ADAPTIVE GRIDS IN SPACE AND TIME FOR PROCESS AND DEVICE SIMULATORS

W.Jüngling, G.Hobler, S.Selberherr, H.Pötzl

Institut für Allgemeine Elektrotechnik
und Elektronik
Technical University of Vienna
Gußhausstraße 27-29, 1040-Vienna, AUSTRIA

Abstract - We present simulations of IC-fabrication steps and of the electric behaviour of devices using fully adpative grids in space and time. The strategies have been developed for finite differences and are independent of the physical models. Critical domains are detected automatically and carefully resolved. The additional amount of code and simulation time caused by the grid strategies is by far compensated by the reduction of CPU-time and memory caused by the optimal exploitation of the computer resources. The grid strategies become important for the development of advanced physical models for process and device simulation when critical domains in space and time cannot be estimated in advance and for simulations in two or three dimensions when a shortage of computer resources impedes the simulation.

1.Introduction

The increasing complexity of physical models in device and especially in process simulation necessitates a secure numerical environment for solving the systems of nonlinear coupled partial differential equations (=PDEs) which usually describe the processes. A spatial and transient grid is required to discretize the operators of the PDEs in space and time. Equidistant meshes have been used in the early days of simulation but turned out to be too inaccurate to match the requirements of process and device simulation of VLSI-devices. Rigid problem oriented grids followed and have contributed a good deal to the success of simulation, especially in two dimensional device simulation. The design of specific grids is mainly based on experience and a qualitative knowledge of the solution of the PDEs. The extension of a space charge layer close to a p-n junction or the migration of a dopant profile during an annealing step can

easily be estimated and a proper grid can be set up by accumulating points in domains of interest.

These methods begin to fail whenever the physical models get more complicated and domains of critical simulation in space and time are difficult to estimate. A similar problem occurs if computer resources reach their limits and the distribution of the grid points in space and time becomes critical. A three dimensional device simulation or the simultaneous solution of a large number of coupled PDEs in one or two dimensions exceeds practically the capacity of today's computers.

This paper presents strategies for fully adaptive grids in space and time for the simulation of process steps during the IC-fabrication and the evaluation of the electrical behaviour of semiconductor devices. The strategies are independent of the physical model under consideration. IC-fabrication steps with a spatial simulation domain of some tenth of a micron and process times of some hours can be handled as well as device simulations with a simulation domain of some hundred mircometers and switching times in the order of nanoseconds. Critical simulation steps in space and time are detected automatically and carefully resolved.

2.Spatial Grids

The design of a spatial grid can be split up into two independent steps: computation of the position of the maximum discretization error and refinement of the grid in the vicinity. The computation of the discretization error depends on the method of discretization. Finite differences are applied to discretize the PDEs and therefore the differential operators are approximated by their differences.

$$\left.\frac{\partial C(x)}{\partial x}\right|_{x=x_i} = \frac{C_{i+1/2} - C_{i-1/2}}{x_{i+1/2} - x_{i-1/2}} \qquad (1)$$

Eq.(1) is only exact if the distribution of C(x) can be described by a local polynomal of second order(2). The discretization error can therefore be defined as the deviation of C(x) from a polynomal

$$p(a_i,x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 \qquad (2)$$

Various methods can be used to compute the deviation between C(x) and p(a,x) but they all should obey two rules to prove in practical applications. The value of the discretization error must be insensitive to small local perturbations of the distribution C(x) as they may occur during interpolation and

the discretization error must decrease with decreasing mesh spacing. The computation of the third derivative of C(x) by numerical differentiation of C(x) can certainly not be used. The comparison between C(x) and a polynomal p(a,x) the parameters a_i of which have been obtained by a least squares fit through values of C(x) at four adjacent grid points $x_k \ldots x_{k+3}$ turns out to be a fairly good method. The large dynamic range of the values of the variables in process and device simulation permits only the control of the relative errors.

$$\text{error} = \sum_{j=k}^{k+3} \left(\frac{p(a_{ij},x)}{C(x)_{max}} - 1 \right)^2 = \text{minimum} \qquad (3)$$

The error can be computed with respect to the local maximum of C(x) (e.g. concentrations of dopants) or with respect to the maximum of C(x) in the whole spatial simulation domain (e.g. potentials).

