Spatial and Transient Grids for Process and Device Simulators

W.Jüngling, P.Pichler, S.Selberherr, E.Guerrero, H.Pötzl

Institut für Allgemeine Elektrotechnik und Elektronik
Abteilung Physikalische Elektronik, TU-Wien
Gußhausstr. 27-29, A-1040 Vienna, AUSTRIA

Abstract - A strategy for fully adaptive spatial and transient grids is presented and applied to typical and critical examples of process and device simulation. The distribution of grid points and the choice of time steps is performed without any user interaction. Transient and spatial domains of interest are detected automatically and are carefully resolved. The additional amount of code and CPU-time required for the adaptive grid is by far compensated by the optimal distribution of grid points and optimal choice of the time steps.

1.Introduction

The improvements in VLSI technology require more sophisticated models for process and device simulation. The development of new models necessitates simulators which can handle the structure of complicated models and provide the numerical background for the evaluation of the models. This paper shows how simulators should be equipped to fulfil the high demands necessary for up-to-date process- and device simulators.

In Section 2 we present some of the numerical background of our grid strategies. A critical example of a coupled boron arsenic diffusion is presented in Section 3. A transient device simulation of a simple diode will be discussed in Section 4. These examples show that the grid strategies cover a wide range of applications from fast processes of about some micro seconds in device simulation to processes of up to several hours in process simulation.

2.How to create a grid

The automatic grid adaption has to face the problems of where to refine the grid, how to refine the grid and how to interpolate the various quantities at inserted grid points.

It is the task of a spatial grid that the discretization error is equidistributed, i.e. that the partial derivatives of the quantities with respect to space are carefully resolved. Continuity equations and current (flux) relations contain spatial operators therefore the independent variables (e.g. electrons, electr. potential) and the dependent variables usually composed by first derivatives (e.g. currents, electr. field) require a fine resolution. We use the finite difference method for the discretization which is accurate if the

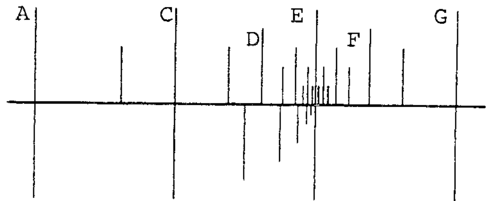
distribution of the quantites can be described locally by a polynomal of 2-nd order. If this is the case the values of the quantities at four adjacent grid points can be described by the polynomal. This will certainly not be the case in typical simulations but we use the deviations from an optimal polynomal of 2-nd order, obtained by a least squares fit, as a measure for the discretization error. We prefer using the deviations from the optimal polynomal to computing the third derivative of the quantities numerically since our method is not sensitive to slightly varying values as they may occur during a simulation.

Having defined the position of the intended grid refinement we have to actually modify the grid in the vicinity. The simplest method splits an interval into two equal ones not considering the grid spacing in the vicinity. Here the ratio between two adjacent grid distances can be far larger than two and the discretization error decreases only linearly with the local grid width. Using a quasiuniform mesh the spatial discretization error decreases with the square of the local mesh spacing as it is shown in /1/. A mesh is a quasiuniform mesh if the maximum ratio of two adjacent grid spacings minus unity is small compared to unity. We prefer a grid refinement according to the "sectio aurea" to obtain a quasiuniform mesh. The differences between the two grid strategies and the "sectio aurea" can be explained by the simple example in Fig.1. (The vertical bars indicate existing grid points. The length of the bars indicates which grid points have to be inserted at the same time to maintain quasiuniformity. The shorter the bar, the later the grid point has been inserted.) The equidistant grid ACEG is refined by D and F so that AC:CD=CD:DE and DE:EF=EF:FG.

