REFERENCES

[1] J. H. Albers, P. Roitman, and C. L. Wilson, "Verification of models for fabrication of shallow arsenic junctions," to be published.

[2] R. E. Bank and A. H. Sherman, "PLTMG users' guide," Center for Numerical Analysis, Univ. of Texas, Austin, Rep. CNA 152, Sept. 1979.

[3] R. E. Bank, "PLTMG users' guide, June, 1981 version," Dep. Math., Univ. of California, San Diego, Tech. Rep., Aug. 1982.

[4] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, pp. 345–355, 1978.

[5] J. A. Cooper, Jr. and D. F. Nelson, "Measurement of high-field drift velocity of electrons in inversion layers on silicon," *IEEE Electron Device Lett.*, vol. EDL-2, pp. 171–173, 1981.

[6] S. C. Eisenstat, M. C. Gursky, M. A. Schultz, and A. H. Sherman, "Yale sparse matrix package I. The symmetric codes," Dep. Comp. Sci., Yale University, New Haven, CT, Res. Rep. 112.

[7] W. Fichtner, E. N. Fuls, R. L. Johnston, T. T. Sheng, and R. K. Watts, "Experimental and theoretical characterization of submicron MOSFET's," in *Proc. 1980 Int. Elec. Dev. Meeting*, pp. 24–27, Dec. 1980.

[8] R. Fair, "Concentration profiles of diffused dopants in silicon," in *Applied Solid State Science Suppl. 2B*, Dawon Kahng, Ed. New York: Academic Press, 1981, pp. 1–108.

[9] P. A. Fox, A. D. Hall, and N. L. Schryer, "The PORT mathematical subroutine library," *ACM Trans. Mathemat. Software*, vol. 4, pp. 104–126, 1978.

[10] R. C. Frye and J. H. Leamy, "Direct observation of the gate oxide electric field distribution in silicon MOSFET's," *IEEE Electron Device Lett.*, vol. EDL-3, pp. 1–3, 1982.

[11] D. P. Kennedy and R. R. O'Brien, "Analysis of the impurity atom distribution near the diffusion mask for a planar p–n junction," *IBM J. Res. Develop.*, vol. 9, pp. 179–186, 1965.

[12] B. W. Kernighhan, "RATFOR—A preprocessor for a rational Fortran," *Software — Practice and Experience*, vol. 5, 1975.

[13] B. G. Ryder, "The PFORT verifier," *Software — Practice and Experience*, vol. 4, pp. 359–377, 1974.

[14] P. Roitman, D. R. Myers, J. H. Albers, J. R. Ehrstein, and J. Lowney, "Direct observation of lateral redistribution profiles of shallow ion implants," *IEEE Trans. Electron. Devices*, vol. ED-28, p. 1240, 1981.

[15] S. A. Schwarz and S. E. Russek, "Simple physical models for bulk and surface electron velocities in silicon," presented at the Device Res. Conf., June 21–23, 1982, Paper VA-1.

[16] A. H. Sherman, "Algorithms for sparse Gaussian elimination with partial pivoting," *ACM Trans. Mathemat. Software*, vol. 4, pp. 330–338, 1978.

[17] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[18] P. Swarztrauber and R. Sweet, "Efficient FORTRAN subprograms for the solution of elliptic partial differential equations," NCAR Tech. Note NCAR-TN/IA-109, 1975.

[19] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley 1981, pp. 16–57.

[20] K. K. Thornber, "Current equations for velocity overshoot," *IEEE Electron Device Lett.*, vol. EDL-3, pp. 69–71, 1982.

[21] D. D. Warner and C. L. Wilson, "Two-dimensional concentration dependent diffusion," *Bell Syst. Tech. J.*, vol. 59, pp. 1–41, 1980.

[22] C. L. Wilson and J. L. Blue, "Two-dimensional finite element charge-sheet model of a short-channel MOS transistor," *Solid-State Electron.*, vol. 25, pp. 461–477, 1982.

[23] P. S. Winokur and H. E. Boesch, Jr., "Interface-state generation in radiation-hard oxides," *IEEE Trans. Nucl. Sci.*, vol. NS-27, p. 1647, 1980.

Finite Boxes—A Generalization of the Finite-Difference Method Suitable for Semiconductor Device Simulation

ANDREA F. FRANZ, MEMBER, IEEE, GERHARD A. FRANZ, MEMBER, IEEE, SIEGFRIED SELBERHERR, MEMBER, IEEE, CHRISTIAN RINGHOFER, AND PETER MARKOWICH

Abstract —**A two-dimensional numerical device-simulation system is presented. A novel discretization scheme, called "finite boxes," allows an optimal grid-point allocation and can be applied to nonrectangular devices. The grid is generated automatically according to the specified device geometry. It is adapted automatically during the solution process by equidistributing a weight function which describes the local discretization error. A modified Newton method is used for solving the discretized nonlinear system. To achieve high flexibility the physical parameters can be defined by user-supplied models. This approach requires numerical calculation of parts of the coefficients of the Jacobian. Supplementary algorithms speed up convergence and inhibit the commonly known Newton overshoot. The advantages and computer resource savings of the new method are described by the simulation of a 100-V diode. We also present results for thyristor and GaAs MESFET simulations.**

Manuscript received November 2, 1982; revised March 29, 1983. This work was sponsored in part by the Fonds zur Förderung der wissenschaftlichen Forschung under Project S22/11, in part by the Research Laboratory of Siemens AG, Munich, Germany, and in part by the United States Army under Contract DAAG-29-80-C-0041.

