Cellular automata studies of vertical MOSFETs

M. Saraniti*, S. Wigger*, G. Zandler**, G.Formicone*** and S. Goodnick*

*Electrical Engineering Department,
Arizona State University, Tempe, AZ 85287-6206, USA, saraniti@asu.edu
**Physik Department and Walter Schottky Institut
TU-München, Garching , Germany, zandler@wsi.tu-muenchen.de
*** Motorola, Tempe, AZ, USA, r37968@email.sps.mot.com

ABSTRACT

This paper presents an overview of the cellular automata (CA) method for semiconductor device simulation. The main advantages of the CA method over the Monte Carlo (MC) approach are presented, and limitations of its modeling capability are discussed. As an application, systematic theoretical CA studies of vertically grown, nanometer scale, MOSFETs are presented. The predicted drain characteristics and output conductance are in excellent agreement with experimental data. The results of the simulations clearly show that in these structures impact ionization is of minor importance. The inclusion of an inhomogeneous p-doping profiles along the channel is investigated, which is shown to improve current saturation and therefore allow the reduction of the device dimensions.

Finally, preliminary results obtained with a new, improved, CA approach are shown as well. This new method includes a full-band representation of the semiconductor's electronic structure.

Keywords: Cellular automata, simulation, MOSFET, multi-grid.

CELLULAR AUTOMATON

As a numerically efficient discrete variant of the MC technique [7], the CA method [8] has been shown to significantly reduce the computational burden required by particle based simulation approaches.

In Fig. 1 the structure of a typical particle-based semiconductor simulation program is shown. The following discussion will focus on the matching of the two main components of the algorithm: the Poisson solver and the carrier dynamics simulation engine. These components are executed sequentially, along a temporal discretization scheme, in order to ensure the consistence between the charge carriers distribution and electric field.

In other words, within the adopted scheme, the field is computed at a given instant of the simulated time, and taken as constant in the course of a time step δt. During δt the dynamics of the carries under the effects of the field itself is calculated. Since the carrier distribution in phase-space changes under the effect of the electric field, and due to the interactions with the host material, a new field is computed at the end of δt, according to the new distribution.

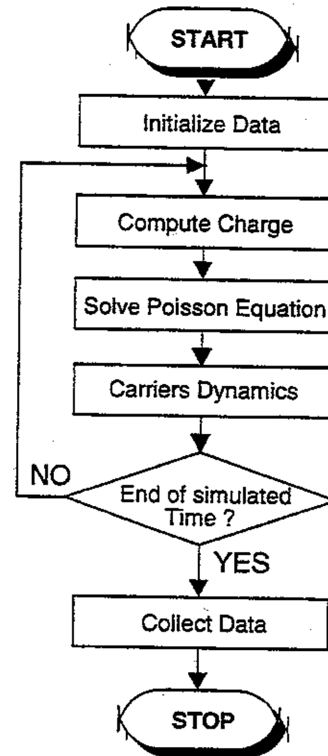

Figure 1: Simplified flow-chart of a typical Monte Carlo or cellular automaton procedure.

Since the field is kept constant during the δt temporal window, the length of such interval is crucial [5] for the correct application of the method.

While several approaches are used to solve Poisson's Equation, only two methods are known within the particle-based framework to simulate the dynamics of charge carriers. Both the CA and the MC approaches track the particle motion in the phase-space. The main difference between the two methods is the selection of the state after a scattering. The MC algorithm is based on a *scattering table* where the probability of scattering is tabulated as a function of the energy or momentum of the charge carrier. The table is accessed at the end of a given time step (shorter than δt) to pick a scattering

event for each electron, then the status after the scattering (i.e. the new position of the carrier in the momentum space) is computed based on the model used for the band structure. This computation is performed whenever a carrier scatters and requires an amount of computational resources which depends on the implemented physical models. In particular, when a numerical, non analytical representation of the band structure is used, as it is within the so–called full band [2], [9] approach, the inversion of the dispersion relation and the computation of the density of states can make the final state selection algorithm computationally burdensome.

In the CA approach a large time saving is achieved *by tabulating the final state* itself, together with the transition probability. In other words, for each initial state, all the possible final states are pre-computed and stored in a table. The final state selection is then reduced to the generation of one or two random numbers, needed to find the final state of the carrier. This method dramatically reduces the time spent to find the final state, while the required memory is consequently increased.