The grid refinement factor computes to be 1.618... = sqrt(2.5)+0.5 which is smaller than two. The simple example in Fig.1 indicates that in general a larger grid domain has to be refined in order to maintain a quasiuniform mesh. The advantage of our grid is the superlinear reduction of the spatial discretization error with respect to the local grid width. The following figures in this paper show that the grid refinement factor of 1.618...

quasiuniform mesh

refinement by 'sectio aurea', $\max(\frac{h_i}{h_{i-1}}) = 1.618...$



refinement by bisection

Fig.1: Comparison between a quasi-uniform mesh (upper mesh) and a grid obtained by bisection (lower mesh)

is sufficient to obtain a short transition from a coarse to a fine grid.

Interpolation of independent variables (e.g. n, p and φ in device modeling) is not very critical if the grid provides a fine resolution. The use of the linear or exponential fitting reveals sufficiently good interpolation. The control of dependent variables (e.g. fluxes, currents) is far more difficult. Rigorous modifications in the spatial grid may lead to perturbations, especially when these secondary quantities are complicated functions of the primary ones.

For the transient integration we use backward difference formulae (=BDF) up to 6-th order as proposed by /2/. For the application in process and device simulations some modifications have been necessary. The large dynamic range of the variables (e.g. electrons may change from 1 to $10^{20} cm^{-3}$) permits only a control of the relative error. We compute the relative error with respect to the local value of the variable if it does not change its sign and is larger than zero (e.g. concentrations) or with respect to the absolute maximum value of the quantity in the whole spatial domain (e.g. potential). We compute the error of the transient integration of every variable at each grid point to control the step width and order with respect to the largest error. The order of integration may be constant or change up or down by one. The order which permits the largest time step is chosen for the integration of the next time step. The maximum increase between two adjacent time steps is limited by an artificial factor of 3.3, the reduction is unlimited.

3. Example 1: Coupled Boron Arsenic Diffusion

Example 1 shows the simulation of a coupled diffusion of boron and arsenic at 1100°C for 120 minutes. The initial distribution has been obtained by the simulation of two ion implantations the data of which are found in Fig.2. In this figure we see four snapshots of the distribution of the boron and arsenic concentrations during the diffusion at 1sec, 900sec, 3600sec and 7200sec. In the beginning boron and arsenic diffuse independently, at about 900sec the p-n junction is formed and a strong interaction between the spreading profiles takes place. At the top of Fig.2 we see

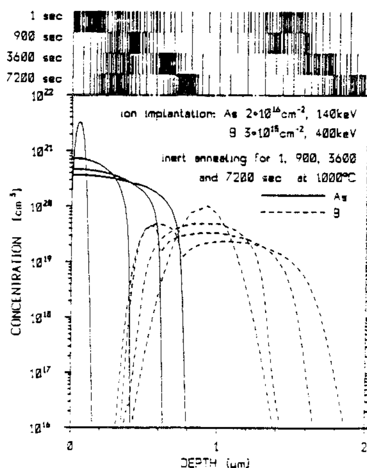


Fig.2

the spatial grids used for the simulations at the indicated times. The grid points are accumulated near the p-n junction and at domains where the impurities show a large change in their concentrations. As the distribution of boron and arsenic changes the grid follows and enables an optimal discretization for the simulation.

Fig.3 shows all the grid transformations during the simulation within the first 2 microns of the simulation domain (which was 20 microns in this example). Existing grid points are indicated by continuous lines, beginning or ending lines indicate changes in the spatial grid. The figure includes three domains of increased density of grid points. The most impressive one indicates the rapidly spreading steep arsenic profile which necessitates a fine resolution in space and time. The other domains indicate the spreading of the boron profiles towards the bulk and towards the surface. The grid indicates that the boron diffuses slower towards the surface than towards the bulk which can be explained by the modeling of the diffusivity of boron. The total number of different grid points which has been used during the simulation has been 1524 points. The average number was 161 grid points the maximum number of grid points at a certain time was about 185. The simulation domain has been 20μm, the ratio between the largest and the smallest grid width has been more than 5000. Using a rigid non equidistant grid would have taken about ten times more computation time than using an adaptive grid.