A. F. Franz, G. A. Franz, and S. Selberherr are with Abteilung Für Physikalische Elektronik, Institute für Allgemeine Elektrotechnik und Elektronik, Technical University of Vienna, Gusshausstrasse 27, A-1040, Vienna, Austria.

C. Ringhofer is with the Mathematics Research Center, University of Wisconsin, Madison, WI 53706.

P. Markowich is with the Institut für Angewandte Mathematik, Technical University of Vienna, Gusshausstrasse 27, A-1040, Vienna, Austria.

I. INTRODUCTION

TWO-DIMENSIONAL finite-difference simulation of complex semiconductor structures, in view of an acceptable amount of computer resources to be used,

0018-9383/83/0900-1070$01.00 ©1983 IEEE

requires an approach that improves the classical methods. Regions where a coarse grid suffices to resolve the solution of the elliptic system of semiconductor equations generally alternate with domains where a fine mesh is necessary. We developed a generalization of the classical rectangular-grid finite-difference method, which we call the concept of "finite boxes," to overcome the drawbacks of the former.

In our simulation program the initial grid is generated automatically according to the specified device geometry, and it is adapted iteratively during the solution process by equidistribution of the local discretization error of Poisson's equation. This approach is applicable to nonrectangular devices and highly complex structures, like multilayer or multiple p–n junction devices. Details of the discretization procedure are outlined in Section II.

In Section III the modified Newton method used for solving the discrete nonlinear system is explained. Supplementary algorithms which speed up convergence and prohibit Newton overshoot are implemented. Emphasis is put on the calculation of the entries of the Jacobian. Since the code should not be restricted to a particular kind of semiconductor, it is necessary to change physical parameters in an easy way without alteration of the code. The constants that characterize the semiconductor (like the permitivity ϵ) are part of the input deck. The doping profile, the generation/recombination term, as well as the carrier mobilities, are defined as external functions. Default models are offered, but the user may replace them with his own formulas. This flexibility induces a diversified approach for the calculation of the Jacobian matrix. Parts of the entries are calculated by evaluating analytically defined formulas. The remaining terms which depend on external functions are generated by numerical differentiation. The resulting advantages and disadvantages will be explained in Section III.

The derived linear system of equations is solved by Gaussian elimination. To minimize fill-in, we developed a reordering based on the "minimum degree" algorithm of the grid points which form the "finite boxes" structure.

In the final section we present typical meshes for three types of devices, namely a high-voltage diode, a thyristor, and a GaAs MESFET. The whole iteration process of the diode is explained with respect to advantages and savings of the new concept. The equidistribution of the value of the error functional is shown. We compare the influence of various ways of modelling the physical parameters of the devices.

II. Discretization

A. The Finite Boxes Concept

In the classical concept of finite differences, an elliptic PDE system is discretized on a rectangular grid throughout the device. The grid spacings depend on the estimate of the truncation error of the discretization scheme along the whole grid lines. Therefore the simulation of complex structures with rapidly varying solutions employs a large number of grid points such that reasonable storage requirements are often exceeded. A program system suitable for an arbitrary device, which is based on the classical finite difference approach, is not executable on most computer systems.

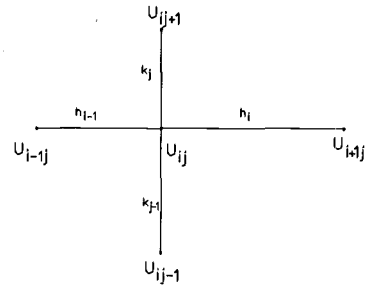


Fig. 1. Classical 5-point discretization.

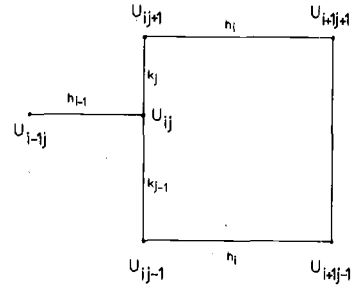


Fig. 2. Finite boxes discretization.

To overcome these difficulties, we developed a new concept for the discretization of the fundamental semiconductor equations which may be interpreted as a consequent generalization of the "terminating-line" approach introduced by Adler [1]. The usual rectangular pattern of grid points is dropped. The mesh is built up by rectangular cells, the so-called "finite boxes." In the case of nonrectangular-shaped devices, triangles are attached to the cells along the outer boundaries (see Subsection II-D). A rectangular box usually links the 4 corner grid points. A fifth point can be inserted at one edge to allow the connection of boxes of different sizes. In this way the dimension of the boxes may vary over decades throughout the simulation area in accordance to the solution of the PDE system and the truncation error estimates. We demonstrate our approach by the discretization of the two-dimensional Laplace operator. Its classical discretization [2] (Fig. 1) leads to the 5-point formula (1).

$$\begin{aligned}\Delta U_{ij} = {} & U_{i+1j}\frac{2}{h_i(h_i+h_{i-1})} + U_{ij+1}\frac{2}{k_j(k_j+k_{j-1})} \\ & + U_{i-1j}\frac{2}{h_{i-1}(h_i+h_{i-1})} + U_{ij-1}\frac{2}{k_{j-1}(k_j+k_{j-1})} \\ & - U_{ij}\left(\frac{2}{h_ih_{i-1}} + \frac{2}{k_jk_{j-1}}\right) \end{aligned} \tag{1}$$

with a local truncation error of at least $0(h+k)$ with $h = \max h_i$ and $k = \max k_j$ assuming that the third derivatives of U are bounded.