The main drawback of this method is the loss of precision related to the discretization of the energy (or momentum) space in the number of states which allows to represent the table in the memory of modern computers.

Recent work [10] has shown that excellent resolution may be obtained when the transition table is defined in energy space and the band structure is represented by analytical expressions within the so–called spherical, non-parabolic approximation [7]. A more accurate approach will be discussed later in the present paper, based on the full-band representation of the electronic structure.

Poisson solver

The computer time required by traditional finite difference Poisson solvers is generally comparable to the time needed to simulate the carriers dynamics in the MC approach. In the CA, because of its intrinsically higher speed with respect to the Monte Carlo, the Poisson solvers may even dominate the total numerical effort. Therefore, the development of fast and highly efficient Poisson solvers is a key issue for particle–based simulators.

Due to the repeated application of the Poisson solver within the time stepping algorithm, methods are preferred which can take advantage of a previous solution using it as initial guess.

Any standard iterative method [13] is based on the repeated application of an algorithm in such a way that a sequence of approximations is defined, which asymptotically converges to the required solution. This approach reduces the error related to each subsequent approximation, with a process which is called error reduction or *relaxation*, in the sense that the sequence of approximations relaxes toward the numerically stable solution. The spectral analysis of the error shows that Fourier components of the error are selectively reduced by the relaxation algorithm, due to the spacing of the grid used to discretize the problem. This fact makes most of the standard iterative methods too slow to be profitably used in device simulation.

On the other hand, the multi–grid method [4], [1] was found to be particularly efficient in the error reduction process, because it solves the problem simultaneously on a set of differently spaced grids. Different Fourier component of the error are thus reduced on different grids.

Table 1: Multi-grid space for the 70 nm vertical MOSFET of Fig. 2.

grid	points	X–points	Y–points	coarsening
Ω_6	12810	61	210	X,Y
Ω_5	3286	31	106	X,Y
Ω_4	864	16	54	X,Y
Ω_3	252	9	28	X,Y
Ω_2	75	5	15	Y
Ω_1	40	5	8	Y
Ω_0	25	5	5	-

The multi–grid approach has been efficiently used [12] in particle–based simulation. Table 1 shows the grid space used for some of the simulations discussed later in the paper. The simulated device considered here is the vertical MOSFET of Fig. 2 (discussed in the next section), which has an initial grid of 12810 (61×210) points. Six coarser sub-grids are built from the initial one applying double coarsening to each direction until a minimum amount of five grid points was reached. Therefore the final, coarsest grid has only 25 points. The time required to solve the Poisson equation amounts to the 25% of the total simulation time.

The implemented multi–grid Poisson solver consists of about 2000 source lines of uncommented code in Fortran 77. All the constant coefficients used in the computation including nearest neighbor tables, are determined in the setup part of the program. This procedure minimizes the time required for each multi–grid iteration, although it increases the required memory to values which are, in any case, not critical in the total balance of the employed storage resources.

The total storage requirement (in bytes) of the present implementation of the Poisson solver is given by

$$B = N_T(17W_I + 26W_R) + 20NW_R, \tag{1}$$

where N is the number of grid points in the finest (input) grid, N_T is the *total* number of grid points on all grid

sets together, while W_I and W_R are the lengths of the integer and real words in bytes, respectively.

VERTICAL MOSFET

Recently, Chemical Phase Deposition (CVD) epitaxy was utilized to grow vertical MOSFETs with channel lengths ranging from 65 to 170nm [6], [11]. This growth method allows to control the channel lenght to the nanometer scale, and use it to tune the device performance.

To investigate the influence of nonlinearmm transport on the electrical characteristics of such ultra short channel devices, we performed systematic calculations using a CA simulator for vertical MOSFETs with various dimensions. The simulated layout corresponds to the geometry shown in Fig. 2. A comparison between simulation results and experiments is shown in Fig. 5 for a 70 nm device. The agreement between CA simulation (points) and the measured currents (dashed lines) is evident.