We emphasize to use and to maintain a quasiuniform mesh in the spatial simulation domain. A grid is called quasiuniform if the ratio between two adjacent grid widths is small compared to unity. It is an important property of these grids /1/ that the discretization error decreases superlinear with the local mesh spacing. Furthermore it guarantees a smooth transition from a coarse to a fine grid. Fig.1 shows a quasiuniform mesh and compares it to an arbitrary mesh which has been created by simple bisection. The "sectio aurea" is used for the grid refinement, i.e. AC:CD=CD:DE and AC=CE, the ratio between two adjacent grid spacings can only be

$$\sqrt{\frac{5}{4}} + \frac{1}{2} = 1.618\ldots, \quad 1 \quad \text{or} \quad \sqrt{\frac{5}{4}} - \frac{1}{2} = 0.618\ldots \qquad (4)$$

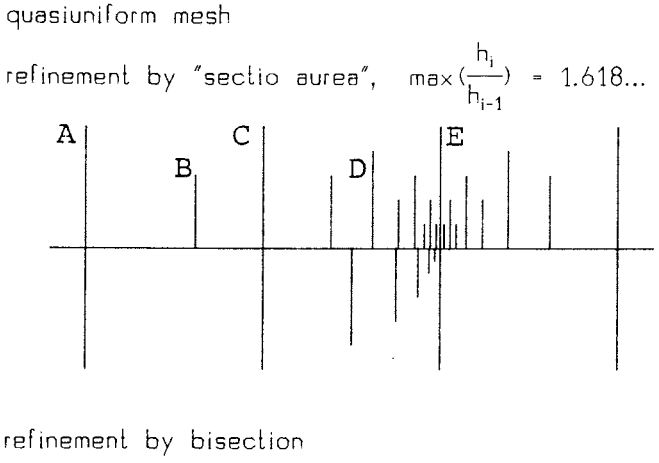


Fig.1

Fig.2 and Fig.3 show a typical application of the adaptive grid during the device simulation of a -10V reverse biased n^+-p diode. Solving the PDEs means solving the discretized system as well as providing an optimal grid for the distribution of the variables. In the example of Fig.2 five grid updates have been necessary until two succeeding grids did not differ significantly. Although all plotted distributions are solutions of the discretized system of the PDEs the corresponding profiles of the variables differ. Fig.3 shows the creation of the final grid in Fig.2. Grid points are mainly concentrated at the boundaries of the space charge layer which are in depths of 2µm and 5.5µm.

We have made the experience that it is easy to enlarge a quasiuniform mesh but it turns out much more difficult to remove single grid points. Therefore we prefer creating a completely new grid starting with an equidistant very coarse initial grid. The problem of interpolation is strongly reduced by the fact that two quasiuniform grids for similar distributions of the variables have many grid points in common. This can be seen at the top of Fig.2, where the different grids are plotted close together. This fact is explained by the use of the same initial grid