Fig. 2 shows a finite box consisting of 5 grid points. The discretization of ΔU_{ij} has to include the function values at

points $i+1j+1$ and $i+1j-1$ since point $i+1j$ is missing. Simple linear interpolation of the missing function value in (1) gives

$$\begin{aligned}\Delta U_{ij} = \frac{2}{h_i(h_i+h_{i-1})}\Bigg[&U_{i+1j+1}\frac{k_{j-1}}{k_j+k_{j-1}}\\ &+ U_{i+1j-1}\frac{k_j}{k_j+k_{j-1}}\Bigg]\\ &+ U_{ij+1}\quad\frac{2}{k_j(k_j+k_{j-1})}\\ &+ U_{i-1j}\quad\frac{2}{h_{i-1}(h_i+h_{i-1})}\\ &+ U_{ij-1}\quad\frac{2}{k_{j-1}(k_j+k_{j-1})}\\ &- U_{ij}\quad\left(\frac{2}{h_ih_{i-1}}+\frac{2}{k_jk_{j-1}}\right).\end{aligned}\tag{2}$$

The discretization (2) will increase the local truncation error proportional to U_{yy} and is therefore not applicable for nonvanishing second derivatives. To obtain a difference equation of the same order of consistency as (1), the first derivatives have to be used for interpolation giving (3)

$$\begin{aligned}\Delta U_{ij} = &U_{i+1j+1}\frac{2k_{j-1}}{h_i(h_i+h_{i-1})(k_j+k_{j-1})}\\ &+ U_{i+1j-1}\frac{2k_j}{h_i(h_i+h_{i-1})(k_j+k_{j-1})}\\ &+ U_{ij+1}\left[\frac{2}{k_j(k_j+k_{j-1})}\right.\\ &\left.-\quad\frac{2k_{j-1}}{h_i(h_i+h_{i-1})(k_j+k_{j-1})}\right]\\ &+ U_{i-1j}\quad\frac{2}{h_{i-1}(h_i+h_{i-1})}\\ &+ U_{ij-1}\left[\frac{2}{k_{j-1}(k_j+k_{j-1})}\right.\\ &\left.-\quad\frac{2k_j}{h_i(h_i+h_{i-1})(k_j+k_{j-1})}\right]\\ &- U_{ij}\quad\left(\frac{2}{h_{i-1}(h_i+h_{i-1})}+\frac{2}{k_jk_{j-1}}\right).\end{aligned}\tag{3}$$

This algorithm is certainly applicable in any direction. The PDE system can be discretized for the finite-boxes structure using (1) and (3). Poisson's equation can easily be obtained by replacing U with the electrical potential. In the case of terminating lines, the electric field is interpolated in (3). For the continuity equations, we use the well-known Scharfetter–Gummel discretization [3] presuming constant current densities between grid points. If one neighbor is misssing, we interpolate the related current densities. Since they are directly related to the first derivatives of the quasifermipotential, this method is consistent with that described previously.

B. Regularity Conditions

The occurrence of a terminating line (Fig. 2) depends on the magnitude of the derivatives of the solution of the partial differential equation (PDE) system. Although the discretization error of (3) is of the same order as that of (1), an additional term is added. For terminating lines in positive or negative x direction, this term reads

$$\frac{k_jk_{j-1}}{h_i+h_{i-1}}\frac{\partial^3U}{\partial x\partial y^2}.\tag{4a}$$

The corresponding term in positive or negative y direction is

$$\frac{h_ih_{i-1}}{k_j+k_{j-1}}\frac{\partial^3U}{\partial x^2\partial y}.\tag{4b}$$

U represents the electrical potential and the quasi-fermi potentials, respectively. Only if this additional error is reasonably small may one permit a line to terminate without a significant loss of accuracy of the solution. Therefore terminating lines only occur in regions of slow variation of the solution.

To bound the local discretization error, it is also reasonable to limit the length to width ratio h_i/k_j of the boxes using a moderate constant c

$$1/c \leqslant h_i/k_j \leqslant c.\tag{5}$$

$c=10$ is used successfully in numerical experiences. Therefore the ratio of all distances between a grid point and its neighbors does not exceed 100.