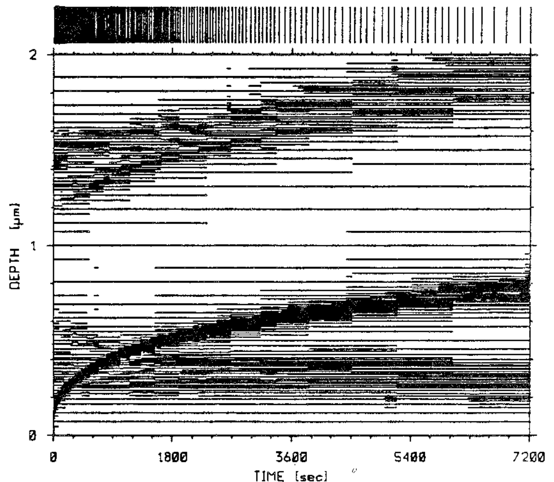


Fig.3

4.Example 2: Switching a n^+p Diode from -4.5V to +0.5V.

The profile of the n^+p diode has been obtained by the simulation of an ion implantation of phosphorus with an energy of 100keV and a dose of $10^{16}cm^{-2}$ into a $6.7\cdot10^{14}cm^{-3}$ boron-doped wafer of 200μ thickness. The implanted phosphorus has been annealed for 20 min at $1100^{0}C$. The final doping profile can be seen in Fig.4 and Fig.6.

We have applied our spatial and transient grid strategies to the simulation of switching a diode from -4.5V reverse bias to +0.5V forward bias within 1 μsec. Fig.4 shows ten snapshots of the hole

concentration during the switching process in the n-domain and in the vicinity of the p-n junction. The distribution of the holes varies during the switching which is also reflected in the grid adaption. In the beginning the -4.5V reverse biased diode necessitates a fine resolution at the boundaries of the space charge layer (i.e. at 3µm and at 5µm) the +0.5V forward biased diode necessitates a fine resolution at the contacts. These demands are reflected in Fig.5 which shows the grid modifications during the switching in the n^+-domain and near the p-n junction. Two domains of fine resolution can be seen which move towards each other. They indicate the boundaries of the space charge layer. (The boundary towards the p-domain moves faster than the boundary towards the n^+-domain). Finally, this accumulation disappears.

