(1-2)

MONTE CARLO SIMULATION OF SEMICONDUCTOR DEVICES

B. Riccò, E. Sangiorgi†, and F. Venturi+

Department of Electronics, University of Bologna, Italy
† Department of Physics, University of Udine, Italy
+ Department of Electronics, University of Parma, Italy

ABSTRACT

This paper presents a brief overview of Monte Carlo simulation of electron devices. A simulator is described that can be used to study sub-micron devices where non-equilibrium effects play an essential role. Selected results obtained with advanced silicon MOS transistors are shown to illustrate important applications where the Monte Carlo approach has no viable alternative.

I. Introduction

This paper is essentially based on the work started in 1985 at the University of Bologna (Italy) in co-operation with the Italian Universities of Udine, Parma and Modena and with AT&T Bell Laboratories (Murray Hill, NJ, USA). The result of such a work is an advanced Monte Carlo (MC) simulator of electron devices, uniquely suitable to study high field transport and carrier heating in VLSI and ULSI devices.

The essential motivations for this project are. a) the growing importance of simulation in advanced technology development; b) the increasing inadequacy of the conventional models of charge transport in semiconductors.

The drift-diffusion (DD) model in fact, derived under simplifying assumptions from Boltzmann transport equation [1], looks only at average (i e. *macroscopic*) properties of the carrier gas and provides an intrinsically *local* description of charge transport with carriers assumed to be in *equilibrium* with the applied fields. Recent works on sub-micron MOSFETs [2,3], however, have clearly shown that non-equilibrium effects play an important role in determining device characteristics.

An almost ideal solution to device simulation is provided by the Monte Carlo (MC) technique [4] to solve directly Boltzmann equation following in space and time the history of each individual particle. This method, however, requires large computation resources, thus it was considered impractical for real device applications. However, recent efforts have changed the situation and MC device simulation is a candidate for future use in technology development.

II. The Monte Carlo method

The MC model of charge transport in semiconductors [4] considers a large number of carriers subject to external forces (electric field) and given scattering mechanisms. A carrier trajectory consists of a sequence of ballistic flights governed by particle dynamics and terminated by collisions (scattering events) described by microscopic models relating the carrier parameters (energy, momentum,) before and after the collision. The scattered particles start a new free flight with a well defined new state. The number of simulated carriers is generally chosen so as to obtain a satisfactory trade-off between CPU time and (granular) noise on calculated statistical properties.

With the MC approach physical modeling is made at the microscopic level with the description of the particle energy-momentum relationship and of the various types of collisions. Since the simplifying assumptions are more justified than those used in macroscopic approaches, MC simulation is more accurate and founded on stronger physical basis than other techniques.

Work supported by CNR MADESS project.

The duration of each free flight and the type of terminating collisions are selected stochastically, according to the probabilities of the microscopic processes. Because each particle is followed individually along its trajectory, the model is intrinsically *non local* and provides detailed microscopic informations, such as carrier position, momentum and energy. Of course, by means of suitable averaging the same quantities as in macroscopic models can be easily obtained.

The MC method is structured in such a way that new scattering models can be easily added to the simulator. Most importantly, it avoids the relaxation time approximation and assumptions on the shape of the carrier energy and momentum distribution. Consequently, energy threshold phenomena (impact ionization, gate current, and hot-carrier degradation, ...) can be most successfully treated [5,6] by means of calculated carrier distributions.

As already mentioned, the advantages briefly discussed above are obtained at the cost of expensive computations, in particular in the case of realistic multidimensional devices.

III. MC simulation of semiconductor devices

In order to apply the MC technique to real devices operating under extreme conditions, several major problems had to be solved. The most relevant among them were the following.

a) *Efficient management of rare events*. When interested in *rare events*, a large amount of resources must be dedicated to the simulation of common ones to achieve significant statistics. A classical example of this problem is given by high energy electrons which normally represent a negligible fraction of the whole carrier population in the high field active region. This difficulty can be tackled by means of multiplication techniques expanding the statistical weight of occurring rare events. In practice, each of these is used to generate proper *initial conditions* for several others.

