Comparison of Cellular Automata and MINIMOS simulations of submicron MOSFETs

A. Rein, G. Zandler and P. Lugli

Physik Department and Walter Schottky Institut, TU München, D-85747 Garching, FRG

Abstract

We present a detailed comparison of a novel cellular automaton (CA) technique and a standard drift diffusion calculation (MINIMOS) of high field transport in semiconductor devices. The CA method may be viewed as equivalent to the Monte Carlo technique but can easily handle ensembles with more than 10^5 particles, can efficiently deal with complex geometries and achieve accelerations on multiprocessor computers that scale linearly with the number of processors. With this new technique Si MOSFETs have been simulated for different gate lengths and gate voltages and the results compared to MINIMOS.

I. Introduction

Device simulation has become a crucial and strategic part of today's microelectronics [1–3]. Drift diffusion approaches [4] are commonly used for device modelling because of their intrinsic speed. They suffer though from the several approximations of the underlying physical model. The Monte Carlo (MC) method [5,6] belongs on the contrary to the most sophisticated, but at the same time the most costly, of all simulators. For this reason, the MC method still remains restricted to university and laboratory research, and has not yet become a common modeling tool. It would therefore be highly desirable to develop a simulator of comparable physical content as the MC, but much faster and also capable of exploiting the potential offered by vector and parallel processors more naturally.

Recently, a new method has been developed which appears to meet these requisites, the Cellular Automata (CA) approach [7,8]. So far, only preliminary tests of its applicability to device modeling have been carried out. In this paper, a critical assessment of the strengths and limitations of the CA for realistic device simulations is given by presenting the first quantitative and detailed comparison between the CA and a standard Drift-Diffusion approach based on the MINIMOS code [1] for a Si MOSFET.

II. The Cellular automaton for device simulations

Like the MC method, the CA is a physical approach to carrier transport in semiconductors based on the simulation of a random walk of classical particles subject to probability scattering events by phonons, impurities, other carriers, whose energy dependence is evaluated from Fermi's Golden rule [5,6].

In general, the electric field acting on the particles is a function of position; in order to fully account for non-homogeneous situations (and therefore to simulate semiconductor devices) the MC and the CA simulations have to be self-consistently coupled to Poisson's equation [5,6]. Recently, the full BE for carrier transport in semiconductors has been transformed to a CA [7]. This constitutes an important improvement, since CA are traditionally only used in the context of transport, for fluid dynamics [9,10] or drift-diffusion simulations [11]. In general, a cellular automaton consists of a lattice with a finite number of states attached to each lattice

site. The population of these states is simultaneously updated according to deterministic or nondeterministic rules in discrete time steps. The dynamics of CA are governed by local rules, i.e. updating site variables involves only a small number of neighbors in each time step. For this reason, CA can optimally utilize massively parallel computer technology. In addition, the locality of the dynamical rules allows an efficient and flexible treatment of complex geometries. The major characteristics of CA are the two length scales they operate on. The first described by the discrete microworld on a lattice obeying a ficticious dynamics of pseudo-particles, whose length and time scales are much shorter than the physical scales. The second is a continuous macroscale with the physical observables, which are obtained in practice by taking averages over many cells.

In its implementation for the solution of the BE, the CA consists of a lattice in position space, each site of which has a finite number of momentum states. The nondeterministic transition rules between these states associated to collision events are determined from the quantum mechanical scattering rates (in the same way as in MC) and from the classical equations of motion. Due to the locality in position space of quantum mechanical scattering events, which is a basic assumption underlying the BE, there is no principle problem to convert these transitions in momentum space into local CA-rules. On the other hand, the drift and diffusion terms in the BE link the distribution function to its value in different position and momentum space locations, being therefore nonlocal in nature. Additionally, a single semiclassical particle trajectory cannot in general be reproduced exactly on the discretized phase space of the automata. Therefore, the kinetic terms of the BE are replaced by hopping probabilities in both position and momentum space in such a way that the equations of motion are fulfilled on the average for an ensemble of pseudoparticles. By an appropriate choice of lattice constant a and timestep dt, the hopping events can be restricted to transitions between nearest neighbors (or second nearest neighbors). This choice is restricted by the desired resolution in position space and by the maximum physically relevant velocity, which must be less than a/dt. This procedure results in a master equation for the state occupancies which contains only on-site transitions between different momentum states and transitions between nearest neighbors with the same momentum state. For a rigorous derivation and a detailed discussion of the basic algorithm, see [7].