A second condition concerns the structure of the "finite boxes." As mentioned before, the discretization scheme for a terminating line is based on interpolation between the 4 points $ij+1$, $i+1j+1$, $ij-1$, and $i+1j-1$. The entries of the points $ij+1$ (6a) and $ij-1$ (6b), respectively, may vanish or change sign if (7).

$$\frac{2}{k_j(k_j+k_{j-1})}-\frac{2k_{j-1}}{h_i(h_i+h_{i-1})(k_j+k_{j-1})}\tag{6a}$$

$$\frac{2}{k_{j-1}(k_j+k_{j-1})}-\frac{2k_j}{h_i(h_i+h_{i-1})(k_j+k_{j-1})}\tag{6b}$$

$$h_i(h_i+h_{i-1}) < k_jk_{j-1}.\tag{7}$$

This may prevent the linearized system of equations to converge. Positive sign of (6a) and (6b) is guaranteed if (8) holds.

$$h_i > k_j + k_{j-1}.\tag{8}$$

For similar reasons we allow only one line to terminate in a box. Therefore the mesh structure in Fig. 3 with a box containing 6 points is illegal. The neighbor $i+1j+1$ of the point ij is missing and point $ij+1$ has no neighbor $i+1j$. The current densities and the electric field along the lines $j+2$ and $j-1$ must get interpolated for both equations ij

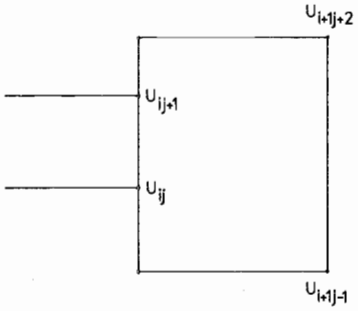


Fig. 3. Illegal mesh structure.

and $ij+1$. This leads to multiple entries of the points $i+1j+2$ and $i+1j-1$ in the system of equations and impairs the numerical stability of the solution methods drastically.

C. Finite Boxes Versus Classical Mesh

Applying formula (3) together with the regularity conditions previously outlined, one can construct a grid point allocation like that demonstrated in Fig. 12 for the discretization of the PDE system. Since this structure is intrinsically different from a classical grid, it is better looked upon as a collection of the so called "finite boxes." Compared to the former, an additional overhead organization of these boxes is necessary. Since no uniform ordering is possible, we use a graph representation. To recognize the connection between two grid points uniquely, the points are stored in a "neighbor list." For each point the list contains 4 pointers to the storage addresses of its neighbors. An ordinary mesh point has 4 neighbors, a terminating line point only 3. Therefore the value zero in the direction of the missing link marks this situation. Together with the coordinate information, this representation of the boxes makes it possible to locate all grid points necessary for the discretization of the PDE operator (c.f. (3) and Fig. 2).

Using this "finite-boxes" strategy, the grid can be refined and coarsened locally. We adapt the mesh spacings to the variation of the solution such that a given global error tolerance is achieved [4]. As demonstrated by Figs. 4 and 5, the number of grid points is drastically reduced compared to classical finite differences since the "finite boxes" strategy allows to allocate only few grid points in regions of weak variation of the solution and to place many points in regions of fast variation (e.g., layer regions occurring at p–n junctions and Schottky contacts).

Fig. 4 shows a grid generated by the "finite boxes" approach for a power thyristor on the blocking characteristic with about 900 points. In Fig. 5 all terminating lines are extended through the whole device giving a rectangular grid (necessary for the classical finite-difference discretization). The number of grid points is increased to about 1700. For this case, our approach requires less than 55 percent of the grid points compared to the classical method. Generally, the saving depends on the specified geometry, the applied voltages, and the resulting current flux; it can exceed 60 percent.

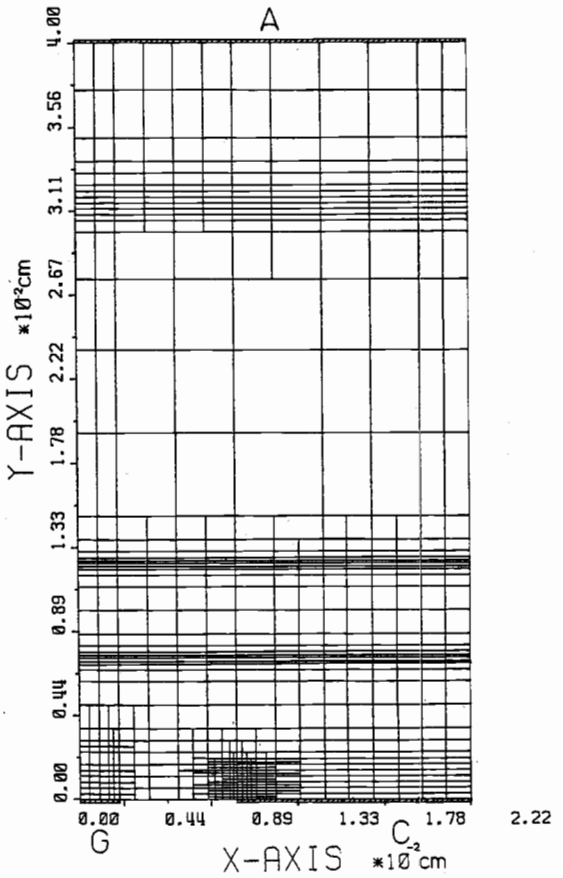


Fig. 4. Finite boxes grid.

