INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING, VOL. 31, 1357–1374 (1991)

SELF-ADAPTIVE SPACE AND TIME GRIDS IN DEVICE SIMULATION

G. NANZ

Institut für Mikroelektronik, Technical University Vienna, Gußhausstraße 27–29, A-1040 Vienna, Austria

W. KAUSEL

Institut für Allgemeine Elektrotechnik und Elektronik, Technical University Vienna, Gußhausstraße 27–29, A-1040 Vienna, Austria

S. SELBERHERR

Institut für Mikroelektronik, Technical University Vienna, Gußhausstraße, 27–29, A-1040 Vienna, Austria

SUMMARY

For the solution of the three basic semiconductor equations in two space dimensions fully self-adaptive grids are a powerful tool to optimize the ratio between the number of unknowns and the accuracy of the solution. For the method of finite differences hardly any mathematically founded criterion for automatic grid control can be given which can be implemented in a computer program with acceptable effort. Therefore most of the strategies for self-adaptive grid control are totally heuristic or derived from physical properties.

In this paper we present criteria for the design of an initial space grid. We discuss some conventional strategies for the automatic control of space grids (e.g. equidistribution of the local discretization error) and present new criteria (e.g. degree of coupling of the equations, physical properties like net generation rate). For the time grid an automatic step length control algorithm is presented and the interaction between space and time grid is analysed. Throughout the paper the results are illustrated by realistic examples.

1. INTRODUCTION

For the solution of the three basic semiconductor equations fully self-adaptive grids are a powerful tool to optimize the ratio between the number of unknowns and the accuracy of the solution. It is well known that an estimation of the magnitude and the distribution of the computational error can be given only by an *a-posteriori* error analysis. Based on this error analysis the mesh can be refined in regions with large errors if this is necessary to reach a desired accuracy.

For a discretization based on finite differences hardly any mathematically founded criterion for the design of both the time and space grid can be given which might be implemented in a computer program with reasonable effort. Therefore most of the strategies to control the self-adaptive grid are based on heuristics or on physical properties.

In this paper various criteria for the design of two-dimensional space grids and of time grids are presented. We discuss strategies for the construction of initial space grids (if the simulation is started from scratch), some conventional strategies for automatic adaption of finite differences grids and we analyse new ones, since the commonly used methods like the control of the potential

0029–5981/91/071357–18$09.00 *Received 5 March 1990*

differences do not provide a sufficiently accurate grid in a computer program, which is applicable to any given device geometry.[1]

The results are illustrated by realistic two-dimensional examples (resistor, diode, lateral DMOS transistor). These have been computed with the two-dimensional device simulation program BAMBI[2,3] which is based on a totally self-consistent solution of the three semiconductor equations utilizing a Finite Boxes grid. In this program all physical models can be defined by the user; this means that in particular all effects of impact ionization can be taken into account. Data for the examples (geometries and doping profiles) can be found in the Appendix.

2. DISCRETE SEMICONDUCTOR EQUATIONS

In this section, we present the basic equations in a very rough form in order to establish the nomenclature. For detailed information about the derivation of the equations including all assumptions (e.g. linear variation of the electrostatic potential between two grid points) and restrictions, boundary conditions and physical models we refer to the book by Selberherr.[4]

Usually the behaviour of a semiconductor is described by the following set of coupled non-linear partial differential equations:

$$\operatorname{div}\operatorname{grad}\psi = \frac{q}{\varepsilon}(n - p - C) \tag{1}$$

$$\operatorname{div}\mathbf{J}_n - q\frac{\partial n}{\partial t} = qR \tag{2}$$

$$\operatorname{div}\mathbf{J}_p + q\frac{\partial p}{\partial t} = -qR \tag{3}$$

$$\mathbf{J}_n = -q(\mu_n n \operatorname{grad}\psi - D_n \operatorname{grad} n) \tag{4}$$

$$\mathbf{J}_p = -q(\mu_p p \operatorname{grad}\psi + D_p \operatorname{grad} p) \tag{5}$$

Equation (1) is Poisson's equation, (2) the continuity equation for electrons, (3) the continuity equation for holes, (4) the carrier transport equation for electrons and (5) the carrier transport equation for holes.

ψ denotes the electrostatic potential, q is the elementary charge and ε the dielectric constant of the device material. n and p denote the carrier concentrations (electrons and holes), C the net doping concentration, $\mathbf{J}_n$ and $\mathbf{J}_p$ the current densities for electrons and holes, R the net generation and $\mu_{n,p}$ the mobilities of electrons and holes, respectively.

For our following investigations we assume the finite differences discretization (6)–(8) of (1)–(3) of Gummel–Scharfetter type in scaled form.[4] For simplicity we have taken only the stationary part and neglected the time dependent terms $\partial n/\partial t$ and $\partial p/\partial t$. It is remarked that this is not a loss of generality of our strategies for space grid control.

$$\psi_{i,j-1}\lambda^2\frac{h_{i-1}+h_i}{2k_{j-1}}$$

$$+\psi_{i-1,j}\lambda^2\frac{k_{j-1}+k_j}{2h_{i-1}}$$

$$-\psi_{i,j}\lambda^2\left(\frac{h_{i-1}+h_i}{2k_{j-1}}+\frac{k_{j-1}+k_j}{2h_{i-1}}+\frac{k_{j-1}+k_j}{2h_i}+\frac{h_{i-1}+h_i}{2k_j}\right)$$

$$+ \psi_{i+1,j}\lambda^2 \frac{k_{j-1}+k_j}{2h_i}$$

$$+ \psi_{i,j+1}\lambda^2 \frac{h_{i-1}+h_i}{2k_j}$$

$$- (n_{i,j} - p_{i,j} - C_{i,j}) \frac{h_{i-1}+h_i}{2} \frac{k_{j-1}+k_j}{2} = 0 \tag{6}$$

$$n_{i,j-1} D_{n|i,j-1/2} B\left(\frac{\psi_{i,j-1}-\psi_{i,j}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_{j-1}}$$

$$+ n_{i-1,j} D_{n|i-1/2,j} B\left(\frac{\psi_{i-1,j}-\psi_{i,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_{i-1}}$$

$$- n_{i,j}\left(D_{n|i,j-1/2} B\left(\frac{\psi_{i,j}-\psi_{i,j-1}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_{j-1}}\right.$$

$$+ D_{n|i-1/2,j} B\left(\frac{\psi_{i,j}-\psi_{i-1,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_{i-1}}$$