We have estimated the number of operations that translate into MFLOPS required on a scalar processor for simplified MC and CA device simulations on a fixed field distribution that includes nonparabolic bands and the standard scattering mechanisms for GaAs. We find that an ensemble MC simulation requires at least 10^3*N operations, where N is the number of particles in the simulated ensemble, while the CA requires about 10^1*N operations, providing a speed-up of two orders of magnitudes with respect to MC. The principal difference is that the CA maps the Boltzmann equation onto a set of discrete rate equations for discrete variables that involve only integer arithmetic. In momentum space, this discrete dynamics allows one to use predetermined scattering tables both for the total scattering rates as well as for the free flights. In the CA, the final k-state is picked by a simple assignment rather than by the algebraic solution of nonlinear equations as in the MC. In addition, the motion in real space is discrete in the CA and consists of deterministic hops between nearest neighbor cells. Thus, the real space motion is automatically synchronized and occurs in a strictly ordered and predictable fashion, in contrast to the MC, where the exact continuous trajectories of each

particle have to be followed. Clearly, very efficient Poisson solvers need to be coupled to the CA in self-consistent device simulations. Our experience shows that SOR algorithms are too slow and can be the bottleneck of a CA simulation, particularly for large doping gradients. Well known alternatives to SOR methods are also FACR and FFT algorithms [12]. It should be mentioned that the initialization phase of a simulation, on the other hand, requires significantly more effort in the CA method since all scattering tables need to be calculated. Fortunately, it needs to be executed only once. Modern massively parallel computers employing either the message passing or shared memory paradigm offer a substantial reduction in turnaround time, provided efficient algorithms can be implemented on such an architecture. Indeed, both the MC and the CA can be implemented on MIMD (multiple instruction multiple data) machines by assigning subdomains of the whole device to the individual processors [13]. Each processor executes one time step of the simulation on its subdomain, collects the properties of those particles that will leave the subdomain at the end of this time step, passes this information to its neighboring processors, and receives information from them about particles entering its subdomain in the next time step. Such collection of information about the particle exchange with the neighboring processors requires more computations than needed on a single processor. This overhead is the price for the distribution of the computational load onto many processors. We have calculated the ratio of this overhead for the CA and MC method. Let N_{pp} be the number of particles assigned to one processor. Within a MC approach, we estimate that it takes every processor $10^1 * N_{pp}$ operations to collect and digest the information about the outgoing and incoming particles. In the CA, this overhead is independent of the number of particles but only depends on the number of real space cells per subdomain N_{site} and the number of discrete k-states per cell N_k through the relation $10^{-2} N_k \sqrt{N_{site}}$. With typical values for device simulations, N_{pp} ~10^4, N_{site} ~ 10^4, N^k ~10^3, one finds that the time for the message-related computations required on each processor is two orders of magnitude smaller in the CA than in MC.

III. CA vs. MINIMOS simulation of Si-MOSFET

In order to test the capability of the CA to handle the complex structure of a Si n-MOSFET, a detailed comparison with a standard drift-diffusion algorithm (MINIMOS [1]) has been performed. The CA simulation employs a non-parabolic band structure for electrons and a parabolic dispersion for one effective hole band. Furthermore, optical and acoustic phonon-scattering as well as impact ionization and impurity-scattering are taken into account. Source and drain contacts are 0.24 μm long, each separated by 10 nm from the gate contact. An oxide layer of 5 nm thickness has been used (Fig. 1). The doping profile is the one calculated by the MINIMOS pre-processor. The high carrier density in a MOSFET requires large particle ensembles in order to reduce the statistical noise and we have used 300000 particles. On several scalar RISC workstations, the execution time for the CA turned out to be 10–15 times larger than MINIMOS.

In figure 2(a) we show the drain characteristics for two different gate voltages (1.5 V and 2.5 V). The gate length (L_G= 0.25 μm) has been chosen close to the limit where MINIMOS simulations can still be expected to be reliable. The drain characteristics agree well with one another, except in the regime of voltages above 2 V, where a slight velocity overshoot (which is not accounted for in the drift-diffusion approach) acts to increase the drain current.