D. Grid Development Process and Local Mesh Refinement

The mesh must be adapted automatically during the solution process in order to achieve a suitable final accuracy of the solution. We start with an initial grid which is constructed with respect to the specified device geometry and the doping profile and which matches the regularity conditions for the "finite boxes." The edges of the device are defined by polygons. Each line is identified by a unique number and its specified boundary condition. If this condition changes along a line (e.g., contact to noncontact area), the edge must be split into two parts with different properties and characteristics, respectively. The rectangle which circumscribes the device geometry is divided by a cross parallel to the coordinate axis. The parts of the cross located inside the specified device form the first grid lines. We check if they represent a regular grid. This will hold only for a convex device domain. Otherwise the already-generated boxes will be divided, and another check is performed. This procedure continues until each line has at least one intersection with another grid line. All specified edges must be represented in that grid. There are two possible connections of a boundary point to interior boxes: 1) edges with a prescribed normal vector into the device parallel to a coordinate axis have only one neighbor, and 2) all the other edges must be linked to two points. Using this method all necessary information about the device geometry is stored in the initial graph. Afterwards, the

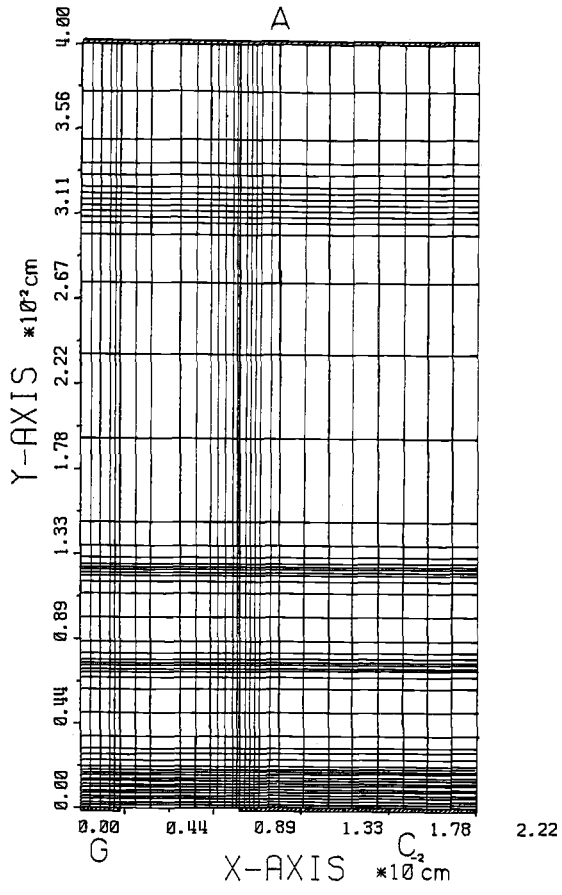


Fig. 5. Classical regular grid.

regularity conditions for the "finite boxes" outlined in Section II-B are checked for each terminating line. In this way a "preinitial" grid is constructed.

Our concept of "finite boxes" is also capable of processing nonrectangular device geometries when boundaries are approximated by polygons. By attaching triangles to the rectangular boxes as shown in Fig. 6, bevelled edges can be discretized. The boundary conditions for the derivatives perpendicular to the bevelled edge involving two neighbors $i+1\,j$ and $ij+1$ are given by

$$A(x,y)\cdot U + B(x,y)\cdot\frac{\partial U}{\partial n} = C(x,y)$$

$$\frac{\partial U}{\partial n} = \frac{\partial U}{\partial x}\cos\alpha + \frac{\partial U}{\partial y}\sin\alpha \quad (9)$$

where α is defined by the angle between the perpendicular to the edge and the positive x-coordinate axis. Equation (9) is a Dirichlet condition if $A=1$ and $B=0$ and a Neumann condition if $A=0$ and $B=1$.

The outlined strategy allows us to define and simulate even rather small structures within a large device. To get the initial grid, the previously defined "preinitial grid" has to be refined corresponding to the specified doping profile and the interior interfaces (e.g., Si-SiO_2). The criterion for this is the ratio of the active impurity concentration at two neighboring points. This ratio must not exceed a prescribed value r if the net concentration is higher than the intrinsic density n_i. We have used successfully a value of $r=10^3$.

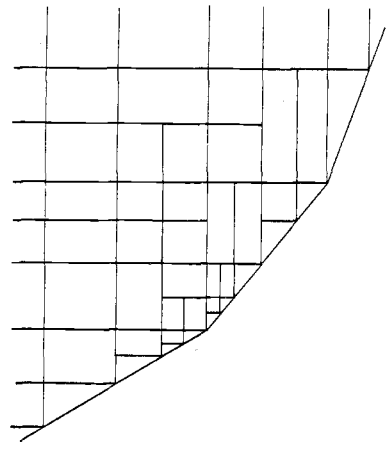


Fig. 6. Bevelled boundary.

Investigations show that the absolute value of the impurity concentration need not be taken into account. Near Schottky contacts and Si–SiO_2 interfaces small grid spacings are necessary. So the described ratio is set to a constant value greater than r for automatic refinement. To stop insertion of grid points the distance between two neighbors is limited to 10^{-6} cm (valid only for the initial grid).