This technique, presented in [7] for the first time, has been implemented in our simulator in an improved version that can recursively treat events of nested rarity and different nature (position, energy, etc.)[8].

b) *Simulation domain boundaries*. In order to save CPU time the MC simulation domain should be restricted to the active device region where high electric fields are normally present. In this way, however, the parameters of the electrons entering the simulation domain are not known *a priori*, while the use of simplifying assumptions can affect significantly the obtained results.

A solution to this problem is to place boundaries only in low-field regions where equilibrium conditions can be invoked.

For a MOSFET this implies to simulate also significant portions of the highly doped source and drain regions although this inevitably increases CPU time, largely spent on electrons within the populated low-field region rather than on the few interesting ones in the MOSFET channel.

In our case, this problem has been solved by considering the presence of each carrier in the device active region as a rare event to be treated as briefly described in the previous point [8].

c) *Self-consistency of potential and carrier concentration*. MC simulation of charge transport directly provides the carrier space distribution to use in Poisson equation. The iterative procedure needed to obtain consistent carrier and potential distributions requires extremely

large CPU time. Furthermore, serious problems arise from the noise associated with the MC calculation that can seriously threaten the convergence of the whole process.

We have developed an original, non linear scheme to couple MC transport and Poisson equation that provides very robust convergence [8].

d) *Efficient use of CPU resources* As MC simulations are very demanding, special attention must be paid to the problem of computation efficiency. Our effort was directed toward: 1) efficient mapping onto vector and/or parallel machine architectures; 2) development of algorithms to increase computation efficiency (rare events, self-consistency, calculation of carrier free flight duration [6], ...). A global CPU time reduction of several orders of magnitude was achieved with complete simulations of MOSFET transistors in $\approx 2h$ of a single processor CRAY-YMP machine (see Table 1).

e) *Accurate microscopic modeling.* In order to consider carrier energies up to few eV the traditional, single energy-band models are inadequate. The alternative is the use of the full silicon band structure which has a dramatic impact on CPU time [9] We have developed an original, simplified model of the energy-momentum relationship that provides a good trade-off between accuracy and efficiency [10]. This model has been recently improved including a consistent and accurate impact ionization model [11].

IV. Relevant applications of the MC simulator

The first example presented here concerns the calculation of *macroscopic* device characteristics (terminal currents, capacitances, intrinsic velocity, ...), while the second deals with the crucial problem of hot electrons in scaled sub-micron devices.

For the former case, deep submicron MOS transistors with gate length (L_{gate}) varying from 0 15μm to 0.75μm (effective lengths of 0.1 – 0.7μm) have been analyzed with the self-consistent MC simulator as well as with an advanced DD program in order to compare the effective capabilities of the two approaches.

Table 2 summarizes a relevant set of calculated device parameters. Comparing device performance obtained with MC and DD simulations, it can be seen that the transconductance (g_m) differs significantly only for gate lengths below 0 25μm, due to non negligible velocity overshoot throughout the channel (in good agreement with experiments [3]). The electron velocity averaged over all carriers (v) obtained from MC and DD simulations are shown (together with the MC average energy) as a function of position along the channel in Fig. 1 for the cases L_{gate} = 0.75μm, 0 25μm and 0 15μm, respectively. From these distributions, the velocity $\overline{v}$ averaged also over the whole channel can be easily computed (Table 2). As can be seen. at 0.75μm (Fig. 1(a)). v_{MC} is substantially higher than v_{DD} only in a very small high-field region near the drain. In the 0 25μm device, instead, $v_{MC} > v_{DD}$ throughout a much higher portion of the channel (Fig. 1(b)); because, however, MC and DD velocities are nearly the same at the source end of the channel, $g_{mMC} \simeq g_{mDD}$ Finally at 0 15μm (Fig. 1(c)), $v_{MC} > v_{DD}$ for the whole channel, consequently $g_{mMC} > g_{mDD}$

Another interesting aspect of the MC calculation is that the inversion charge is significantly smaller than for the DD case, so is also the MC gate capacitance (C_G) Since, as already discussed, $g_{mMC} > g_{mDD}$, MC simulation predicts significantly higher velocity performance than the DD model, as indicated by the calculated cutoff frequency ($f_T = g_m/2\pi C_G$) As expected, the MC and DD f_T's are nearly identical for relatively long devices, but differ significantly with decreasing gate length Finally, Fig. 2 shows the difference in output characteristics of the 0 15μm device calculated by the MC and DD approaches