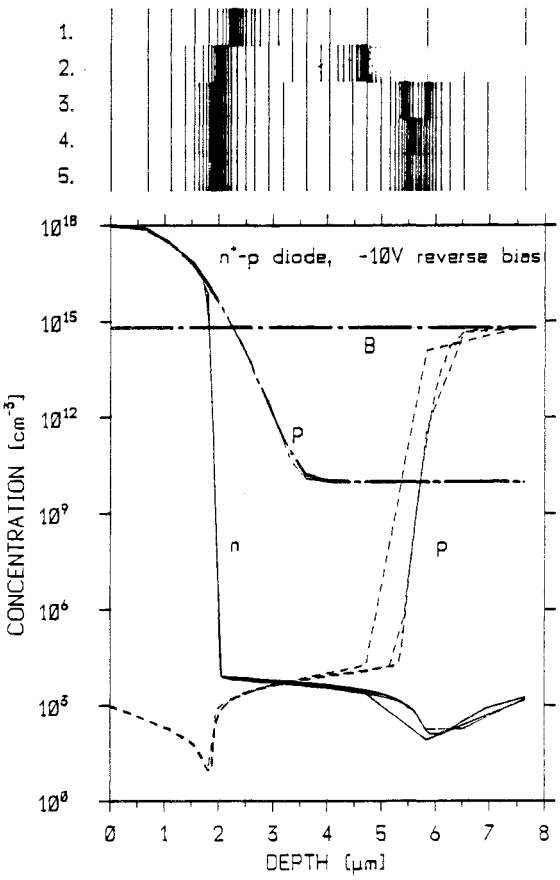


Fig.2

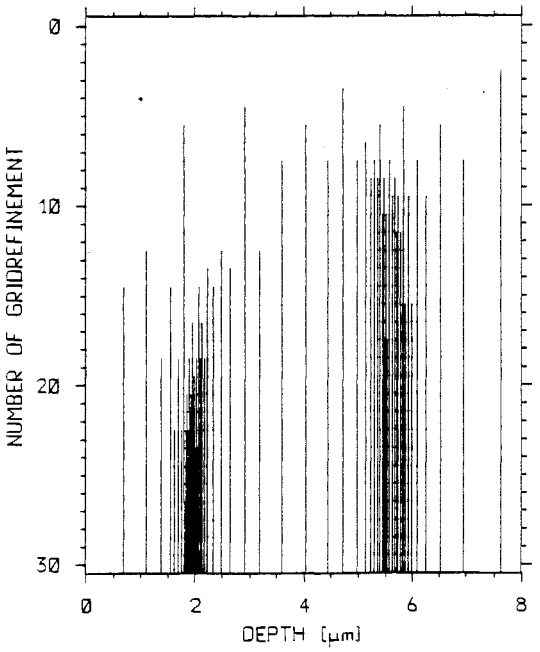


Fig.3

(eleven equidistributed points within 200μm in this example) and the technique of splitting the intervals.

3.Transient Grids

Backward Differentiation Formulas (=BDF) up to 6^{th} order are used to discretize the transient operators. BDF are nicely summarized and applied to simulations of electric networks in /2/. Although the basic ideas can be copied for process and device simulation the implementation requires major modifications. After the spatial discretization we treat the discretized PDEs as a system of 'ne'·'nx' coupled algebraic equations ('ne' is the number of PDEs, 'nx' the number of spatial grid points) and concentrate on solving

$$\frac{\partial c}{\partial t} = L(C). \qquad (5)$$

Although there is no explicit coupling between the spatial and the transient grid we observe that the automatically chosen time steps get smaller with decreasing spatial grid widths.

The large dynamic range of the variables which is usually on the order of twenty decades permits only a relative error control. The error is computed at every depth for every variable. The maximum error of all variables at a certain grid point determines the error of the transient integration at this point. Since the BDF is a predictor corrector method, the error of the transient integration is computed from the deviation of these two values. These computations are performed for the actual order of integration and the higher and lower order. The order which computes the lowest discretization error and permits therefore the largest next time step is chosen for the next integration. The comparison between the specified error of the integration and the computed discretization error determines whether the next time step will be larger or smaller than the prior one.

Interpolation of the values of the variables at inserted points during a transient integration must be carried out very carefully. We have tested four possibilities. From the mathematical point of view the interpolation using the PDEs itself is the best way to solve the problem. The PDEs are solved on a three point problem, the boundary conditions are specified as Dirichlet conditions using the values of the old grid. Unfortunately this method is far too time consuming for an implementation into a process or device simulator. A simple linear or logarithmic interpolation works well if the grid is sufficiently fine and grid points are inserted only. During grid updates starting with the coarse initial grid (typically 11 equidistributed points) when grid points are also deleted the use of this method often causes errors in the

dose of the simulated concentrations. This method can therefore not be recommended for typical process simulations where constancy of dose is important.

Computing the values of the variables on the new grid from the old distribution in such a way that the dose of the old and new distribution remains constant turns out to be fairly good to interpolate during a transient integration.