The initial guess of the solution is derived from the charge-neutral approximation. In [5] it is proven that this gives a reasonable approximation of the solution, even if the mesh is too coarse for the desired final accuracy. An experimental "proof" is given in Section IV-A. After solution of the system of equations (c.f. Section III), the local discretization error is calculated at each point. An optimal discretization can be obtained by equidistribution of this error [4]. The value of the error functional (referenced as weight function) determines whether new grid points have to be inserted or old points can be deleted without loss of accuracy. If it exceeds a given tolerance at a mesh point, we refine the grid locally by dividing the four boxes around that point (or three in the case of a terminating line). The mesh points with weight-function values far less than that limit are deleted if the following conditions hold: 1) the deletion of a point must not split the graph (e.g., the mesh) into two separated subgraphs; and 2) each remaining interior point must be linked to at least three other points (inner corners are prohibited). Regarding these conditions, only a reasonably small number of mesh points can be removed at every refinement. The function values at the new grid points are interpolated using the PDE's. Then the total system is solved again. The error estimates are recalculated and checked. This procedure terminates if no more grid points have to be inserted. A typical series of automatically generated and adapted grids are shown in Section IV.

III. Solution of the Discretized System

The linear system of equations obtained by the "finite boxes" discretization is solved simultaneously. Although decoupled methods are easier to implement and require

less overhead, they converge very slowly (at most linear). Their application is only of interest as far as computer storage saving is concerned.

At first we present the simultaneous solution algorithm. Since this approach has the disadvantage that all three equations are solved simultaneously—and therefore the storage requirements are fairly large—we also suggest a Block–Newton–SOR method outlined afterwards.

A. Simultaneous Newton Algorithm

The discretized nonlinear system is solved by a modified Newton method with A as iteration matrix

$$F(x) = 0 \tag{10}$$

$$x_{k+1} = x_k - A \cdot F_k. \tag{11}$$

As commonly known, the classical Newton method with $A = (F'(x_k))^{-1}$ tends to overshoot especially if the iteration is started with a bad initial guess. Therefore we use an iteration matrix as suggested by Meyer [6] and Bank and Rose [7].

$$A = (\lambda I + F')^{-1}. \tag{12}$$

The value of the parameter λ is directly proportional to $\|F(x_k)\|$. If the calculated solution x_{k+1} does not decrease the norm of the system of equations (10), λ has to be increased. So the matrix A is updated, and a new solution is computed. This procedure terminates if

$$\|F(x_{k+1})\|_2 < \|F(x_k)\|_2. \tag{13}$$

If condition (13) is not satisfied for a particular λ, it is necessary to recalculate the Jacobian. In order not to waste the last Newton step, we calculate a parameter δ proposed by Deuflhardt [8] such that

$$\|F(x_k + \delta(x_{k+1} - x_k))\|_2 < \|F(x_k)\|_2, \quad 0 < \delta \leqslant 1. \tag{14}$$

This leads to a minimization of the number of Gaussian eliminations which are rather time consuming. By calculating δ in a way that

$$\|F_k\|_2 - \|F_{k+1}\|_2$$

becomes a maximum, faster convergence can be achieved.

B. Block–Newton–SOR

The main disadvantage of the simultaneous Newton algorithm is the large amount of computer resources for the factorization of the iteration matrix. Therefore an alternative to that has been investigated. We implemented a Block–Newton–SOR (successive overrelaxation) method. Under the assumption that the Jacobian is definite, one can use a classical block-iteration scheme (iteration index m) (15) for the solution of the kth Newton step of system (10) with $F = (F_1, F_2, F_3)^T$.

$$\begin{vmatrix} \frac{\partial F_1}{\partial \psi} & 0 & 0 \\ \frac{\partial F_2}{\partial \psi} & \frac{\partial F_2}{\partial n} & 0 \\ \frac{\partial F_3}{\partial \psi} & \frac{\partial F_3}{\partial n} & \frac{\partial F_3}{\partial p} \end{vmatrix}^{k} \cdot \begin{vmatrix} \delta\psi^k \\ \delta n^k \\ \delta p^k \end{vmatrix}^{m+1}$$

$$= - \begin{vmatrix} F_1(\psi^k, n^k, p^k) \\ F_2(\psi^k, n^k, p^k) \\ F_3(\psi^k, n^k, p^k) \end{vmatrix} - \begin{vmatrix} 0 & \frac{\partial F_1}{\partial n} & \frac{\partial F_1}{\partial p} \\ 0 & 0 & \frac{\partial F_2}{\partial p} \\ 0 & 0 & 0 \end{vmatrix}^{k} \cdot \begin{vmatrix} \delta\psi^k \\ \delta n^k \\ \delta p^k \end{vmatrix}^{m} \tag{15}$$

where F_1 is the Poisson equation, F_2 is the continuity equation for electrons, and F_3 is the continuity equation for holes.