$$+ D_{n|i+1/2,j} B\left(\frac{\psi_{i,j}-\psi_{i+1,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_i}$$

$$\left. + D_{n|i,j+1/2} B\left(\frac{\psi_{i,j}-\psi_{i,j+1}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_j}\right)$$

$$+ n_{i+1,j} D_{n|i+1/2,j} B\left(\frac{\psi_{i+1,j}-\psi_{i,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_i}$$

$$- n_{i,j+1} D_{n|i,j+1/2} B\left(\frac{\psi_{i,j+1}-\psi_{i,j}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_j}$$

$$- R_{i,j} \frac{h_{i-1}+h_i}{2} \frac{k_{j-1}+k_j}{2} = 0 \tag{7}$$

$$p_{i,j-1} D_{p|i,j-1/2} B\left(\frac{\psi_{i,j}-\psi_{i,j-1}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_{j-1}}$$

$$+ p_{i-1,j} D_{p|i-1/2,j} B\left(\frac{\psi_{i,j}-\psi_{i-1,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_{i-1}}$$

$$- p_{i,j}\left(D_{p|i,j-1/2} B\left(\frac{\psi_{i,j-1}-\psi_{i,j}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_{j-1}}\right.$$

$$+ D_{p|i-1/2,j} B\left(\frac{\psi_{i-1,j}-\psi_{i,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_{i-1}}$$

$$+ D_{p|i+1/2,j} B\left(\frac{\psi_{i+1,j}-\psi_{i,j}}{U_T}\right) \frac{k_{j-1}+k_j}{2h_i}$$

$$\left. + D_{p|i,j+1/2} B\left(\frac{\psi_{i,j+1}-\psi_{i,j}}{U_T}\right) \frac{h_{i-1}+h_i}{2k_j}\right)$$

$$+ p_{i+1,j} D_{p|i+1/2,j} B\left(\frac{\psi_{i,j} - \psi_{i+1,j}}{U_T}\right) \frac{k_{j-1} + k_j}{2h_i}$$

$$- p_{i,j+1} D_{p|i,j+1/2} B\left(\frac{\psi_{i,j} - \psi_{i,j+1}}{U_T}\right) \frac{h_{i-1} + h_i}{2k_j}$$

$$- R_{i,j} \frac{h_{i-1} + h_i}{2} \frac{k_{j-1} + k_j}{2} = 0 \tag{8}$$

where h_i and k_j are the distances of the grid lines in the x- and y-directions, $\psi_{i,j}$, $n_{i,j}$, $p_{i,j}$, $C_{i,j}$ and $R_{i,j}$ are the electrostatic potential, the electron and the hole concentrations, the impurity concentration and the net generation, respectively, at the meshpoint (i, j), λ^2 is the scaling factor[4], $U_T = kT/q$ is the thermal voltage (k Boltzmann constant, T absolute temperature), $D_{n|i+1/2,j}$ and $D_{p|i+1/2,j}$ are the effective diffusivities of the electrons and the holes, respectively, in the middle of the interval $[(i, j), (i + 1, j)]$, and $B(x)$ denotes the Bernoulli function, which is defined by $B(x) = x/(e^x - 1)$.

By the Einstein relation we can write $D_n = \mu_n U_T$ and $D_p = \mu_p U_T$.

3. SELF-ADAPTIVE SPACE GRIDS

In a computer program which is applicable to arbitrary device geometries the optimal mesh can be obtained only by self-adaptive grid control. Thereby the time consuming part of mesh design shrinks to the simple definition of the geometry.

3.1. Initial space grid

The design of an initial space grid is a very important task even if it does not seem so on first glance. A large number of experiments has shown that the numerical process is very sensitive to changes in the mesh at the beginning of a simulation. This means that the number of unknowns at the end of the computation may vary by about 25 per cent, depending on the start grid. Furthermore the convergence behaviour is strongly dependent on the initial mesh.

We have tested and implemented strategies which take care of the information about the behaviour of the solution given by the geometry, the doping profile, some numerical properties of grids and a physical criterion.

The methods for the construction of an initial grid in an arbitrary geometry to the demands of the discretization formulae have been treated by Fischer.[5]

3.1.1. Geometry information. For the design of an initial grid the device geometry has obviously to be taken into account. In many cases it is sufficient to use only the pure geometric values (co-ordinates, boundary conditions, etc.), but in some other cases it is necessary to take into account the behaviour of the solution, depending on the geometry.

From theory it is known that, at points with a change of the boundary condition (Neumann–Dirichlet boundary (contact), origin of cylindrically symmetric co-ordinate system), a singularity in the electric field will arise. This knowledge is of major importance if there is no further criterion to construct an initial grid (e.g. in MESFET's with a constant doping profile, cf. Section 3.1.2).

In these cases we introduce one new grid line in every co-ordinate direction close to such a 'singular' point to account for the influence of the singularity. This is also a possibility to isolate

the singularity, i.e. in a neighbourhood of such a point new points must not be inserted, since it is not necessary for the accuracy of the solution to refine the mesh to 'infinity' in these regions.

In Figures 1(a), (b) the initial meshes of the simple resistor ((a) of Appendix) can be seen. In Figure 1(a) the knowledge of the singularity has not been taken into account, in Figure 1(b) the three lines close to the point with the change in the boundary condition have been introduced.

Only by this modification of the grid could an accurate solution of the problem be achieved, otherwise the grid did not account for the singularity in the electric field.[6] It should be mentioned that a similar method can be used at Neumann boundaries in order to reduce the discretization error if Neumann boundary conditions of first order are implemented in a program.