4.Examples

Process and transient device simulation are challenging examples for fully adaptive grids in space and time. Fig.4 specifies the most important process parameters of Ex.1. Fig.5 and Fig.6 show the results of the simulations, Fig.7 to Fig.9 demonstrate the corresponding grid modifications for the simulation. The physical model for the simulation is the dynamic cluster model of /3/ for arsenic and a diffusion equation for boron. The arsenic spreads from 0.25μm to 1μm and shows a steep gradient in the profile which is caused by the strongly concentration dependent diffusivity. Fig.7 shows the grid modifications in the first 40min of the annealing. The horizontal lines represent existing grid points, beginning or terminating lines indicate the insertion of a new point or the deletion of an existing point. The vertical bars at the top of the figures indicate the transient grid used for the simulation. In the very beginning the initial profile is carefully resolved which can be seen by the accumulation of grid points close to the surface and in a depth of about 0.15μm. The fine transient grid is necessary to resolve the tranformation from electrically active arsenic into clustered arsenic. Within the first 5min BDF speed up the simulation but then reduce the time steps size since the simulation of an annealing at 1000^{o}C requires a finer discretization than an annealing at 800^{o}C. During the diffusion the maximum arsenic concentration gets smaller. Since the diffusivity of As reduces with decreasing concentration the simulation becomes less critical which is indicated by the enlarging time step size. The strong accumulation of grid points in the figure indicates the spreading of the arsenic profile. Grid points are inserted and removed due to the actual spatial distribution of the variables and always guarantee an accurate discretization. The spatial grid is enlarged after every time step. Completely new grids are installed after a certain number of time steps has elapsed or after a certain number of additional grid points has been inserted. This strategy turns out to be the most effective one. The times when the grid has been updated can be detected by the "steps" in the domains of fine spatial grids.

After 30min the simulation has been interrupted. It starts again with a small step width and speeds up since the inital step width has been chosen too small for the required accuracy. As soon as the temperature reaches the 1100^{o}C the