Since the coefficient matrix of (15) is block lower-triangular, we decouple the elimination process into three linear systems (16)–(18), which have to be solved sequentially.

$$\frac{\partial F_1^k}{\partial \psi} \cdot \delta\psi^{km+1} = - F_1(\psi^k, n^k, p^k) - \frac{\partial F_1^k}{\partial n} \cdot \delta n^{km} - \frac{\partial F_1^k}{\partial p} \cdot \delta p^{km} \tag{16}$$

$$\frac{\partial F_2^k}{\partial n} \cdot \delta n^{km+1} = - F_2(\psi^k, n^k, p^k) - \frac{\partial F_2^k}{\partial \psi} \cdot \delta\psi^{km+1} - \frac{\partial F_2^k}{\partial p} \cdot \delta p^{km} \tag{17}$$

$$\frac{\partial F_3^k}{\partial p} \cdot \delta p^{km+1} = - F_3(\psi^k, n^k, p^k) - \frac{\partial F_3^k}{\partial \psi} \cdot \delta\psi^{km+1} - \frac{\partial F_3^k}{\partial n} \cdot \delta n^{km+1}. \tag{18}$$

The cut Taylor series on the right-hand side of (16)–(18) are resubstituted and a relaxation parameter ω is introduced giving (19)–(21).

$$\frac{\partial F_1^k}{\partial \psi} \cdot \delta\psi^{km+1} = - \omega \cdot F_1(\psi^k, n^k + \delta n^{km}, p^k + \delta p^{km}) \tag{19}$$

$$\frac{\partial F_2^k}{\partial n} \cdot \delta n^{km+1} = - \omega \cdot F_2(\psi^k + \delta\psi^{km+1}, n^k, p^k + \delta p^{km}) \tag{20}$$

$$\frac{\partial F_3^k}{\partial p} \cdot \delta p^{km+1} = - \omega \cdot F_3(\psi^k + \delta\psi^{km+1}, n^k + \delta n^{km+1}, p^k). \tag{21}$$

The parameter ω accelerates the rate of convergence. It is calculated by [9]

$$\omega = \frac{2}{1+\sqrt{1-\varrho}} \tag{22}$$

where ϱ is the spectral radius of the iteration matrix ($\omega = 1$) given by

$$\varrho = \lim_{k \to \infty} \varrho_k$$
$$\varrho_k = \frac{\|x_{k+1} - x_k\|}{\|x_k - x_{k-1}\|}. \tag{23}$$

An estimate for ϱ is obtained if two consecutive ϱ_k values do not differ by more than a given tolerance ϵ. Starting with an initial value $\omega_0 = 1$, ω is updated for faster convergence whenever a sufficiently good estimate for ϱ is obtained [10].

$$\omega_{\text{new}} = \frac{2}{1+\sqrt{1-(\varrho+\omega_{\text{old}}-1)^2/(\varrho\,\omega_{\text{old}}^2)}}. \tag{24}$$

The increments $\delta\psi$, δn, and δp are only accepted if (25).

$$\|F_i^{k+1}\|_2 < \|F_i^k\|_2, \quad i = 1,2,3. \tag{25}$$

If (25) cannot be satisfied, we use the damping algorithm (14).

The sequence of the equations depends on the specified device. Poisson's equation F_1 is always solved first. The equation for the majority carriers preceeds the equation for the minority carriers. For devices with no unique majorities (e.g., diodes), the continuity equations are solved alternately. Let k be the iteration count. Then for even k we solve the sequence a), for odd k sequence b).

a)	b)
F_1	F_1
F_2 (electrons)	F_3 (holes)
F_3 (holes)	F_2 (electrons)

C. The Calculation of the Entries of the Jacobian

There are in principle two possibilities to calculate the elements of the Jacobian. At first they can be calculated analytically which allows the implementation of formulas straight forward without additional overhead. Nevertheless, there are issues which may give preference to the second method, namely the numerical calculation of the entries. If one has implemented the formulas analytically, it is impossible to investigate structurally different models for physical parameters because this changes the code of the program. Various algorithms for the numerical approximation of derivatives have been tested. We chose an automatically generated step-width control method similar to that given by Curtis and Reid [11]. The value of the functional including the parameter models varies over many decades. Therefore simple evaluation using constant step width will produce inaccurate results. Since the numerical calculation is more complex and computer time consuming, we have tried to take advantage of both methods as follows.

The application should not be restricted to special assumptions on physical parameters. Therefore the doping profile as well as the recombination/generation term and the carrier mobilities are defined as external functions which are user-supplied and can even consist of numerical data. To process these models correctly we split the coefficients of the Jacobian for the two continuity equations. One part is calculated analytically, the other containing the derivatives of the physical parameters numerically using the algorithm described earlier. The discretized and linearized Poisson's equation contains no derivatives of externally defined functions, so the corresponding entries of the Jacobian are calculated analytically.

The calculation of the Jacobi matrix F_k' is rather time consuming because of the numerical evaluation of external functions. Therefore we perform a rank one update of the Jacobian [12] whenever the convergence rate is sufficiently large (26).

$$F_{k+1}' = F_k' + \frac{\left((F(x_{k+1}) - F(x_k)) - F_k' \cdot x_k\right) \cdot x_k^T}{x_k^T \cdot x_k}. \tag{26}$$

In addition, it is possible to bypass some equations of the system if the following condition holds for the grid point i:

$$|F(x_k)^j| < \epsilon \cdot \max |F(x_k)|_\wedge$$
$$|x_{k+1}^j - x_k^j| < \epsilon \cdot \max |x_{k+1} - x_k|$$
$$j \epsilon \text{neighb}(i). \tag{27}$$

This decreases the rank of the system, and therefore increases the speed of calculation.

The linear systems (11) and (19)–(21) are solved using a sparse Gaussian elimination algorithm. The special structure of the equations (11) allows block elimination (for each point the three equations are combined), which speeds up the LU factorization.