3.1.2. Analysis of the doping profile. It is a conventional method for the design of an initial space grid to analyse the doping profile. In order to resolve *pn*-junctions in a proper way the gradients of the doping profile are evaluated. In regions with large gradients new points are inserted, in regions with small gradients a coarse grid is assumed to be sufficient. This may cause troubles if abrupt junctions occur. In this case a refinement to infinity would occur because of the singular gradient. Therefore we have introduced a minimum distance between grid points which

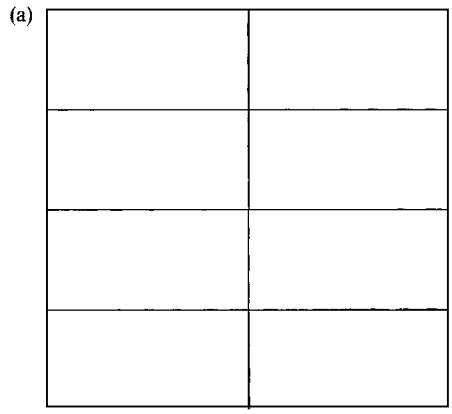

Figure 1(a). Initial space grid of a resistor without the knowledge of the singularity

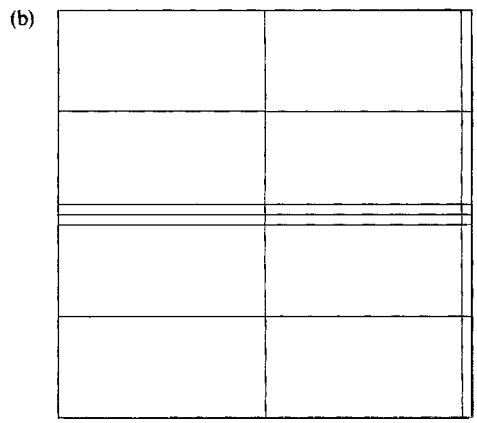

Figure 1(b). Initial space grid of a resistor accounting for the singularity

is controlled by the Debye length of the local doping. In the program BAMBI we have chosen the value $10\sqrt{[(\varepsilon U_T)/(q|C(x, y)|)]}$ for the minimum distance of points.

This general criterion provides a good initial grid if the doping is not assumed to be constant (cf. Section 3.1.1). Depending on the complexity of the device and the desired accuracy of the simulation, the upper limit for the doping gradient above which points are inserted may be varied by the user.

3.1.3. Quasiuniform grid. In the paper by Jüngling *et al.*[7] it is emphasized to use a quasiuniform space grid. This property guarantees superlinear behaviour of the local truncation error of the discretized equations.[4] Furthermore, a smooth transition from a coarse to a fine grid is obtained which significantly increases the convergence speed of the Newton iteration cycle for the solution of the non-linear system.

The well known *sectio aurea* (*golden section*), however, is applicable only for one-dimensional grids, since for two-dimensional meshes the number of points will grow too fast.

Therefore we have implemented a criterion which is weaker than the *sectio aurea*. It checks for the ratio of adjacent grid spacings, and the limit is set to 3 and 1/3, respectively, thus avoiding a too rapidly increasing number of points. In Figure 2 the result of this criterion can be seen. The values mentioned above have been established by numerical experiments.

3.1.4. Physical criterion. One additional strategic criterion for the design of an initial grid should be mentioned. From device physics it is known that near the boundary between the oxide and the semiconductor region (interface) an inversion layer may be built up. If the space grid near interfaces is too coarse this effect may even be lost in the simulation. Therefore it is necessary to introduce a line very close to the interface. We have taken a maximum distance of $1{\cdot}0 \times 10^{-7}$ cm from the interface. This additional line must also be inserted even if at the opposite side of the oxide no contact is specified. This can be seen from various trench-structures.[8]

3.1.5. Example. In this paragraph we present an example of the construction of an initial grid for a lateral DMOS transistor. In Figure 3 the start grid after the analysis of the pure geometry can be seen. In this case the singular points have not been refined since the doping profile is not

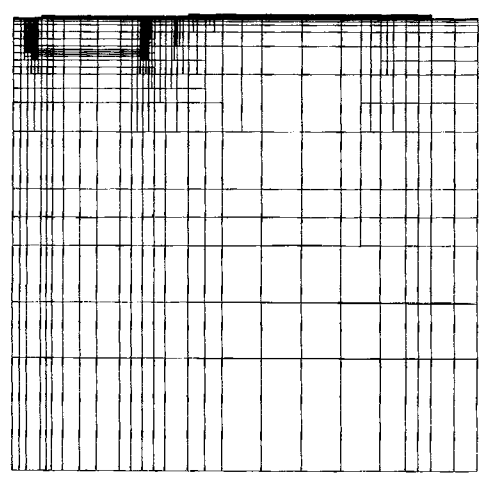

Figure 2. Initial grid of DMOS transistor after the analysis of all criteria from section 3

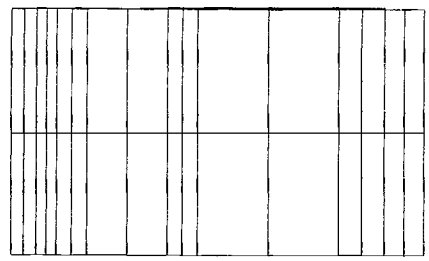

Figure 3. Initial grid of DMOS transistor after the analysis of the geometry

assumed to be constant. In Figure 2 the initial mesh of the simulation is shown. One can nicely see how parallel to the interface there is an accumulation of grid lines. The proper treatment of *pn*-junctions and the smooth transition from a coarse to a fine mesh can also be seen.

3.2. Self-adaptive space grid refinement

In order to reach a desired accuracy of the solution the space grid is successively adapted to provide an optimal mesh for the particular operating point which has to be calculated. For the finite element method various criteria for grid refinement have been developed;[9] for the finite differences method, however, only few investigations and criteria have been published. In the following we will discuss some strategies.

3.2.1. Local discretization error. It is a conventional and mathematically founded way in grid processing to equidistribute the local discretization error of the solution over the whole integration domain.[10,11] Usually only Poisson's equation is considered since mathematical tools fail in treating the whole system or the numerical effort will be very time consuming. The order of magnitude of the local discretization error also indicates in some sense the global accuracy of the solution.

The discrete Poisson equation (6) has a local truncation error linearly proportional to the mesh spacing and the third partial derivatives of the electrostatic potential on a non-uniform mesh.

For the discrete continuity equations (7), (8) we get a local truncation error linearly proportional to the mesh spacing and the sum of the first and second partial derivatives of the current density components (for the considered equation). This means that the Gummel–Scharfetter discretization scheme is a first order method which provides linear convergence.

In the following, $\rho(x, y) = (q/\varepsilon)\,(n(x, y) - p(x, y) - C(x, y))$ denotes the space charge.