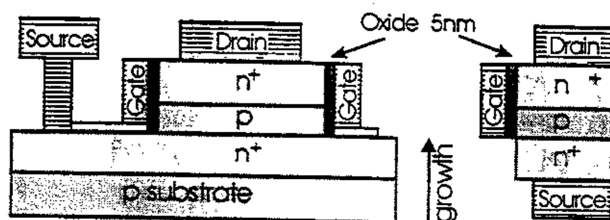

Figure 2: Layout of the vertical CVD–EPI MOS. The growth direction is shown, while the equivalent simulated region is on the right side.

The average electron velocityof a 65 nm device is shown in Fig. 3, for bias values of V_G=2V and V_D=2V. A pronounced velocity overshoot is evident over the entire length of the conduction channel, the maximum velocity being reached at the drain end of the p–buffer, where it peaks to a value of three times the saturation velocity in homogeneous Silicon.

Effects of the doping profile

The saturation behavior of the simulated devices is clearly affected by bulk parasitic conduction which occurs in the source–drain direction, away from the inversion layer. This parasitic $n-p-n$ structure acts as a barrier for electron current until its spatial extension is larger than the depleted regions at the $n-p$ and $p-n$ sides, i.e. when the central p region is not depleted.

When the dimension of the p region is reduced or, analogously, when the source-drain bias extends the aforementioned depletion regions, electrons can flow through the $n-p-n$ region, and the influence of the gate bias is negligible on the output current.

To improve the saturation characteristics of the simulated vertical MOSFET and possibly reduce its channel

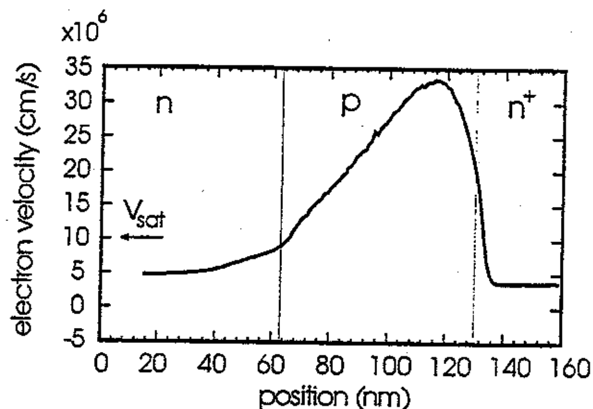

Figure 3: Average drift–velocity of electrons in the CVD–EPI MOSFET of Fig. 2; a pronounced velocity overshoot is evident in the p–buffer.

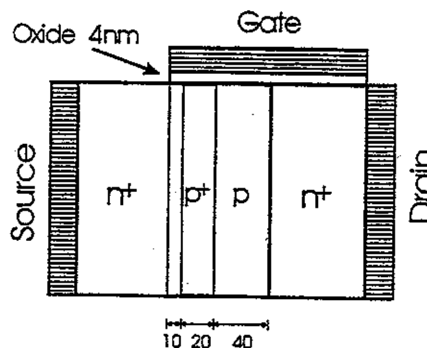

Figure 4: Cross–section of the saturation optimizing geometry. A 3×10^{18} cm^{-3} p^+–buffer is included within the central p region.

length, an additional 20 nm p^+–buffer was introduced in the central p–region. The chosen doping concentration in this p^+–buffer is 3×10^{18} cm^{-3}. The buffer was placed 10 nm from the n–p source junction, as shown in Fig. 4.

The buffer separation from the n–p source junction keeps the value of the source–junction field low enough to avoid impact ionization. Additionally, punch–through is reduced due to the higher doping. The drain current reduction due to the buffer was compensated by reducing the oxide thickness to 4 nm.

Simulation results for the device with the p^+–buffer are compared in Fig. 5 with measurements (dashed lines) and simulations (points) made on a transistor with a homogeneously doped channel. The simulated saturation, shown with continuous lines, is improved compared to the measured characteristics shown with dashed lines. The reduction of the punch–through due to the p^+–buffer allows a further reduction of the channel length to 50 nm. The simulated drain current curves of a 50 nm transistor are shown in Fig. 6.