In the second example, we investigated the effects of different scaling rules on electron heating and intrinsic device performance in deep submicron MOSFETs. To this purpose, the technological parameters of a reference 1μm process (L_{eff} = 0.75μm) have been scaled down to L_{eff} = 0 25μm according with the conventional rules [12], shorter devices, with L_{eff} down to 0.075μm, have also been simulated with the same technological parameters as the 0.25μm process Two different voltage scaling strategies have been investigated: the traditional approach [12] (F_{av} scaling) which makes the average lateral electric field ($F_{av} = 80kV/cm$) in the channel independent of device dimensions, and a different one (F_{max}) designed to keep the same value of the maximum lateral electric field at the $Si - SiO_2$ interface ($F_{max} = 390kV/cm$). In the latter case the voltages must be scaled less than the dimensions because of the different shape of the potential along the channel and the scaling rule must be obtained empirically because of the lack of analytical relationship between F_{max} and applied voltages

Figs. 3,4 illustrate the calculated electron heating and device performance obtained with the F_{max} scaling rules. Fig. 3 indicates that a drastic reduction in hot carriers is expected for ultra short MOSFETs, though without any abrupt threshold, in particular the presence of few electrons at energies above qV_{ds} (see curves D and E) and their low T_e agrees with experiments on low voltage hot electrons exhibiting enhanced dependence on V_{ds} [13]. Fig. 4 shows the transistor current as predicted by MC and DD models as a function of channel length. the F_{max} scaling leads to increased current capability for smaller devices In addition, at the shortest L_{eff}, velocity overshoot becomes important and the DD model significantly underestimates the achievable performance. The same analysis described above performed with the F_{av} (i e linear) voltage scaling clearly shows (Fig 5) that hot electron effects are drastically reduced (mainly because F_{max} decreases for the shortest device), but the current capability is badly affected (Fig 6) and so is the device dynamic performance.

The results briefly illustrated above clearly indicate that linear voltage scaling is too conservative for future deep submicron technologies, at least from the point of view of electron heating. In fact, supply reductions designed for constant maximum lateral electric field provide better device performance with no increase of hot electron effects.

ACKNOWLEDGMENTS The Authors are grateful to the *Monte Carlo Group* of the University of Modena, Italy (Profs C. Jacoboni, P. Lugli and R. Brunetti), the *Monte Carlo Group* of the University of Bologna, Italy (Drs. C. Fiegna, M. Lanzoni, and R. Pioggia), and the *Device Simulation Group* of *AT&T* Bell Labs. (Drs M. Pinto and R. K. Smith) for the intense and very fruitful cooperation.

REFERENCES

[1] K. Bløtekjær, *IEEE Trans. Electron Devices*, vol. ED-17, p 38, 1970.
[2] G.G Shahidi, D A. Antoniadis, and H.I. Smith, *IEEE Elect. Dev. Lett.*, EDL-9, p 497, 1988
[3] M C Jeng, J. Chung, A. T. Wu, T. Y. Chan, J Moon, G May, P. K Ko, and C Hu, in IEDM Tech. Dig , pp 710-713, 1987
[4] C. Jacoboni and L. Reggiani, *Rev. Mod Phys* , 55. p 645, 1983.
[5] E. Sangiorgi, M R Pinto, F. Venturi, and W Fichtner, IEEE Electron Device Letters, vol EDL-9, pp.13-16, 1988.
[6] E. Sangiorgi, B. Riccò, and F. Venturi, IEEE Trans. on CAD, vol. CAD-7, pp 259-271, 1988.
[7] A Phillips, P J. Price, Appl Phys. Letters, vol. 30, pp 528-530, 1977.
[8] F. Venturi, R.K Smith, E. Sangiorgi, M.R. Pinto, and B Riccò, *IEEE Trans on CAD*, CAD-8, p 360, 1989.
[9] M Fischetti and S Laux, *Phys Rev. B*, 38, p 9721, 1988
[10] R. Brunetti, C. Jacoboni, F Venturi, E Sangiorgi, and B. Riccò, *Solid-State Electronics*, vol. 32, p 1663, 1989.
[11] F. Venturi, E Sangiorgi, R. Brunetti, W. Quade, C Jacoboni, and B Riccò, NUPAD III, Hawaii, June 1990.
[12] R H Dennard, F H Gaensslen, L Kuhn, N Y. Yu, V.L. Rideout, E. Bassous, and A. Le Blanc, *IEEE J. Solid-State Circuits*, SC-9, p.256, 1974
[13] B Riccò, E Sangiorgi, and D. Cantarelli, *IEDM Tech. Dig* , p 92, 1984.