Fig.4

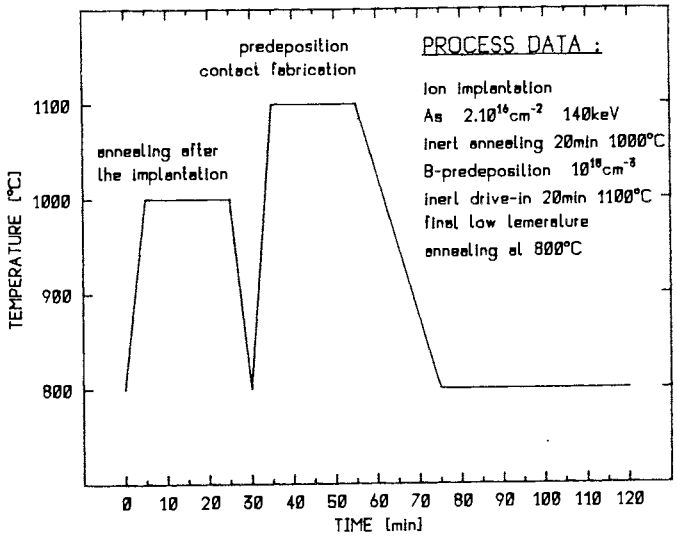


Fig.5

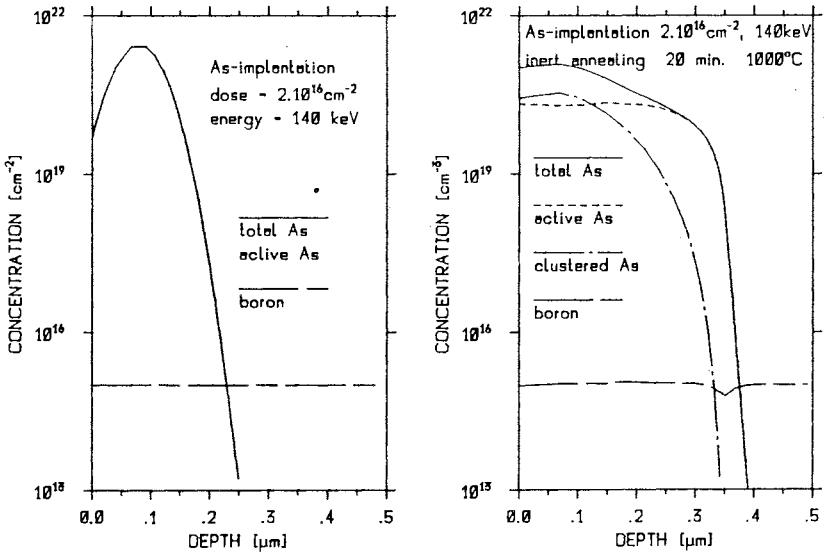


Fig.6

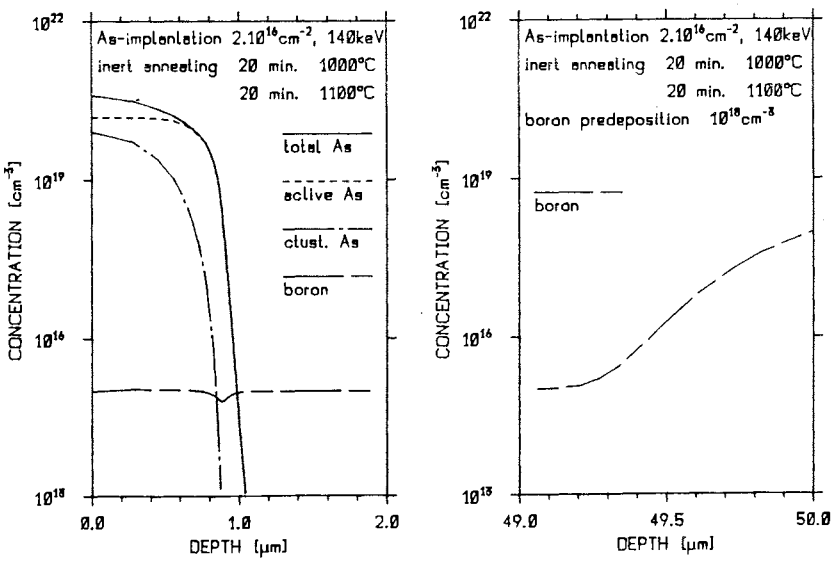


Fig.7

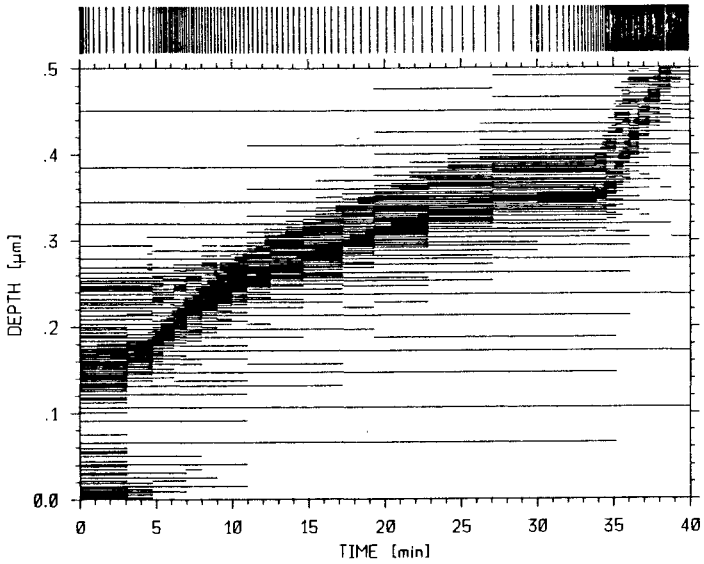


Fig.8

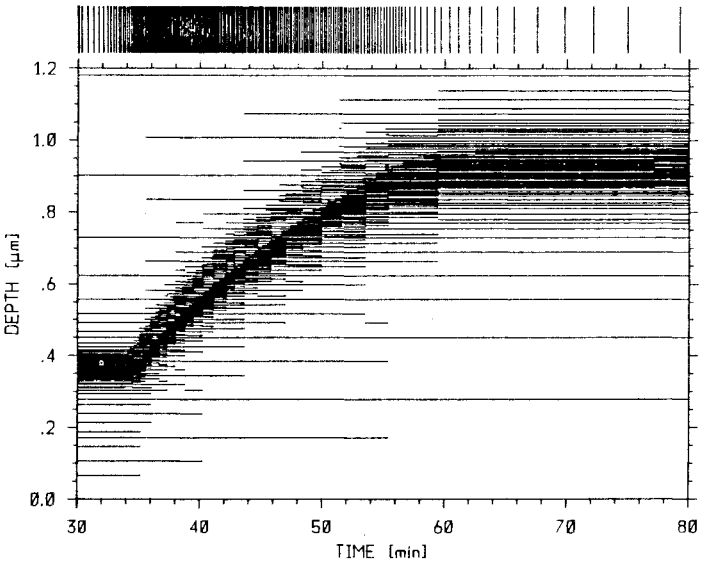
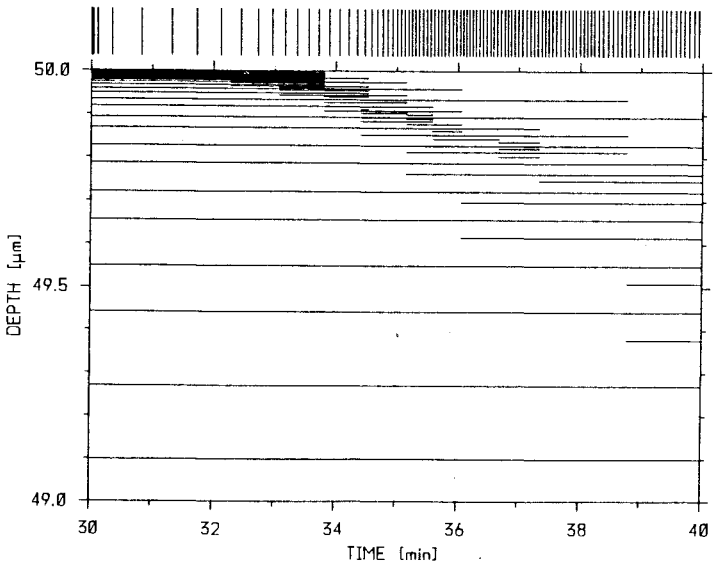


Fig.9



diffusion speeds up significantly and therefore a much more accurate simulation is required. These phenomena are demonstrated by the dense transient grid at the right top of the figure and the larger grid modifications after 35min.

Fig.8 shows the grid modifications between 30min and 80min. The fine discretization and the strong grid modifications indicate the high temperature annealing from 35 to 65min, the coarse transient grid and the static spatial grid are caused by the final low temperature annealing at $800^{\circ}C$.

The grid modifications during the boron predeposition at the rear side of the wafer are shown in Fig.9. In the beginning a fine grid is required very close to the surface. As the boron spreads from the surface into the bulk the accumulation of grid points gets less distinct and moves from the surface into the bulk.

Ex.2 shows the simulation of the switching of a n^+-p-p^+ diode from +0.5V forward bias to -4.5V reverse bias within 1µs. Fig.10 and Fig.11 show the doping profile and some snapshots of the electron and the hole concentration during the switching. Fig.12 shows the corresponding grid modifications. The figures show the creation of the space charge layer which is nicely reflected in the grid modifications. The simulation of the forward biased diode requries a fine grid close to the contacts. During the switching grid points are removed from the contacts and are inserted close to the p-n junction which is in a depth of about 2µm. The extension of the space charge layer into direction of the n^+-domain remains nearly unchanged while the extension into the p-domain increases strongly. The transient grid at the top of the figure indicates again the domains of a critical transient integration, the black domains in Fig.12 show domains of critical spatial simulation.

The ratio between the largest and the smallest grid width is usually in the oder of some thousands (c.f. /4/). The ratio between all different depth ever used during the simulation and the average number of grid points used is about four for typical device simulations and up to ten for process simulations. Since the CPU-time for the modification and creation of adaptive grids takes typically only 10% of the total simulation time the use of adaptive grids saves memory as well as CPU-time.