A slight increase in the doping concentration in the

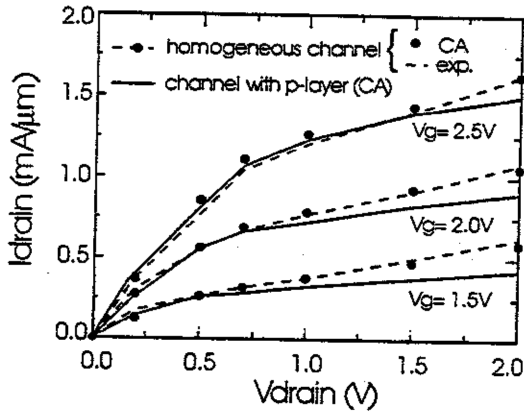

Figure 5: Improved saturation behavior of the simulated MOSFET (lines) after adding an additional 20 nm p^+ buffer to a device with homogeneousely doped channel (dashed lines, points).

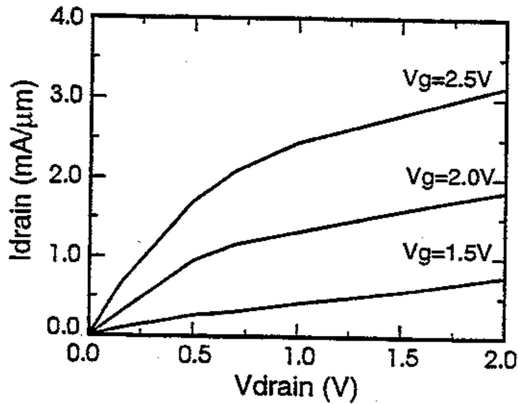

Figure 6: Current–voltage curve for a 50nm vertical MOS with inhomogeneously doped p^+–buffer. The doping concentration in the channel is 4×10^{18} cm^{-3}, whereas a density of 5×10^{18} cm^{-3} acceptors is set in the p+–buffer.

channel was used to improve the output conductance, where values of 4×10^{18} cm^{-3} and 5×10^{18} cm^{-3} were used within the channel and in the p^+–buffer, respectively.

The position of the p+ buffer shown in Fig. 4 was found to be the best in reducing punch–through and avoiding impact ionization at the drain junction. Attempts were performed to further improve this layout, in particular several graded doping profiles were tested, in order to achieve a similar performance with a lower doping concentration. Due to the short geometry, these attempt did not bring appreciable improvement of the device saturation behavior.

FULL BAND CA

The CA approach for semiconductor device simulation allows the implementation of the basic algorithms in fast and accurate programs, reducing the trade-off between modeling capabilities and computational efficiency which is typical of particle–based simulation techniques. On the other hand, the method may now be considered mature enough to consider improvements in the physical models used therein.

A full–band approach [2] is required to allow better modeling of hot electron effects at high carrier energies. The simple spherical, non-parabolic approximation of the band structure used in the previousely discussed CA is not sufficient to model the portion of carriers at high energy. In fact, the density of states (DOS) resulting from such an approximation is too simple to properly simulate the behavior of high energy carriers. A more realistic band structure is then required to obtain the correct DOS (see Fig. 7), which plays a crucial role in the scattering rate.

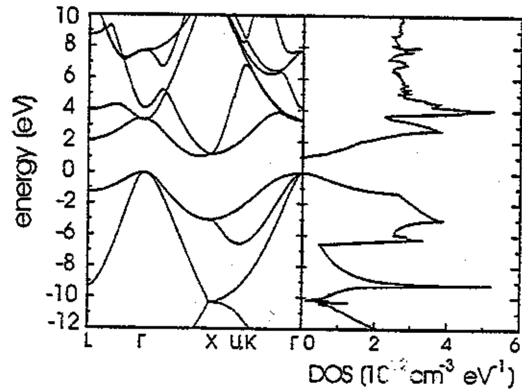

Figure 7: Silicon band structure (left) computed with empirical pseudopotentials, and density of states (right).

The full–band phonon dispersion has to be included as well in the scattering rate computation, and all phonon modes have to be considered, such that the asymmetric nature of scattering rates in momenta space is accounted for [9].

Based on such considerations, a new approach has been developed, which combines the performance of the CA with the modeling capabilities of the full–band MC.