Device	MC-Poisson loops	CPU time (s)	Vector speed-up	Scattering events/sec (s^{-1})
$N^+ - N - N^+$	3	135	4.5	45,000
0.25μm MOSFET	10	7,680	4	16,000
0.15μm MOSFET	10	6,960	4	20,000
0.075μm MOSFET	10	6,200	4	22,000

Table 1 Timing results for a single processor CRAY-YMP

L_{mask} (μm)	g_m (mS/mm)	g_m/C_{ox} (cm/sec)	$\bar{v}$ (cm/sec)	C_G (pF/cm)	$f_T = g_m/2\pi C_G$ (GHz)
0.75 (DD)	230	4.00×10^6	4.00×10^6	32.9	11.1
0.75 (MC)	200	3.47×10^6	4.00×10^6	28.2	11.2
0.50 (DD)	300	5.20×10^6	5.25×10^6	22.2	21.5
0.50 (MC)	330	5.73×10^6	5.48×10^6	22.2	23.7
0.25 (DD)	370	6.42×10^6	7.32×10^6	11.7	50.3
0.25 (MC)	400	6.94×10^6	8.50×10^6	10.7	59.5
0.15 (DD)	420	7.29×10^6	8.50×10^6	7.1	94.1
0.15 (MC)	600	1.05×10^7	1.11×10^7	7.05	135.4

Table 2. Intrinsic device characteristics as given by DD and MC models. $V_{GS} = 1.5V$, $V_{DS} = 2.0V$, $T = 300K$

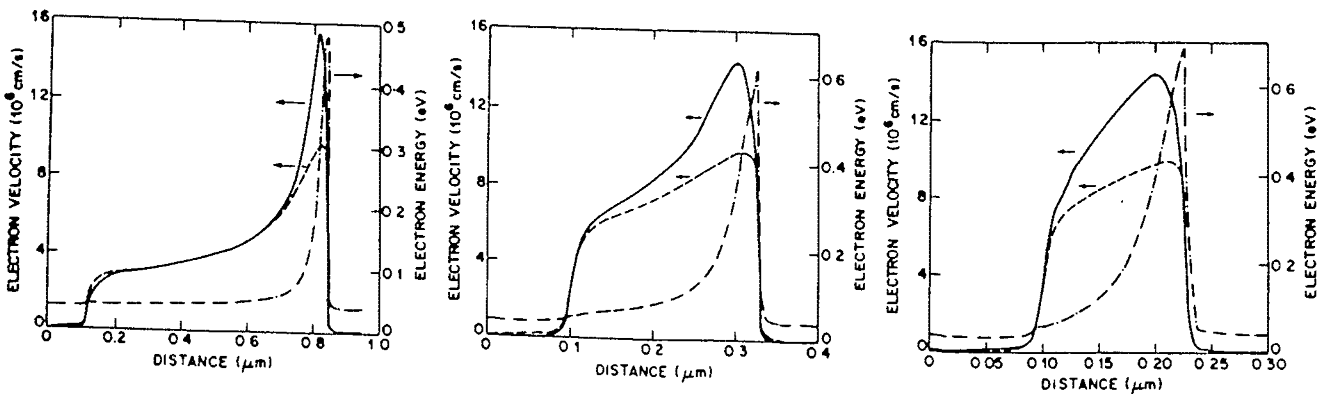

Fig. 1. Average electron velocity in the MC (solid) and DD (dashed) simulations and surface electron energy in the MC solution (dashed-dotted) for $V_{GS} = 1.5V$, $V_{DS} = 2.0V$ and $L_{gate} = 0.75\mu m$ (a), $L_{gate} = 0.25\mu m$ (b), and $L_{gate} = 0.15\mu m$ (c).