For the numerical treatment of this criterion the right-hand side of (1) is considered. $\rho(x, y)$ is expanded in a Taylor series around the point $(x + h_i, y + k_j)$, where h_i and k_j again indicate the ith and jth mesh spacing in the x- and y-directions:

$$\begin{aligned}\rho(x + h_i, y + k_j) = {} & \rho(x, y) \\ & + h_i\rho_x(x, y) + k_j\rho_y(x, y) \\ & + \frac{h_i^2}{2}\rho_{xx}(x, y) + \frac{h_ik_j}{2}\rho_{xy}(x, y) + \frac{k_jh_i}{2}\rho_{yx}(x, y) + \frac{k_j^2}{2}\rho_{yy}(x, y) \\ & + O(g_{i,j}^3)\rho'''(x, y) \qquad (9)\end{aligned}$$

where $g_{i,j} = \max(h_i, k_j)$, $\rho_x = \partial\rho/\partial x$, $\rho_y = \partial\rho/\partial y$ and ρ''' derivatives of ρ of order 3 and higher.

In the discretization formulae (6)–(8) only the first term $\rho(x, y)$ is considered. The truncation error of (9) will be integrated over a rectangular domain around each point, thus providing the simple formula (10) for a first order approximation of the local truncation error Δ_ψ.

$$\begin{aligned}\Delta_\psi &= \int_{-k_{j-1}/2}^{k_j/2}\int_{-h_{i-1}/2}^{h_i/2} (\rho(x+\zeta, y+\eta) - \rho(x, y))\,\mathrm{d}\zeta\,\mathrm{d}\eta \\ &= A(\tfrac{1}{4}(h_i - h_{i-1})\rho_x + \tfrac{1}{4}(k_j - k_{j-1})\rho_y \\ &\quad + \tfrac{1}{24}(h_i^2 - h_i h_{i-1} + h_{i-1}^2)\rho_{xx} + \tfrac{1}{32}(h_i - h_{i-1})(k_j - k_{j-1})\rho_{xy} \\ &\quad + \tfrac{1}{32}(k_j - k_{j-1})(h_i - h_{i-1})\rho_{yx} + \tfrac{1}{24}(k_j^2 - k_j k_{j-1} + k_{j-1}^2)\rho_{yy})\end{aligned} \tag{10}$$

where $A = \frac{1}{4}(h_i + h_{i-1})(k_j + k_{j-1})$ denotes the integration area.

It should be remarked that the differentiation of ρ with respect to the co-ordinate directions x and y is a disadvantage in the numerical treatment of this formula. From theory $\rho_{xy} = \rho_{yx}$ holds, but in the discrete case very high differences for these values may occur, strongly depending on the mesh spacing. These differences may be taken directly as a criterion for grid refinement but then the first derivatives are neglected and it becomes much more difficult to find a rigorous mathematical explanation. The resulting grids, however, are almost the same for both strategies. In Figure 4 the *distribution* of the local discretization error of the diode from Figure 5 (0·3 V in backward direction) can be seen. The values are scaled to unity therefore it contains no information about the *quantity* of the error. It turns out (as is expected) that the error is concentrated at the junctions (boundaries of the space charge regions) and at the singularity at the change of the boundary condition (cf. Section 3.1.1).

The equidistribution of the local discretization error is a basic property that must be fulfilled to reach a given accuracy. However, various numerical experiments have shown that this criterion is not sufficient. In the following some new criteria for automatic grid adaption will be proposed and discussed.

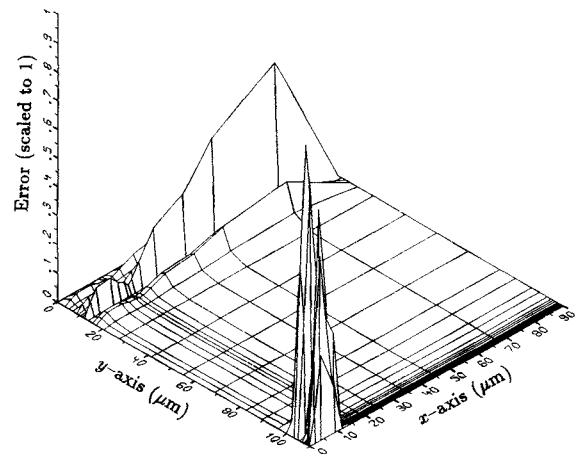

Figure 4. Distribution of the local discretization error in a diode

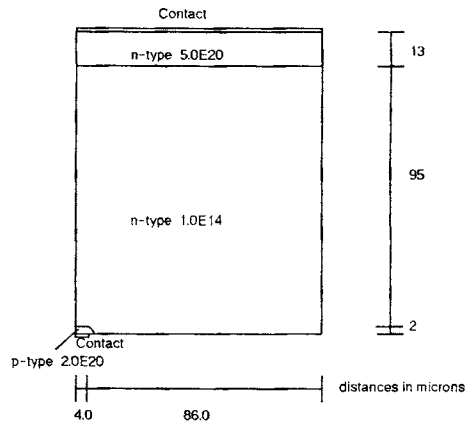

Figure 5. Geometry of the diode

3.2.2. Degree of coupling of the equations. The degree of coupling of (6), (7), (8) is mainly determined by the magnitude of the Bernoulli function

$$B\left(\frac{\psi_i - \psi_{\mathrm{NB}_i}}{U_\mathrm{T}}\right)$$

where ψ_i denotes the electrostatic potential at the grid point i with its neighbour NB_i. As the difference of the potential values increases, B tends to 0. In this case either the electron or the hole concentration at the point i becomes almost independent of the potential at the neighbouring point NB_i in the grid, as can easily be recognized from (7) and (8).

It should be noted that this effect becomes even worse for low temperatures since U_T will be decreased. Furthermore, some coefficients in the matrix of the linearized equation system may become zero, thus resulting in a singular matrix.

Limiting the Bernoulli function B by insertion of new lines is (under equithermal conditions) in fact a restriction to the electrostatic potential ψ. It is well known that a rigorous restriction to the electrostatic potential is too strong because the local truncation error of Poisson's equation is linearly proportional to the mesh spacing and the *third* partial derivatives of ψ. This means that the differences of the potential values do not affect the local discretization error if the space charge does not vary in this region.

On the other hand the magnitude of the Bernoulli function is of fundamental importance for elliptic and mixed (elliptic–parabolic) problems where the behaviour of the solution is determined by the boundary conditions.

For the first inner neighbour of a Dirichlet point both $B(x)$ and $B(-x)$ appear in the discretization formulae either in (7) or in (8). Thus for these points the magnitude of $B(|x|)$ has to be checked. If decoupling occurs, unconditionally new lines have to be inserted otherwise the linear equation system becomes singular. Vanishing values of B at inner points in regions with varying space charge and at contact points have to be avoided. Furthermore, in some cases saddle points (punch-through) and peaks in the electrostatic potential (transient simulations) may occur. In these cases decoupling has also to be avoided.