Figure 2(b) shows the I-V characteristics for a shorter gate length of 0.16 μm. Short

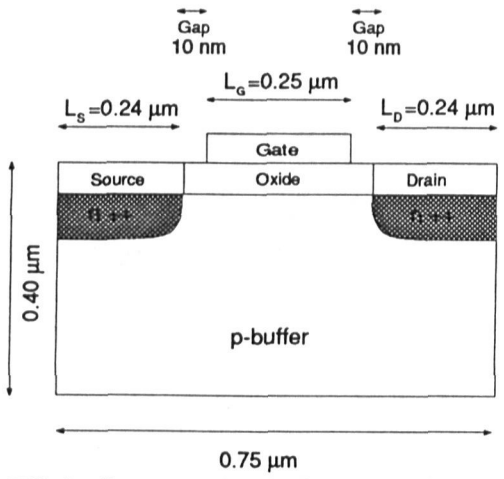

FIG 1. Geometry of a 0.25 μm MOSFET. In x-direction the device is subdivided into 126 and in y-direction into 78 blocks, respectively. The n^{++} doping for the contact region is $10^{19}cm^{-3}$.

channel effects become more important in this regime, and the two methods give significantly different drain characteristics. This originates almost totally from a significant velocity overshoot that is accounted for by the CA simulation but not by MINIMOS. In fact, as shown in Fig. 3 (a), the electron channel density of the 0.16 μm MOSFET is very similar for the two methods up to an applied voltage of 4.0 V and a gate voltage of 1.5 V. In contrast, the drift velocity as obtained in the CA shows a significant overshoot for drain voltages above 0.6 Volt, in contrast to the MINIMOS results (Fig. 3 (b)). At last we present our investigation about impact ionisation. For the cellular automaton we used a microscopic impact ionisation model for high-field energy electron transport [14] and adapt the homogeneous ionisation rate to experimental data [15]. In figure 4(a) and 4(b) position resolved distributions of impact ionisation events are given for the 0.25 μm MOSFET at a drain voltage U_D = 4.0 V and gate voltage U_G = 1.5 V has been used. Figure 4(a) illustrates the MINIMOS simulation, where the maximum of the impact ionisation is located between gate and drain. This also holds for the distribution of impact ionisation using the cellular automaton (Fig. 4(b)), but the maximum is shifted closer to the intersection of gate and drain. The overall agreement is still very satisfactory.

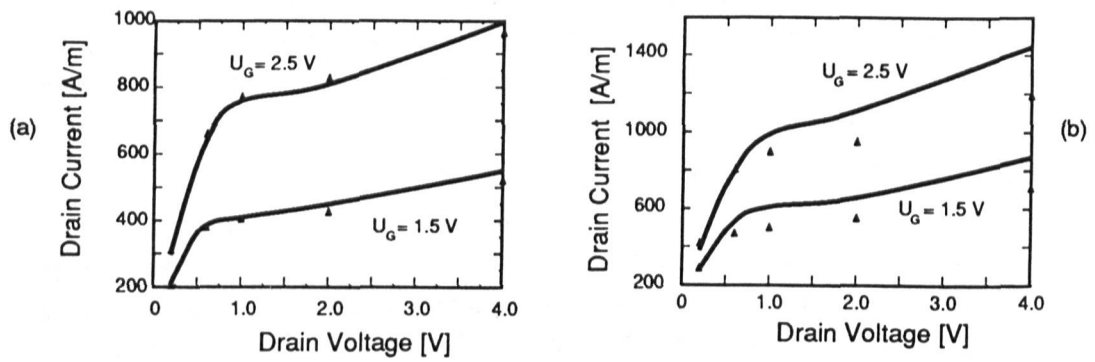

FIG 2. (a) Drain current vs. drain voltage for a 0.25 μm Si - MOSFET calculated with the drift diffusion model MINIMOS (triangles) and with the cellular automaton (full line). (b) Drain current vs. drain voltage for a 0.16 μm Si - MOSFET.

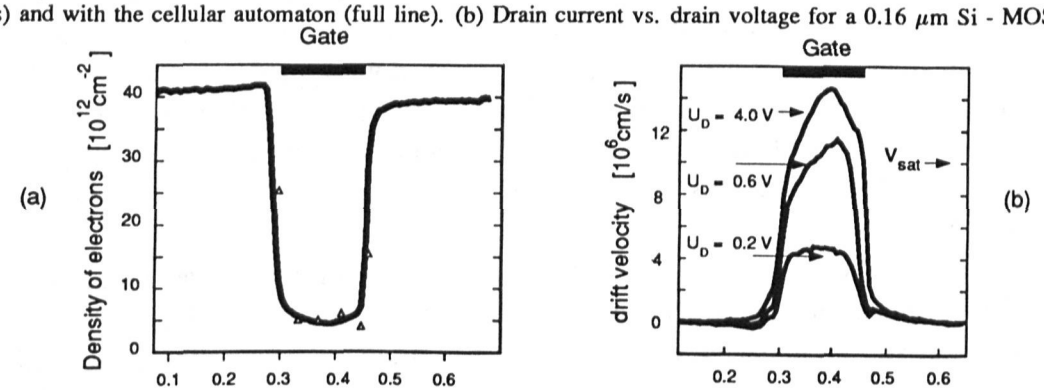

FIG 3. (a) Vertically integrated sheet density of electrons in the 0.16 μm MOSFET (U_D=4.0 Volt; U_G= 1.5 Volt). The full line represents the CA calculation and triangles results of MINIMOS. (b) Vertically weighted drift velocity under the gate (v_{sat} indicates the saturation velocity of bulk silicon).

