{\Gamma} \log r$
6. if $t_r > t_0 + \Delta t$ goto 1 (the i-th particle is arrived at the final time without any scattering, forget t_r , continue with the next particle i $\leftarrow$ i + 1) else goto 7
7. move again the particle in $[t_0, t_r]$
8. check if a scattering (self or real) occurs
9. until $(t_0 + \Delta t) - t_r > 0$, $t_0 \leftarrow t_r$ goto 2 .

In IG algorithm no further error is introduced and the number of self-scatterings reduces, as will be shown in Section 6.

5. Constant Time Technique.

In CCT the total simulation time is subdivided into time intervals Δt. The probability that a particle will survive without scattering during a ballistic flight of duration Δt is

$$\exp\left\{-\int_t^{t+\Delta t}\frac{dt'}{\tau[\mathsf{k}(t')]}\right\}\simeq\exp\left\{-\frac{\Delta t}{\tau(\mathsf{k})}\right\}$$

having assumed that the time step Δt is small enough so that $\mathsf{k}(t')$ can be taken as constant during the free flight.
If Δt $/\tau(\mathsf{k})\ll 1$, the probability that the particle will scatter at the end of the free-flight of duration Δt can be approximated as:

$$1-\exp\left\{-\frac{\Delta t}{\tau(\mathsf{k})}\right\}\simeq\frac{\Delta t}{\tau(\mathsf{k})}\quad .$$

Therefore, for each particle the total scattering rate is evaluated at the final wavevector $\mathsf{k}(t+\Delta t)$ and the following comparison is made:

$$\frac{\Delta\mathrm{t}}{\tau(\mathsf{k})}\geq< r_1 \tag{5}$$

where r_1 is a random number $\in$ [0,1].
If in eq. (5) the operator $\geq$ holds the particle suffers a scattering, otherwise no scattering occurs.
It should be underlined that, in addition to the splitting error, in CCT the operations of approximations and series expansion of the integral introduce a **systematic error**, which influences the number of scatterings.
The positive aspects of the use of the CTT are the following: (i) is very simple to implement, because the particles are synchronized in time; (ii) is simple to obtain a parallel version of the algorithm; (iii) is very efficient, in terms of CPU time, because the scattering check is done every Δt.

6. Results.

For this paper, a one-dimensional n^+-n-n^+ silicon diode was used. This diode consists of two highly doped regions n^+, called cathode and anode, connected by a less doped region n, called channel.
In the simulations, the n^+ regions are 150 nm-long doped to a density of $2\times10^{17}\mathrm{cm}^{-3}$, while the channel is 250 nm-long doped with a density n of $10^{15}\mathrm{cm}^{-3}$. The device is considered at room temperature T_0 =300K, and the applied bias is 1 V. The particle number was fixed and is $N=62000$.
In the first step, the average number of scattering events was calculated for

DOI: 10.1685/CSC09223

CG, PWCG and IG schemes. The perfomance of the PWCG scheme was evaluated for different values of the threshold energy ε_1.
As you can see in table (1), in PWCG scheme the average number of self-scattering events is smaller than the average number of self-scatterings in CG scheme.
Moreover, in the PWCG this number reduces when the value of the energy ε_1 decreases.
However, the best result is obtained by the IG scheme, in which the self-scattering percentage is only the 2%, which predicts a gain in the CPU time.

Table 1. Average number of scattering events per particle per time (in ps)

Algorithm	RealScat	SelfScat	SS percentage
Constant-Gamma	6.3	113	95 %
PW-C-Gamma $\Gamma_1(0.4)$	6.3	45.1	88 %
PW-C-Gamma $\Gamma_1(0.2)$	6.3	26.5	81 %
PW-C-Gamma $\Gamma_1(0.1)$	6.3	16.4	72 %
Individual-Gamma	6.3	0.12	2 %

In a second step, a comparison between the Constant Gamma and the Piecewise-Constant-Gamma schemes was performed. As shown in Fig.(2), the PWCG algorithm introduces an error in the average velocity and in the average energy, wich decreases when the value of the threshold energy ε_1 is increased.

