The Status of Process and Device Simulation
Erasmus Langer and Siegfried Selberherr
Institute for Microelectronics, Technical University Vienna,
A-1040 Wien, AUSTRIA.

I. INTRODUCTION
This contribution is intended to review the state-of-the-art in numerical simulation of semiconductor processes and devices. Concerning the technological processes, the main direction for ongoing research is the extension of the simulation tools to three spatial dimensions. This is necessary for modern devices in order to provide sufficiently accurate doping profiles and device geometries for the three-dimensional device simulation. There are two major problems for three-dimensional simulations [20]: First, three-dimensional computations are very CPU-time intensive. Second, in some fields (e.g. for diffusion processes) it is simply the lack of really good and accurate models. In this paper efficient approaches to accelerate three-dimensional process simulation are sketched. Regarding the simulation of submicron devices much emphasis has to be laid on the used carrier transport model. This contribution presents a transport model suitable for ultra small MOS transistors.

II. TOPOGRAPHY SIMULATION
Topography simulation deals with the basic processes of pattern definition and pattern transfer which ultimately change the shape of the wafer surface. The numerical algorithms for surface movement play a key role in those simulators and lead to major differences in accuracy, robustness, and efficiency of the simulation tools. Basically there are two types of algorithms used for modeling three-dimensional topography processes. Volume removal methods which have been successfully used in three-dimensional lithography simulation [10, 26] can easily handle arbitrary geometries, but unfortunately they suffer from inherent inaccuracy, because they favor certain etch directions [12]. Surface advancement methods on the other hand offer highly accurate results, though with potential topological instabilities such as erroneous surface loops which result from a growing or etching surface intersecting with itself. The surface loops must be removed before they become too complex which is a rather complicated task in three-dimensional simulation [9].

A new and very promising approach [30] for the surface movement is based on fundamental morphological operations used in image and signal processing [4] and allows accurate simulation of arbitrary three-dimensional structures without loop formation. The simulation geometry is basically considered as a two-valued image (material or vacuum), and an array of square or cubic cells is used where each cell is characterized as etched or unetched. Additionally, a material identifier is defined for each cell, therefore material boundaries need not be explicitly represented. The surface or etching boundary consists of unetched cells that are in contact with fully etched cells. Cells on the surface are exposed to the etching medium or to the deposition source, and etching or deposition proceeds on this surface. A linked surface cell list stores dynamically etch or deposition rates of exposed cells. To advance the etch front spatial filter operations based on the erosion or dilation operation [4] are performed along the surface boundary. During an etching process, all cells within a filter are etched away, while cells outside stay unchanged. Usually, for anisotropic three-dimensional simulations filters are ellipsoids, for isotropic movement of surface points filters are spheres, although there is no restriction on the filter shape. The spatial dimension of an applied filter determines how far a surface point moves. The main axes of an ellipsoid are given by the local etch or deposition rates multiplied by the time step. After each time step the exposed boundary has to be determined by scanning all the cells in the material. Material cells are surface cells if at least one cell side is in contact with an already etched cell. The exposed sides of the detected surface cells describe the etch or deposition front at a certain time step.

III. ION IMPLANTATION
Ion implantation is currently the most important technique for introducing dopants into semiconductors. As modern annealing methods (RTA) do not alter the implanted profile very much anymore, the determination of the initial implantation profile has become a very important task. Thus the simulation of ion implantation gained in significance tremendously. For the simulation three main techniques can be used: The analytical description of the doping profile [25], the solution of the Boltzmann transport equation [5] or the Monte Carlo method [14, 35].
The analytical method usually has the benefit of minor demands on CPU-time consumption. One-dimensional profiles can be modeled accurately by the analytical description of profiles. Already for two-dimensional computations problems arise, because of the lack of an underlying physical base for multidimensional extensions of this technique. The limitations can be seen on examples with abrupt changes of the simulation geometry or for tilted implantations [28]. Finally, in the three-dimensional space, the CPU-time requirements increase dramatically, and this last advantage against the other mentioned methods looses its weight.

The solution of the Boltzmann transport equation is very efficient and accurately possible for one-dimensional applications [5]. For two-dimensional simulations the CPU-time and memory requirements increase significantly. Nevertheless, this method can be still advantageous compared to the two-dimensional Monte Carlo methods, when demands on accuracy are not too high. For three-dimensional simulations this method is despite today's computer power not applicable.