Within the CA framework, the full Brillouin zone has been discretized on a set of *different* tensor-product non uniform grids, one for each "band", and all possible final states (in all considered bands) are tabulated for each initial cell. The grid spacing is adjusted to minimize the unacceptable [2] error due to the discretization. It is evident that the key problem of this approach is the trade–off between energy "resolution" and the amount of memory used to store the transition table. Initial results show that this problem can positively addressed.

The non-polar transition rate from a region centered in the point $\mathbf{k}$ to a region $\Omega_{\mathbf{k}'}$ centered around the point $\mathbf{k}'$ is approximated by

$$P_\nu(\mathbf{k},\mathbf{k}') = \frac{1}{\tau(\mathbf{k},\nu;\Omega_{\mathbf{K}'},\nu')} = \frac{\pi}{\rho\omega_{\eta\mathbf{q}}}|\Delta_{\eta,\nu'}(\mathbf{q})|^2$$
$$\times|\mathcal{I}(\nu,\nu';\mathbf{k},\mathbf{k}')|^2 D_{\nu'}(E',\Omega_{\mathbf{K}'})(n_{\eta\mathbf{q}}+\frac{1}{2}\pm\frac{1}{2}) \quad (2)$$

where ρ is the semiconductor density, $\omega_{\eta\mathbf{q}}$ is the frequency of a phonon of type and polarization η and wave vector $\mathbf{q}=\mathbf{k}'-\mathbf{k}$, $\Delta_{\eta,\nu'}(\mathbf{q})$ is the non-polar matrix element as defined (and approximated) in [2], $\mathcal{I}$ is the overlap integral, $D_{\nu'}(E',\Omega_{\mathbf{K}'})$ is the density of states at energy $E'=E(\mathbf{k})\pm(\hbar\omega_{\eta\mathbf{q}})$ in band ν', and, finally, $n_{\eta\mathbf{q}}$ is the phonon occupation number at the lattice temperature. In these formulas, the upper and lower signs correspond to emission and absorption of a phonon. The transition rate of Eq. 2 is computed and tabulated for each cell in the inhomogeneous grids representing each of the energy bands included in the model.

In order to compare the computed values with data of the literature, the tabulated transition rates can be summed over the final states to obtain the total scattering rate for a given band ν as a function of the momentum:

$$P_\nu(\mathbf{k}) = \sum_{\mathbf{k}'} P_\nu(\mathbf{k},\mathbf{k}'), \quad (3)$$

this can then be averaged on equienergetic surfaces to obtain the familiar scattering rates as a function of the energy:

$$P(E) = \frac{1}{D(E)}\sum_{\mathbf{k},\nu} P_\nu(\mathbf{k})\delta[E-E(\mathbf{k}')] \quad (4)$$

where $D(E)$ is the density of states at energy E. The rate in Eq. 4 represents the total electron–phonon scattering rate in function of the energy.

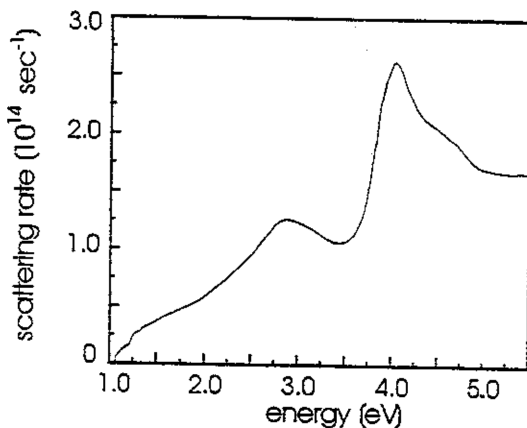

Figure 8: Electron–phonon scattering rate for Silicon at room temperature as computed from the CA transition table.

A plot of the total electron—phonon scattering rate as computed from the values in the transition table using Eq. 3 and Eq. 4 is shown in Fig. 8. The material is Silicon at room temperature. The influence of the density of states (Fig. 7) is evident. The agreement with data shown in literature [2], [3], [9] is excellent.