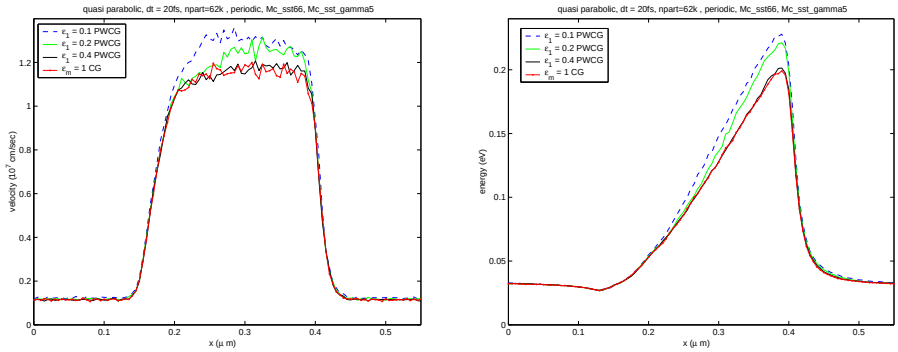

Fig. 2. The average velocity (on the left) and the average energy (on the right) versus x for some value of ε_1.

By comparing the results obtained with the CG scheme and the IG scheme, it is evident that IG algorithm does not introduce errors, apart from the

splitting error, neither in the average velocity nor in the average energy, as shown in Fig. (3). This happens also for different values of the time step Δt.
In Fig. (4) the behaviour of the CTT is shown. It can be seen that, by varying the value of the time step, a big error is introduced in the average velocity and also in the average energy.

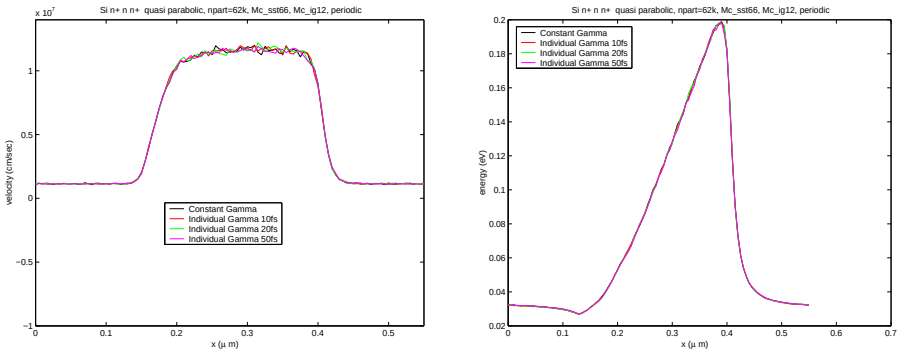

Fig. 3. The average velocity (on the left) and the average energy (on the right) versus x, obtained with the IG algorithm.

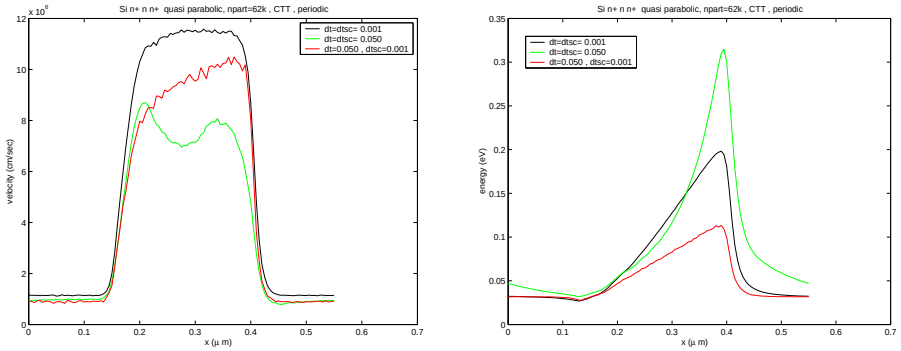

Fig. 4. The average velocity (on the left) and the average energy (on the right) versus x, obtained with the CCT algorithm.

The last considerations are about the CPU times and the precision of the implemented algorithms.
For SST the splitting error vanishes for $\Delta t = 60fs$, which corresponds to a CPU time of $\simeq 300s$ (see Fig. (5)).
The same amount of CPU is used in CTT case for $\Delta t = 4fs$ and in IG case for $\Delta t = 10fs$.