For the above listed reasons the Monte Carlo method is the choice for three-dimensional problems, although some special considerations are necessary to reduce the otherwise tremendous CPU-time consumption. Two main techniques have been developed to accelerate the simulation: First a superposition method is used to decrease the number of collision events to be evaluated [14], and second an octree has been introduced for the discretization of the geometry to simplify the point location problem [29].

A problem is the incorporation of the real crystalline structure of semiconductor materials. Although crystalline simulators exist [13] the time required to get three-dimensional results is very high, because the superposition method can not be used for crystalline targets. Nevertheless, Fig. 1 and Fig. 2 clearly show that the effect of a crystalline target has to be taken into account: Boron was implanted with a tilt angle of 0° and an energy of 30 keV into an amorphous target (Fig. 1) and a crystalline one (Fig. 2). There is a significant difference in the mean projected range. The ions penetrate the crystalline target about twice as deep as the amorphous material.

IV. DIFFUSION

Modeling realistic diffusion processes requires incorporation of as much physics as possible to obtain sufficient accuracy. To design a new simulation model one has to make a compromise between the number of parameters and the underlying physical relationships. As an example a robust, physically based model which is easy to understand is presented below in form of a two-dimensional simulation model for dopant diffusion in polysilicon.

Polysilicon layers are used in modern IC fabrication processes as diffusion sources, for instance for out-diffusion processes by forming a n-polysilicon-gate MOSFET or for emitter- and graft-base formation in high performance bipolar LSI's technology [31]. Any advanced polysilicon diffusion model must include various phenomena such as clustering due to the excessively high dopant concentrations and segregation kinetics to handle the exchange of dopants in the grain/grain-boundary network. To determine the impurity profile in the complex lattice polysilicon-
silicon structure, it is also necessary to include generation/recombination mechanism and grain growth kinetics. The two-dimensional coupled PDE's for the active dopant concentration in the grain interior \((C_{ga})\) and the grain boundaries \((C_{gb})\) are given in Eq. (1)-(3), where \(s\) denotes the charge state of the dopant and \(r\) is the effective grain size which can be seen as the reciprocal grain boundary area per unit volume. \(\gamma\) denotes a geometric factor taking into account the structure of polysilicon material. Diffusion and segregation kinetics is followed after [19].

\[
\frac{\partial C_{ga}}{\partial t} + G_{seg} = \\
= \text{div} \left( D_{ga} \left( \text{grad} \ C_{ga} + s \cdot \frac{C_{ga}}{U_T} \cdot \text{grad} \ \psi \right) \right) \\
\frac{\partial C_{gb}}{\partial t} - G_{seg} = \\
= \text{div} \left( \gamma \cdot D_{gb} \left( \text{grad} \ C_{gb} + \frac{C_{gb}}{r} \cdot \text{grad} \ r \right) \right) \\
G_{seg} = t \cdot \left( \frac{T_{b}^{\text{max}}}{r} - C_{gb} \right) \cdot C_{ga} - \\
- e \cdot \left( C_{g}^{\text{col}} - C_{ga} \right) \cdot C_{gb} \\
(3)
\]

Eq. (3) describes the generation/recombination term of the exchange of dopants between grain interiors and grain boundaries by use of trapping \(t\) and emission \(e\) factors, where \(T_{b}^{\text{max}}\) denotes the maximum number of free states in the grain boundary. \(C_{g}^{\text{col}}\) is the solubility limit for the dopant species; it is also taken to calculate the active grain interior concentration \(C_{ga}\) from the total interior concentration \(C_{g}\) in the static clustering model Eq. (4) after [23].

\[
\frac{C_{g}}{C_{g}^{\text{col}}} = \frac{C_{ga}}{C_{g}^{\text{col}}} + m_g \cdot \left( \frac{C_{ga}}{C_{g}^{\text{col}}} \right)^{2m_g} \\
(4)
\]

During the thermal treatment grain growth occurs. In this model the grains of polysilicon are assumed to be squares \((\gamma = 1)\) growing from initial grain size \(r_0\). The calculation of the migration of the grain boundaries is based on thermodynamic concepts of surface energy anisotropy and secondary grain growth [15]. From basic rate theory, the net rate of atomic transfer of dopants from lattice sites from one grain to those of a neighbor site is given by an complementary Arrhenius law

\[
\Delta K = K^+ \cdot \left(1 - e^{-\frac{\Delta \phi}{k_B T}} \right) \\
(5)
\]

where \(K^+\) denotes a jump frequency for atoms at the boundary and \(\Delta \mu\) is the difference in the electrochemical potential on either site of the boundary. Under constant pressure and volume the electrochemical potential is given by Eq. (6). The boundary migration \(G\) is obtained from Eq. (7).

\[
\frac{\partial r}{\partial t} = \frac{D_{ga}}{\lambda} \cdot \left[1 - e^{-\frac{\Delta \phi}{k_B T}} \right] . \\
(8)
\]

The grain growth depends on the local dopant concentration via the diffusion coefficient \(D_{ga}\), so the grain size becomes non-uniform along the vertical and lateral direction.