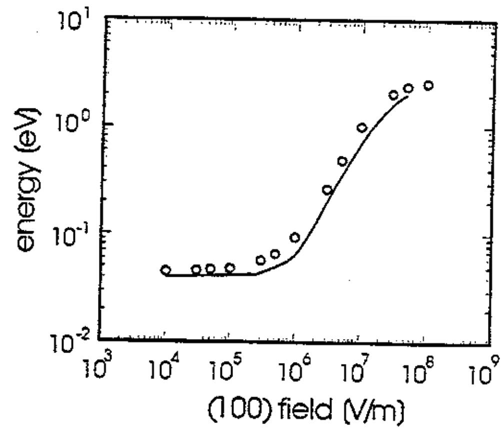

Figure 9: Simulated average electron energy as a function of the electric field. Values computed with the full-band CA (circles) are compared with full band MC results [2] (line).

A test was performed in order to evaluate the dependency of the electron energy to the electric field. The values of energy as a function of electric field for Silicon at room temperature were computed with the new full–band CA; they are compared in Fig. 9 with previously published MC data. The agreement is very good, and it can be improved by a more appropriate grid.

CONCLUSIONS

The CA approach for semiconductor device simulation was successfully used to simulate a family of CVD epitaxial vertical MOSFETs. Excellent agreement with the experiment was found and some simulations were performed in order to predict which geometry and doping profile can improve the performance of the device.

A new approach based on the CA basic idea was also proposed, such that the modeling capabilities of the CA simulation programs can include the full-band representation of the band structure of the semiconductor. A new way to compute the CA transition table was proposed, initial results were shown which agree with full-band MC predictions.

REFERENCES

[1] A. Brandt. Guide to multigrid development. In Wolfgang Hackbusch and Ulrich Trottemberg, editors, *Multigrid Methods; Proceedings of the Conference Held at Köln-Porz, November 23–27, 1981*,

number 960 in Lecture Notes in Mathematics, pages 220–312, Berlin, 1982. Springer–Verlag.

[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. *Physical Review B*, 38(14):9721–9745, November 1988.

[3] M.V. Fischetti, S.E. Laux, and E. Crabbé. Monte carlo simulation of high-energy electron transport in silicon: is there a short-cut to happiness? In Karl Hess, Jean-Pierre Leburton, and Umberto Ravaioli, editors, *Hot Carriers in Semiconductors (HCIS-9) Illinois, July 31 - August 4, 1995*, pages 475–480, New York, 1996. Plenum Press.

[4] W. Hackbush. *Multi-Grid Methods and Applications.* Springer–Verlag, Berlin, 1985.

[5] R.W. Hockney and J.W. Eastwood. *Computer Simulation Using Particles.* Adam Hilger, Bristol, 1988.

[6] F. Hofmann, W.H. Krautschneider, L. Risch, and H. Schaefer. CVD-EPI MOS transistor with a 65 nm vertical channel. In *Extended Abstracts of the 1995 International Conference on Solid State Devices and Materials, Osaka*, pages 46–48, 1995.

[7] C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Equations.* Springer–Verlag, Wien, NewYork, 1989.

[8] K. Kometer, G. Zandler, and P. Vogl. Lattice–gas cellular–automaton method for semiclassical transport in semiconductors. *Physical Review B*, 46(3):1382–1394, July 1992.

[9] T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K.Taniguchi, and C. Hamaguchi. A Monte Carlo simulation of anisotropic electron transport in Silicon including full-band structure and anisotropic impact-ionization model. *Journal of Applied Physics*, 75(1):297–312, Jan 1994.

[10] A. Rein. *Zelluläre Automaten in der Transporttheorie: Konzepte und Anwendungen.* Phd thesis, Technische Universität München, Walter Schottky Institut, Am Coulombwall; D85747 Garching - Germany, June 1995.

[11] L. Risch, W.H. Krautschneider, F. Hofmann, H. Schäfer, T. Aeugle, and Wolfgang Rösner. Vertical MOS transistors with 70 nm channel length. *IEEE Transactions on Electron Devices*, 43(9):1495–1498, September 1996.

[12] M. Saraniti, A. Rein, G. Zandler, P. Vogl, and P. Lugli. An efficient multigrid poisson solver for device simulations. *IEEE Transaction on Computer–aided Design of Integrated Circuits and Systems*, 15(2):141–150, February 1996.

[13] D.M. Young. *Iterative Solution of Large Linear Systems.* Computer Science and Applied Mathematics. Academic Press, New York, 1971.