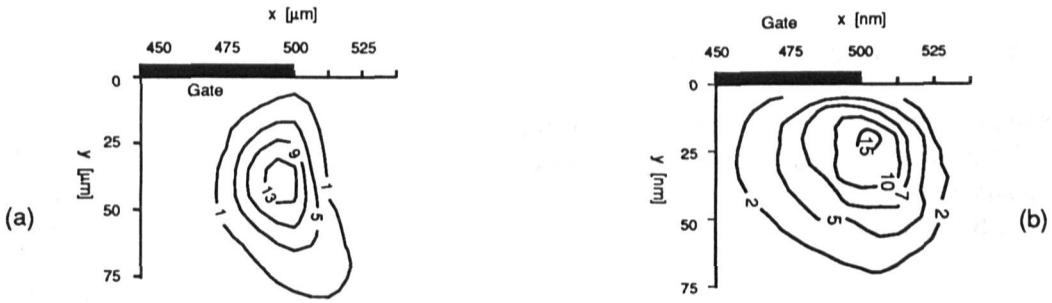

FIG 4. Contour plot of impact ionisation rate per volume [$10^{16}cm^{-3}ps^{-1}$] for a 0.25 μm MOSFET (U_D=4.0 V, U_G=1.5V) in the gate region of the device. The black wide line along the x-axis indicates the gate contact. (a) MINIMOS (b) CA.

IV. Conclusion

We have presented a detailed comparison of a novel cellular automaton (CA) technique and a standard drift diffusion calculation (MINIMOS) of high field transport in semiconductor devices. Good agreement between both methods is found for simulations of a submicron MOSFET within the regime of validity of MINIMOS, while for very short gate length nonlocal transport effects significantly influence the results of the CA simulation. In particular, the occurence of velocity overshoot in the channel of the device is responsible for the enhanced drain current in the CA simulations.

References

1. S. Selberherr, "Analysis and Simulation of Semiconductor Devices", (Springer, Wien, 1984)
2. W. L. Engl, editor, "Process and Device Modeling" (Elsevier, Amsterdam, 1986)
3. K. Hess, J. P. Leburton, U. Ravaioli, "Computational Electronics" (Kluwer Academic, Norwell, MA, 1991)
4. W. Haensch, "The Drift Diffusion Equations and its Applications in MOSFET Modeling", (Springer, 1991, Vienna)
5. M. Fischetti and S. E. Laux, *Phys. Rev.* B **38**, 9721 (1988)
6. C. Jacoboni and P. Lugli, "The Monte Carlo Method for Semiconductor Device Simulation", (Springer, 1989, Vienna)
7. K. Kometer, G. Zandler and P. Vogl, *Phys. Rev.* B **46**, 1382 (1992)
8. G. Zandler, A. Di Carlo, K. Kometer, P. Lugli, P. Vogl and E. Gornik, *IEEE Electron Dev. Letters* **14**, 77 (1993)
9. U. Frisch, B. Hasslacher, and Y. Pomeau, *Phys. Rev. Lett.* **56**, 1505 (1986)
10. U. Frisch, D. d'Humieres, B. Hasslacher, P. Lallemand, Y. Pomeau, and J. P. Rivet, *Complex Systems* **1**, 649 (1987)
11. M. G. Ancona, Solid State Electronics **33**, 1633 (1990)
12. H. Chen, W. H. Matthaeus and L.W. Klein, unpublished
13. R. W. Hockney and J. W. Eastwood, "Computer Simulation Using Particles", (McGraw-Hill, New York, 1981)
14. R. Thoma, H. J. Pfeifer, W.L. Engl, W. Quade, R. Brunetti, and C. Jacoboni, *J. Appl. Phys.* **69**, 2301 (1991)
15. C.A. Lee, R.A. Logan, R.L. Batdorf, J.J. Keimack, W. Wiegmann, *Phys. Rev.* A **134**, 769 (1964)