DOI: 10.1685/CSC09223

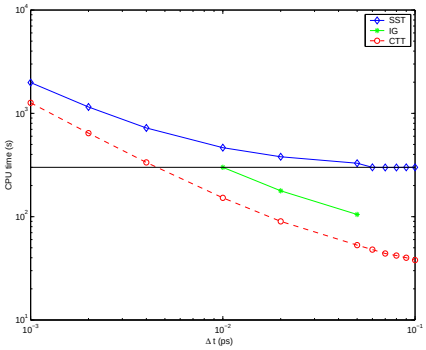

Fig. 5. CPU times.

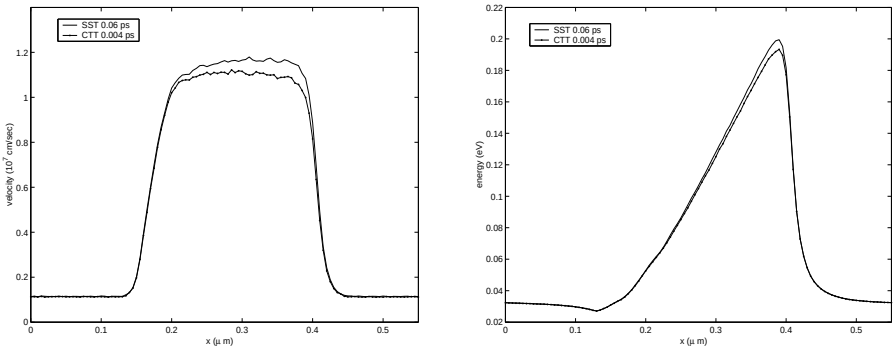

Fig. 6. Error in the velocity (on the left) and in the energy (on the right).

But, in the case of the CTT, for $\Delta t = 4fs$, the error is still significant and is: (i) 3.44% for the energy; (ii) 8.85% for the velocity, as shown in Fig. (6). In the IG case, in correspondence of $300s$ of CPU time, there is the same precision of the STT.
But, by comparing the performance of the two algorithms in correspondence of $\Delta t = 60fs$, is evident that in IG there is a gain factor which is about 3.

Acknowledgements.

This work makes use of results produced by the PI2S2 Project managed by the Consorzio COMETA, a project co-funded by the Italian Ministry of University and Research (MIUR) within the Programma Operativo Nazionale "Ricerca Scientifica, Sviluppo Tecnologico, Alta Formazione". The author was supported by "Progetti di Ricerca di Ateneo", Università degli Studi di Catania.

REFERENCES

1. A.M. Anile and O. Muscato, Improved hydrodynamical model for carrier transport in semiconductors, *Phys. Rev. B*, **51**(23), pp. 16728–16740, 1995.
2. M. V. Fischetti and S. E. Laux, Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects, *Phys. Rev. B*, **38**(14), pp. 9721–9745, 1995.
3. C. Jacoboni and L. Reggiani, The Monte carlo method for the solution of charge transport in semiconductors with applications to covalent materials, *Rev. Modern. Phys.*, **55**(3), pp. 645–705, 1983.
4. S. E. Laux and M. Fischetti, Numerical aspects and implementation of the DAMOCLES Monte Carlo device simulation program, K. Hess editor, *Monte Carlo Device Simulation: Full Band and Beyond*, pp. 1–26, Kluver, Boston, 1991.
5. M. Lundstrom, Fundamentals of carrier transport, Cambridge: Cambridge University Press, 2000.
6. P. A. Markovic, C. A. Ringhofer and C. Schmeiser, Semiconductor equations, Vienna: Springer-Verlag, 1990.
7. O. Muscato, V. Di Stefano and W. Wagner, Analysis of the Error in Direct Simulation Monte Carlo for Submicrometric Semiconductors using the PI2S2 Grid Infrastructure, in Proceedings of the Symposium *GRID Open Days at the University of Palermo*, edited by R. Barbera, Consorzio COMETA, pp. 277–284, 2008.
8. O. Muscato, V. Di Stefano and W. Wagner, Numerical study of the systematic error in Monte Carlo schemes for semiconductors, to appear on M2AN, 2009.
9. H. D. Rees, Calculation of steady state distribution functions by exploiting stability, *Phys. Lett. A*, **26**(9), pp. 416–417, 1968.
10. R. M. Yorston, Free-flight time generation in the Monte Carlo simulation of carrier transport in semiconductors, *J. Comput. Phys.*, **64**(1), pp. 177–194, 1986.