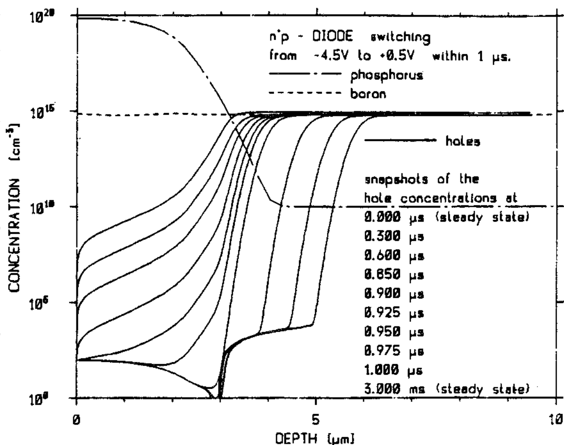


Fig.4

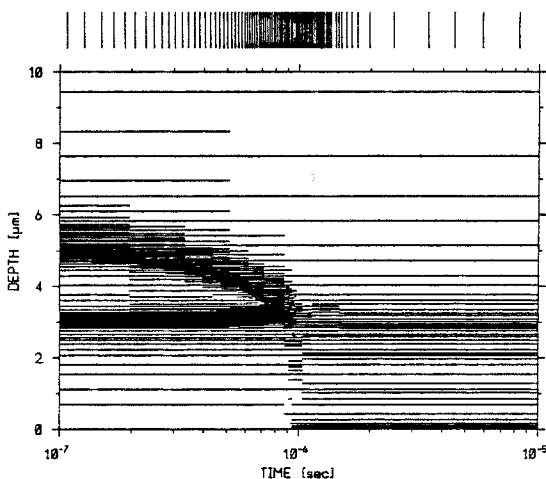


Fig.5

Thereafter, grid points are inserted at the n-contact where they become required to resolve the large gradient of the holes in a forward biased diode. The grid points between 0 and 3µm are necessary to resolve the gradient of the electrons. Fig.6 shows fourteen snapshots of the electron concentration during the switching of the diode. We observe the decreasing of the space charge layer and the injection of electrons into the low doped p-domain. After about 1.2µsec the electrons reach the p-contact. Fig.7 shows the grid modifications in the whole simulation domain. The injection of electrons and their spreading is detected and for the short period from 0.9 to 1.2 µsec grid points are inserted into the p-domain to resolve this spreading accurately. After the electrons have reached the p-contact a fine resolution at this contact is necessary to resolve the large gradient. Thereafter, the

electron distribution in the p-domain becomes constant and the grid points are therefore evicted. At the top of Fig.5 and Fig.7 the transient grid is indicated by the vertical bars. The time step width becomes shorter between 0.5µs and 1.2µs and increases afterwards. The transient grid shows extremely small time steps shortly before 1µs and about 1.2µsec. At these times the injected holes and, delayed, the electrons arrive at the contacts. The fast increasing concentrations at the boundaries necessitate fine transient resolution. After the transient processes disappear (about 2µsec) there is no reason for maintaining a fine transient grid. The time step control gets aware of this fact and speeds up the integration until 1msec is reached. The total number of different grid points during the simulation has been 436 the average number 116.

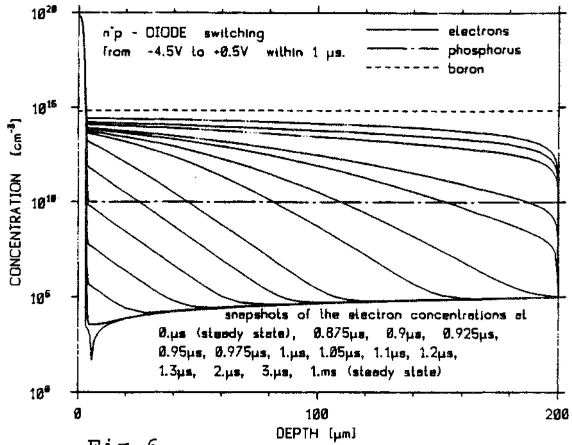


Fig.6

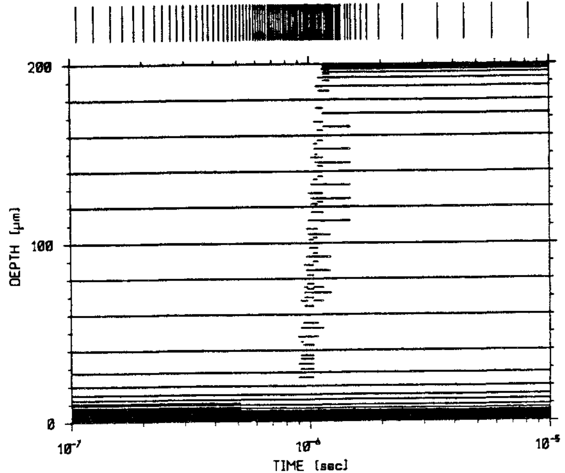


Fig.7

The ratio between the largest and the smallest spatial grid width has been about 44000 (!) caused by the large simulation domain (200µm).

5.Summary

The paper presents an approach for the implementation of adaptive spatial and transient grids for process and device simulation. The strategy takes advantage of theoretical findings such as the use of a quasiuniform spatial mesh or the use of backward difference formulae for the transient integration. We have shown that the amount of additional code and simulation time is highly compensated by the gain of CPU-time and decreased memory.

References

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

/2/ Brayton R.K. et. al., "A New Eff. Algorithm for Solving Diff. Algebraic Systems Using Implicit BDF", Proc.IEEE, Vol.60, pp.98-108

Acknowledgement - This work has been sponsored by SIEMENS Research Lab. and the "Fonds zur Förderung der wissenschaftlichen Forschung" Project S43/10.