In the program BAMBI we have implemented two strategies. *During* the Newton iteration cycle (set-up of coefficient matrix, elimination process) the main diagonal elements of the

linearized equation system are checked for zero values. If a zero main diagonal element occurs this indicates local decoupling. The value is set to a very small number to avoid singularity of the linear equation system. Thereby a slightly changed problem is solved. In this case the solution has to be computed again on a grid which avoids these effects. *After* each Newton iteration cycle the electrostatic potential is checked for high gradients and—if the space charge is varying more than a given limit—new lines are inserted. By repeated application of this criterion difficulties due to vanishing values of B can be avoided for the next iteration cycle.

It should be remarked that we have proposed a criterion that provides not only the region where new lines have to be inserted but also the direction from the considered point.

3.2.3. Physical criteria. Various physical criteria for grid refinement may be implemented. In Section 3.1.4 a strategy for the initial grid has been proposed. We have extensively tested one particular criterion that is based on the user defined net generation. In regions with a net generation of more than $|10^{22}|\,\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$ new points have to be inserted to take all effects in the device into account. This 'magic' number of $|10^{22}|\,\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$ has been established by a large number of numerical experiments. For these new points a minimum distance of approximately $60 \times 10^{-8}\,\mathrm{cm}$ has been introduced, which is about the mean free path of electrons in silicon. Using this criterion a significant gain of accuracy is obtained; furthermore, the convergence speed of the Newton iteration scheme is increased.

Another physical criterion should be mentioned. The smoothness of the current densities can be taken as a measure for the quality of the solution. Since the current densities are assumed to be smooth the differences between two grid points should be limited. This again provides a criterion that indicates the direction of inaccuracy since the current densities can be split into the co-ordinate directions.

The smoothness of the contact currents and of the potential energy from one grid to the next can be used to estimate the quality of the solution. If the contact current values or the potential energy change more than a defined limit from one grid to the next a new iteration cycle has to be started with a refined grid. In the program BAMBI these are additional stopping criteria for the simulation.

3.2.4. Deletion of lines. In the previous paragraphs only criteria for the *insertion* of points have been discussed. Particularly in time dependent calculations the regions of interest move, therefore it is necessary that the grid also moves. This means that during a simulation grid points lose their importance for the accuracy of the solution while in other regions points have to be inserted. In order to minimize the number of grid points for a desired accuracy these points have to be deleted.

It is quite difficult to derive a mathematically founded criterion for the deletion of points. Indeed it is not possible to use the 'inverse' strategies for insertion of points. Only the local discretization error gives some information where points might be removed.

The Gummel–Scharfetter discretization scheme provides linear convergence since it is a first order method. As already mentioned, it implies only linear variation of the electrostatic potential between two neighbouring grid points. Furthermore, the local discretization error is proportional to the mesh spacing. From this consideration we can establish a criterion for the deletion of lines:

If for all points on a grid line the local discretization error is less than half the defined limit for insertion of new points and if it is less than half of the middle of error of the two neighbours in a perpendicular direction of the line, this line may be removed.

Some advantage of the reduced number of points will be destroyed since by the criterion of Section 3.1.3 new points will be inserted in order to obtain smooth transitions from a coarse to a fine grid.

Even for stationary calculations savings of about 25 per cent of points could be achieved because lines which have been inserted by the analysis of the doping profile will be removed if they are not needed for a special operating point.

In Figures 6 and 7 the grids of the diode (0·5 V in backward direction) without and with removal of lines can be seen. For this operating point the nn^+-junction is not of importance, therefore, the lines introduced by the analysis of the doping profile have been removed.

For transient simulations a 'moving grid' will save CPU-time and storage requirements. In most cases a reduction of the run time could be observed using alternate insertion and deletion of lines in spite of the additional effort which has to be undertaken because of the results which are described in Section 4.3.

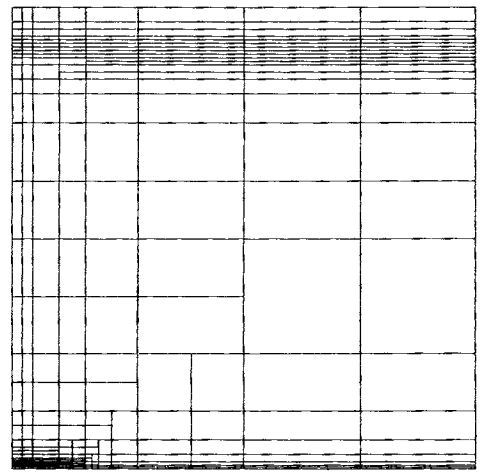

Figure 6. Grid of a diode without deletion of lines

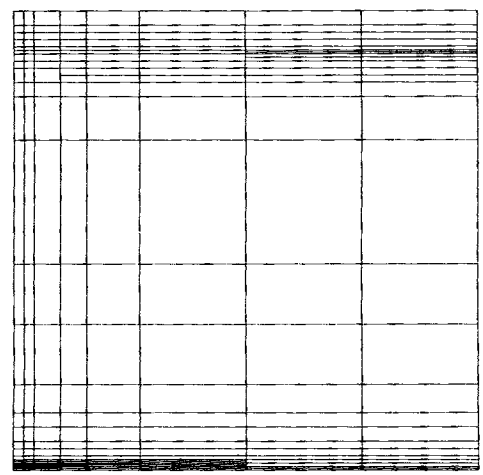

Figure 7. Grid of a diode after deletion of lines

3.2.5. Case of no convergence. The solution is either done by Gummel's algorithm (decoupled solution of the equations) or by a modified version of Newton's algorithm (simultaneous solution of the equations). In both cases it may occur that the iteration cycle does not converge.

From our knowledge it is not possible to give a generally applicable criterion for grid refinement for this case. If the residual norm of the solution is near (e.g. two orders of magnitude) to the desired accuracy and further improvement of the solution is not possible because of numerical reasons (e.g. weak coupling, high net generation, roundoff errors) all the above mentioned criteria (Sections 3.2.1–3.2.4) might be used. But if the residual norm of the solution is far away from the desired accuracy it becomes much more difficult. A commonly used method is an analysis of the residual vector introducing new points in regions with large residuals. This sometimes enables convergence, but often fails.

We have implemented another strategy. On the existing space grid we solve a set of partial differential equations:

$$\operatorname{div} \operatorname{grad} w_i = 0$$

for each contact i with the following boundary conditions:

$w_i = 1$ at the ith contact;
$w_i = 0$ at all other contacts;
$\partial w_i / \partial n = 0$ at all other boundaries.
∂n denotes the derivation with respect to the outer normal vector.