Fig.10

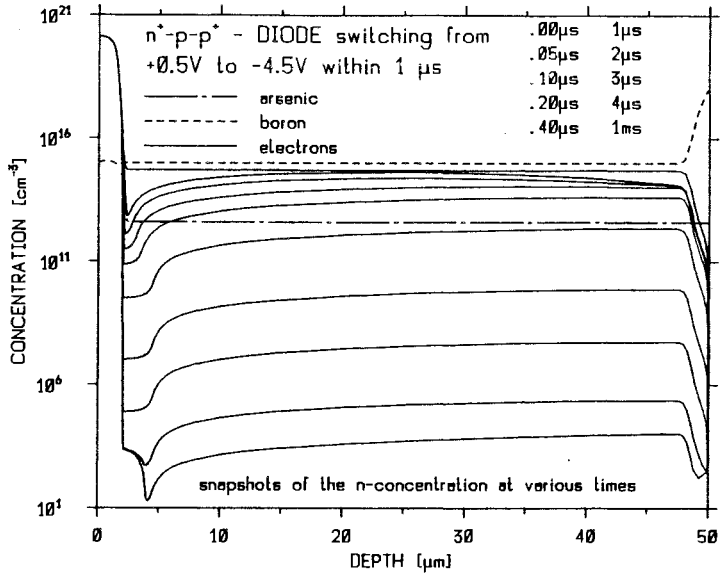


Fig.11

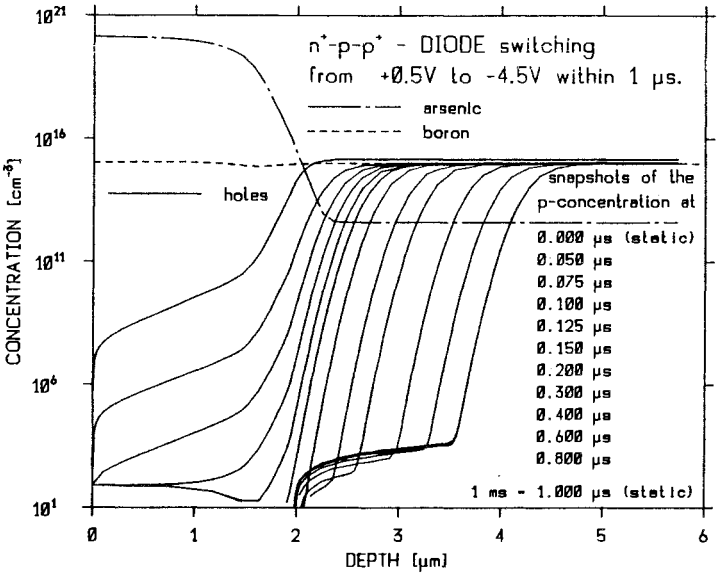
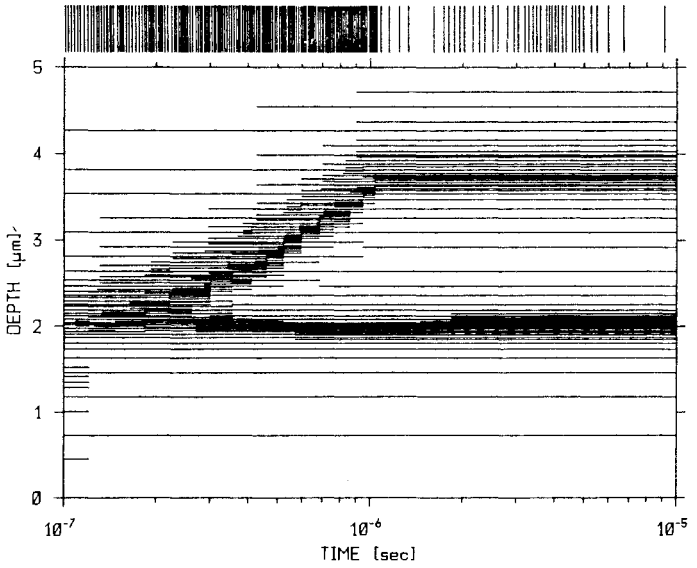


Fig.12



References

/1/ Selberherr S., "Analysis and Simulation of Semiconductor Devices", Springer Verlag Wien NewYork, 1984, ISBN 3-211-81800-6.

/2/ Brayton R.K., Gustavson F.G., Hachtel G.D., "A New Effective Algorithm for Solving Differential-Algebraic Systems Using Implicit Backward Differentiation Formulas", Proc.IEEE, Vol.60, 1972.

/3/ Tsai M.Y., Morehead F.F., Baglin J.E.E., Michel A.E., "Shallow Junctions by High-Dose As Implants in Si", J.Appl.Phys., Vol.51, No.6, 1980.

/4/ Pichler P., Jüngling W., Selberherr S., Guerrero E., Pötzl H., "Simulation of Critical IC-Fabrication Steps", IEEE Transactions on Electron Devices, Vol.ED-32, No.10, 1985.

Acknowledgment - This work has been sponsored by the SIEMENS Research Laboratories, Munich, Western Germany and the Fond zur Förderung der wissenschaftlichen Forschung, Projekt S43/10.