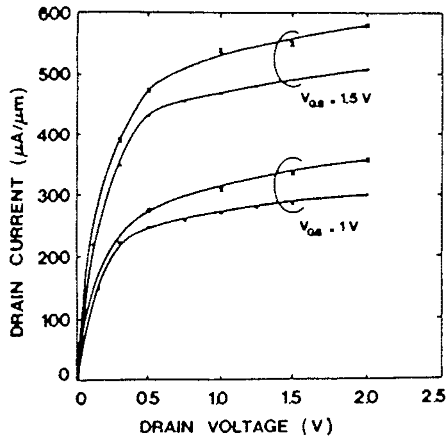

Fig 2 Output characteristics for a MOSFET with $L_{gate} = 0.15\mu m$ obtained with the DD (•) or the MC (x) model

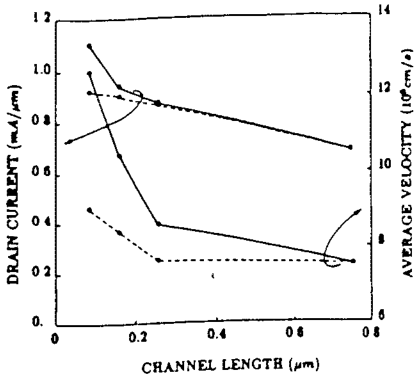

Fig. 4. Drain current and average electron velocity in scaled MOS transistors with applied voltages reduced so as to maintain the same value of the maximum electric field in the channel, calculated with the MC and DD model (solid and dotted line, respectively)

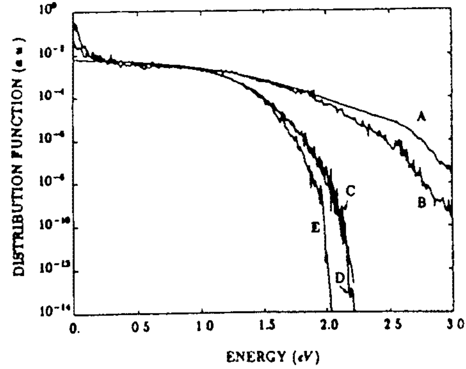

Fig. 3. Electron energy distribution (at the point of maximum carrier heating) calculated in scaled MOS transistors with applied voltages reduced so as to maintain the same value of the maximum electric field in the channel. The curves represent the cases of L_{eff} = 0.75 (B), 0 25 (C), 0 15 (D) and 0.075 μm(E), respectively. The case (A) of a constant field is also illustrated for comparison.

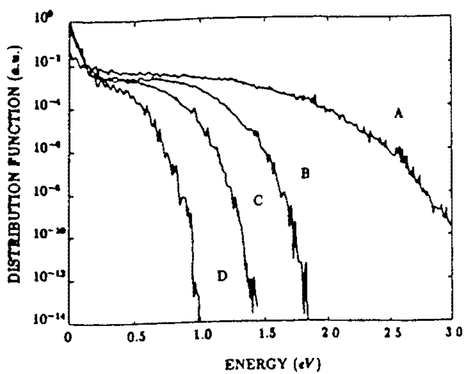

Fig. 5. Electron energy distribution (at the point of maximum carrier heating) calculated in scaled MOS transistors with applied voltages reduced so as to maintain the same value of the average electric field in the channel. The curves shown here represent the cases of L_{eff} = 0.75 (A), 0.25 (B), 0.15 (C) and 0.075 μm (D), respectively.

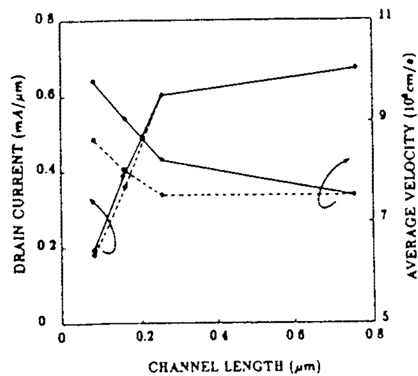

Fig. 6. Drain current and average electron velocity in scaled MOS transistors with applied voltages reduced so as to maintain the same value of the average electric field in the channel, calculated with the MC and DD model (solid and dotted line, respectively)