The sum of all so called *weights* must be 1 at each point of the grid; this means that the difference between 1 and the sum of the weights directly indicates the numerical error at each point. This error indicates that in regions with large differences new points have to be inserted. In Figures 8(a), (b) the weights for the diode can be seen.

It should be mentioned that the local discretization error of the solution of these Laplace equations can also be taken for an estimation of the quality of the space grid, leading to almost the same results. In our simulations it turned out that this method based on the solution of the Laplace equations is superior compared with the strategy of analysing the residual vector.

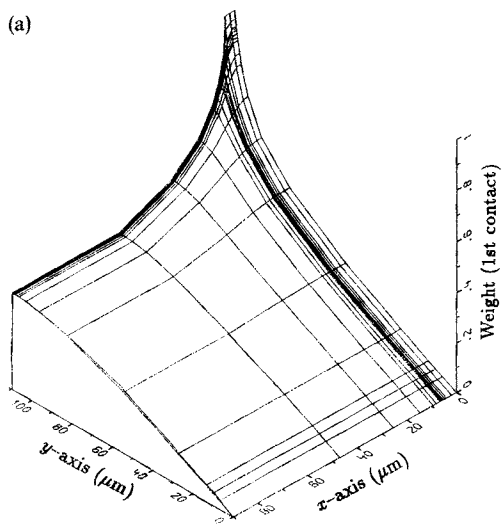

Figure 8(a). Weight function for a diode (1st contact)

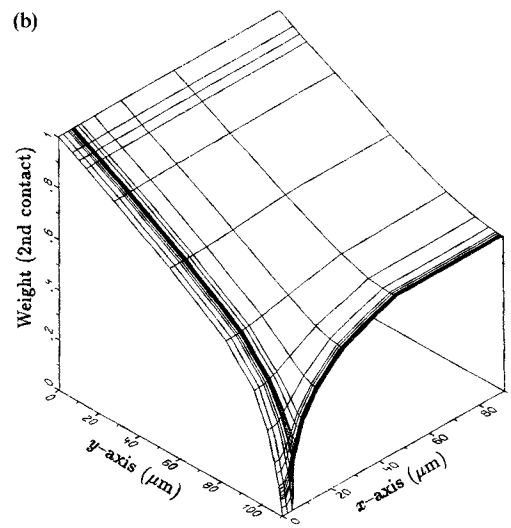

Figure 8(b). Weight function for a diode (2nd contact)

4. SELF-ADAPTIVE TIME GRIDS

In Section 3 we have discussed self-adaptive space grids. All considerations are also applicable to the space grid in transient simulations. In this section we discuss the automatic control of the time grid.

Since the Gummel–Scharfetter space discretization scheme provides linear convergence on a non-uniform space grid a time discretization scheme of order one is sufficient to guarantee consistency. In BAMBI we have implemented the backward Euler method, because it is absolutely stable. This means that no restrictions on the time step are required.

It is more difficult to derive criteria for a self-adaptive time grid than for space grids. This means that all considerations are based on heuristics and on physical properties.

4.1. Initial time step

The calculation of an initial time step is based on a physical consideration. Since the effects by the dielectric relaxation time are not analysed, we choose $\frac{1}{100}$ of the maximum life time of the carriers as initial time step dt_0. This value has been established by numerical experiments. However, if there is no change in the boundary conditions for a longer time than dt_0 the greatest possible time step is chosen while the boundary conditions are kept constant.

Since it is impossible to give a generally applicable strategy for the automatic computation of an initial time step the user may define it by himself in BAMBI.

4.2. Automatic step width control

Automatic time step width control is almost equivalent to estimating the time and the space discretization error simultaneously. We have dealt with the space discretization error in Section 3.2.1; here we will discuss the estimation of the time discretization error.

4.2.1. Time discretization error. Once more (as is done for space discretization) an estimation of the time discretization error is done by analysing the transient behaviour of the space charge. The change of the space charge densities from one time step to the next indicates the intensity of the time dependent process. It can be written as

$$\Delta\rho = \|\rho_{t+1} - \rho_t\| = \|p_{t+1} - p_t - (n_{t+1} - n_t)\| \tag{11}$$

where the index t denotes the number of the time step. Equation (11) can be computed very easily and it has proved to provide a good estimation of the time dependent computational error.[12]

4.2.2. Step width algorithm. We have implemented a sophisticated time step algorithm in BAMBI. The computation of the initial time step has been discussed in Section 4.1, the calculation of the time step during the simulation is done as follows:

Take the old time step dt_{old}. If $\Delta\rho$ is less than a given limit (e.g. 10 per cent) and there is no change in the boundary conditions take $dt_{new} = 4\,dt_{old}$.

Take the old time step dt_{old}. If $\Delta\rho$ is less than a given limit (e.g. 10 per cent) and there is a moderate change in the boundary conditions take $dt_{new} = 2\,dt_{old}$.

Take the old time step dt_{old}. If $\Delta\rho$ is greater than a given limit (e.g. 10 per cent) and/or there is a significant change in the boundary conditions take $dt_{new} = s\,dt_{old}$, where $0{\cdot}1 \leqslant s \leqslant 1$ depends on the change in the boundary conditions.

Furthermore, $\Delta\rho$ is extrapolated from two previous time steps. If 'overshoot' would occur the proposed time step is cut, thus guaranteeing that no important information will be lost.

To analyse the changes of the boundary conditions each proposed new time step is divided into 10 subintervals to check for the applied voltages and currents. Thereby also discontinuities can be resolved in a proper way. In order to reduce the number of iterations in the next Newton cycle the solution for the next time step dt_{new} is extrapolated from the two previously calculated solutions.

However, it may happen that the Newton cycle does not converge. Then the proposed time step is divided by 4 and computed again. If this fails a grid refinement is performed.

It should be mentioned that for transient calculations the space discretization error is not controlled after every time step because this might lead to an exploding number of points. Instead we check the time discretization error after every third time step to guarantee the accuracy of the space grid.

4.3. Interaction of time and space grids

The interaction of the time space grid is of major importance in transient simulations and it may cause numerical trouble if it is not handled properly.

If a space grid refinement has been performed the solution for the new points is interpolated from the old ones. The last time step is calculated again handling the old points as Dirichlet points. Intuitively it is obvious that this method introduces an error. Moreover, new points are inserted in regions with big changes in the solution, thus a smooth fitting of the new values will result in a bad solution which, however, is assumed to be correct in the next Euler step.