V. MOS TRANSISTOR SIMULATION

Enabled by the progress in the fabrication processes, the miniaturization of semiconductor devices forces improvements of the physical models as well as of the numerical procedures which the simulation of semiconductor devices is based on.

On the one hand, the numerical tools have been extended to three spatial dimensions in order to address geometry-dependent effects [1, 8, 22, 32]. On the other hand, actual devices are characterized by large electric fields in conjunction with steep gradients of the electric field and of the carrier concentrations. In many cases the cases over which the variations occur are comparable to the mean free path of the carriers — under these conditions, the widely used drift-diffusion model is losing validity. More sophisticated models, such as the hydrodynamic and energy-transport models [2, 18, 24], the spherical harmonics expansion method [6, 21, 33], and the Monte Carlo technique [3, 11, 34], overcome these limitations. However, the increased physical rigour of these models comes at the expense of increased computing times.

The so-called hybrid transport model represents a combination of the Monte Carlo method and the drift-diffusion equations [16] which is based on the following considerations: The Monte Carlo method is well suited to describe the non-equilibrium transport occurring under conditions appearing in ultra small MOS transistors (i.e. very high electric field in the active region with rapid changes over distances comparable to the mean free path of the carriers). On the other hand, for the description of low field transport the drift-diffusion model with local transport coefficients provides sufficient accuracy. Moreover, the drift-diffusion model has turned out to be even superior to the Monte Carlo method in regions with retarded fields.
Starting point for the refinement of the transport model are the so-called "enhanced drift–diffusion equations" consisting of Poisson equation (9), continuity equations for electrons and holes (10), (11), and the current relations for both carrier types (12), (13).

\[
\text{div}(\vec{J}) = -\rho \tag{9}
\]

\[
\text{div} \vec{J}_n - q \cdot \frac{\partial n}{\partial t} = q \cdot R \tag{10}
\]

\[
\text{div} \vec{J}_p + q \cdot \frac{\partial p}{\partial t} = -q \cdot R \tag{11}
\]

\[
\vec{J}_n = q \cdot \mu_n \cdot n \cdot \left( \vec{E} + \frac{1}{n} \cdot \text{grad} \left( \frac{n \cdot k_B \cdot T_n}{q} \right) \right) \tag{12}
\]

\[
\vec{J}_p = q \cdot \mu_p \cdot p \cdot \left( \vec{E} - \frac{1}{p} \cdot \text{grad} \left( \frac{p \cdot k_B \cdot T_p}{q} \right) \right) \tag{13}
\]

At the first glance these equations look like the fundamental semiconductor equations which have already been used in the famous work of Gummel in 1964 [7]. The enhancement lies in the existence of different temperatures \( T_n \) and \( T_p \) within the diffusion terms of the current relations (12), (13) as well as in the appropriate models for the carrier mobilities \( \mu_n \) and \( \mu_p \) [17].

The idea of the hybrid model now is to estimate the carrier mobilities and carrier temperatures by means of the Monte Carlo method in regimes with non-equilibrium transport only. In regions with low-field transport these physical parameters are supplied by local models thus saving computation time.

Fig. 3 and Fig. 4 show the potential and the lateral electric field at the surface of an n-channel MOSFET with \( L_{\text{gate}} = 0.15 \mu m, \tau_\text{x} = 5 \text{nm} \) simulated self-consistently by MINIMOS [27] once with the compound model (solid lines) and with the drift-diffusion model (dashed lines). The device has a metallurgical channel length of \( L_{\text{eff}} = 0.12 \mu m \) and exhibits a threshold voltage \( U_{\text{th}} = 0.24 \text{V} \); the bias conditions are \( U_G = U_D = 2.0 \text{V} \). In the Monte Carlo case, in the high field region the potential profile (Fig. 3) becomes smoother so that a significant lower lateral electric field (Fig. 4) is predicted.

REFERENCES