The contact currents are very sensitive to small changes in the solution, therefore they provide a good indicator for this interpolation error, as can be seen in Figure 9 (arrow) (turn-on of a MOS transistor).[12] In order to reduce this error two quasi-zero time steps are introduced. These two

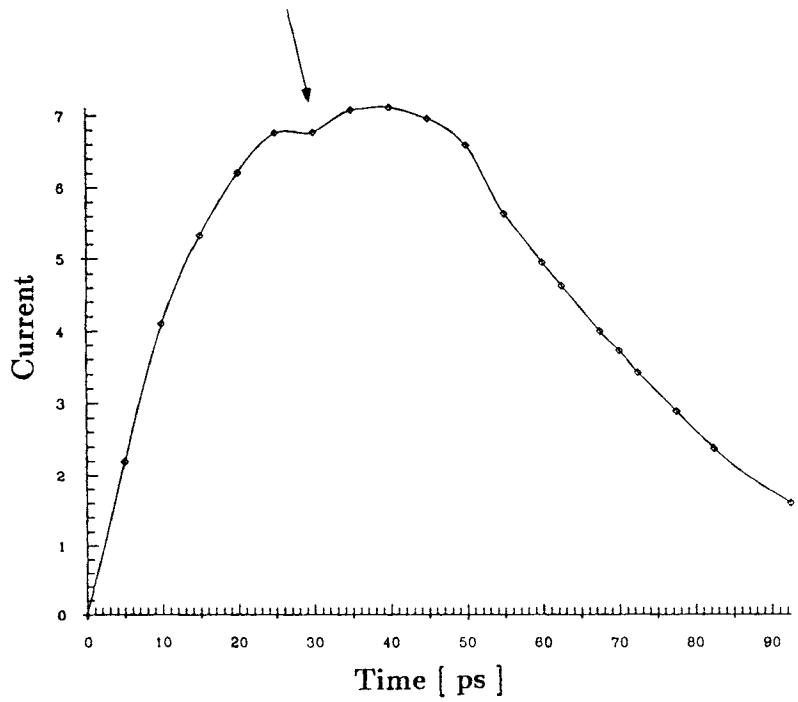

Figure 9. Contact currents without quasi-zero time steps

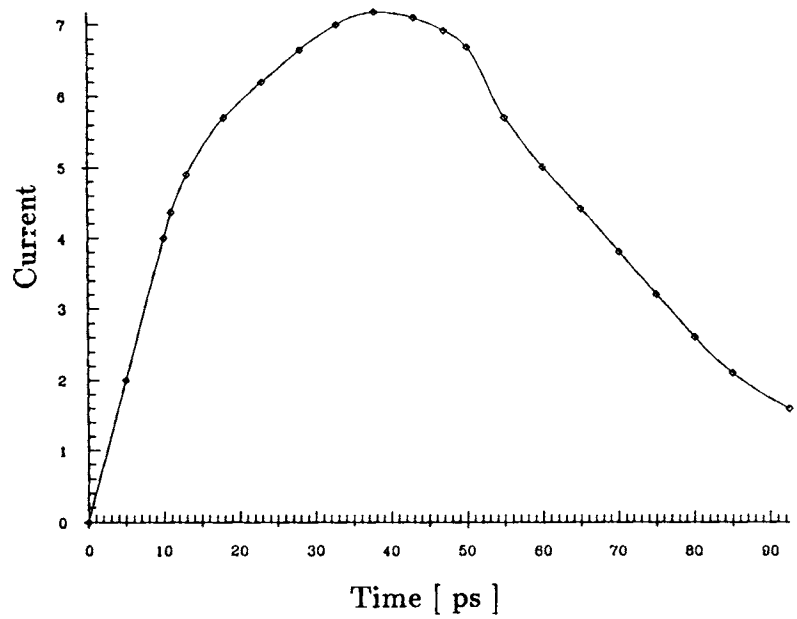

Figure 10. Contact currents with two quasi-zero time steps

steps cause wrong and oscillating results but they allow a relaxation of the space charge, thus leading to a significant improvement in the smoothness of the terminal currents Figure 10. It should be mentioned that the number of quasi-zero time steps depends on the order m of the time discretization scheme. From theory m intermediate time steps are sufficient, but it has turned out in numerical experiments that $m + 1$ steps give even better results.

5. CONCLUSION

We have presented criteria for fully self-adaptive two-dimensional space grid design. Besides the equidistribution of the local discretization error the degree of coupling of the equations is controlled and physical properties like net generation are taken into account. The automatically computed initial time step is derived from the life times of the carriers. The time step algorithm controls the applied boundary conditions (contact voltages, contact currents) and the time discretization error which is proportional to the change in the space charge. To reduce the influence of changes in the space grid on the computational error in transient simulations quasi-zero time steps provide a relaxation of the space charge after insertion or deletion of points.

By the presented methods for fully self-adaptive space and time grid control nearly optimal meshes for all operating points of a device can be achieved, thus minimizing the number of unknowns for a desired accuracy. Furthermore, an estimation of the distribution of the computational error can be given.

ACKNOWLEDGEMENTS

This work was supported by the Siemens AG Research Laboratories at Munich, Germany, by the Siemens AG, Villach, Austria, by Digital Equipment Corp. at Hudson, U.S.A., and by the 'Fonds

zur Förderung der wissenschaftlichen Forschung', project P7485-PHY. The authors are indebted to Prof. H. Pötzl for many helpful discussions.

APPENDIX

Geometries and doping profiles of examples

(*a*) *Resistor.* This example is a simple potential problem with a constant doping. It is taken from the thesis by Rank.[9]

In Figure 11 the geometry can be seen; the doping profile is assumed to be constant.

(b) p^+nn^+-diode. This two-dimensional diode is taken from the thesis by Franz.[13]

In Figure 5 the geometry can be seen; in Figure 12 the doping profile is shown.

(c) DMOS transistor. This lateral DMOS transistor has been analysed previously.[14–16]

In Figure 13 the geometry can be seen; in Figure 14 the doping profile is shown.

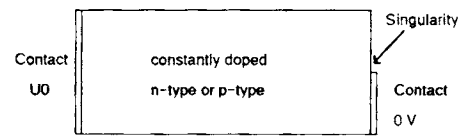

Figure 11. Geometry of the resistor

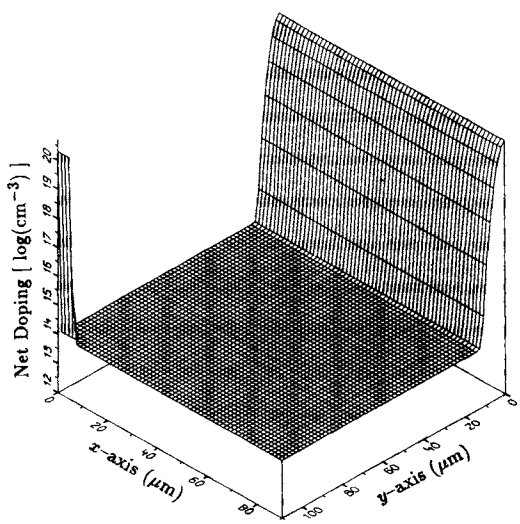

Figure 12. Doping profile of the diode

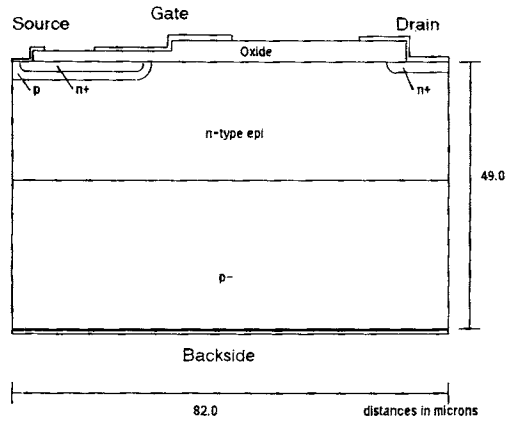

Figure 13. Geometry of the DMOS transistor

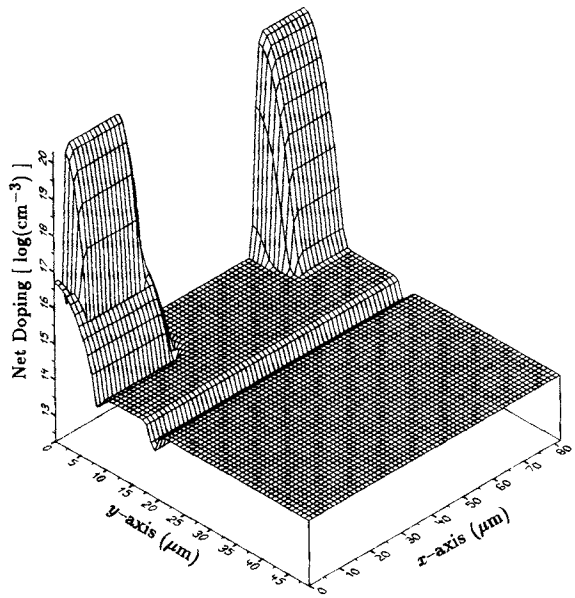

Figure 14. Doping profile of the DMOS transistor

REFERENCES

1. G. Nanz, W. Kausel and S. Selberherr, 'Automatic grid control in device simulation', in S. Sengupta *et al.* (eds.), *Numerical Grid Generation in Computational Fluid Dynamics '88*, Pineridge Press, Swansea, U.K., 1988.
2. A. F. Franz and G. A. Franz, 'BAMBI—A design model for power MOSFETs', *IEEE Trans. Comp. Aided Design Integrated Circuits Syst.*, **CAD-4**, 177–189 (1985).
3. A. F. Franz, G. A. Franz, W. Kausel, P. Dickinger and G. Nanz, *BAMBI 2.1 User's Guide*, Technical University Vienna, 1989.
4. S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, Wien, New York, 1984.

5. C. Fischer, 'Gittererzeugung in der zweidimensionalen Halbleiter-Bauelemente-Simulation', *Master Thesis*, Technical University Vienna, 1989.
6. G. Nanz and P. Dickinger, 'Selbst-adaptive Finite-Differenzen-Gitter in der Halbleiterbauelementesimulation', presented at *GAMM 90*, Hannover, Apr. 1990.
7. W. Jüngling, G. Hobler, S. Selberherr and H. Pötzl, 'Adaptive grids in space and time for process and device simulators', in J. Häuser and C. Taylor (eds.), *Numerical Grid Generation in Computational Fluid Dynamics*, Pineridge Press, Swansea, U.K., 1986.
8. J. Faricelli, Private communication, 1989.
9. E. Rank, '*A-posteriori* Fehlerabschätzungen und adaptive Netzverfeinerung für Finite-Element- und Randintegral-element-Methoden', *Mitteilungen aus dem Institut für Bauingenieurwesen I*, Technical University Munich, 1985.
10. P. A. Markowich, C. A. Ringhofer, S. Selberherr and M. Lentini, 'A singular perturbation approach for the analysis of the fundamental semiconductor equations', *IEEE Trans. Electron Devices*, **ED-30**, 1165–1180 (1983).
11. H. A. Dwyer, M. D. Smooke and R. J. Kee, 'Adaptive gridding for finite difference solutions to heat and mass transfer problems', in J. F. Thompson (ed.), *Numerical Grid Generation*, North-Holland, Amsterdam, 1982.
12. W. Kausel, G. Nanz, S. Selberherr and H. Pötzl, 'BAMBI—A transient two-dimensional device simulator using implicit backward Euler's method and a totally self adaptive grid', *NUPAD II Workshop*, San Diego, CA, Technical Digest 105/106, 1988.
13. G. Franz, 'Numerische Simulation von bipolaren Leistungshalbleiterbauelementen', *Thesis*, Technical University Vienna, 1984.
14. G. Nanz, P. Dickinger, W. Kausel and S. Selberherr, 'Punch-through in resurf devices', *Proc. Int. AMSE-Conf. on Modelling & Simulation*, Karlsruhe, 1987, pp. 63–70.
15. G. Nanz, P. Dickinger, W. Kausel and S. Selberherr, 'ON-resistance in the ALDMOST', *Journal de Physique, Colloque C4, Suppl. No. 9 (ESSDERC 1988)*, Les Ulis CEDEX (France): Les Editions de Physique, 629–632 (1988).
16. G. Nanz, P. Dickinger, W. Kausel and S. Selberherr, 'Avalanche breakdown in the ALDMOST', *Proc. Int. Conf. on Simulation of Semiconductor Devices and Processes Vol. 3*, Bologna, 1988, pp. 175–181.