The big problem when using Gaussian elimination is the fill-in of the upper triangular matrix U during forward elimination. Minimization of this fill-in can reduce the demand for virtual storage and the number of operations drastically. For that purpose a reordering of the grid points is necessary. We chose the "minimum-degree" ordering similar to that published by George and Liu [13]. Their algorithm however can only be applied on structurally symmetric matrices. The classical finite-difference discretization generates such a matrix. In the "finite boxes" approach the discretization scheme of regular grid points is identical to the previously mentioned 5-point formula. At terminating lines, however, 6 points are used. This leads to a structurally nonsymmetric matrix. The total storage requirement of the algorithm [13] is known before execution because of the representation of the matrix by an undirected graph. This advantage is lost when applying it to a structurally nonsymmetric matrix. A new strategy has been developed for this purpose using directed graphs [14]. We adapted the neighbor list concept (cf. Section II-C) for the bidirectional graph representation. The number of entries is variable and equals the current degree of the particular

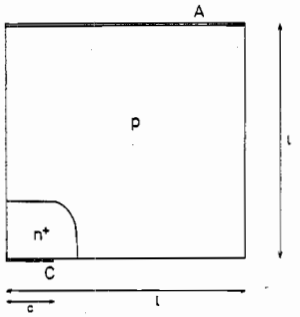


Fig. 7. Geometry diode.

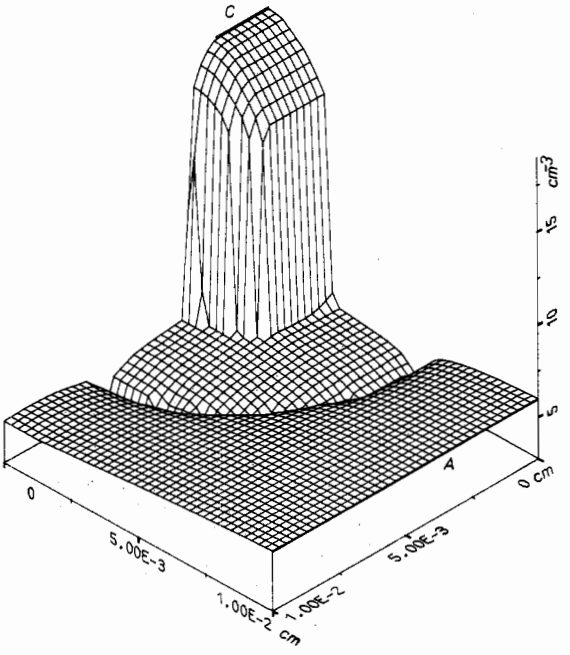


Fig. 8. Electron distribution of the diode.

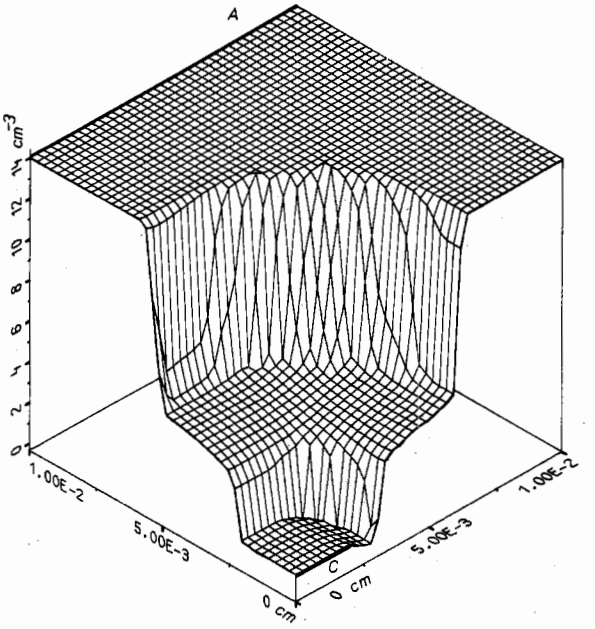


Fig. 9. Hole distribution of the diode.

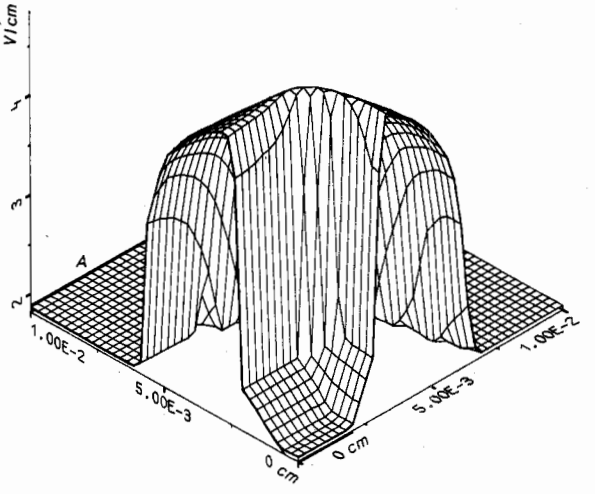


Fig. 10. Modulus of the electric field of the diode.

node. Contrary to [13], there is no unique upper boundary value for the degree of a node, and the maximum storage requirements cannot be predicted. Therefore we have to switch to out-of-core storage.

IV. Results

In this section we discuss some applications of the "finite boxes" method. The processing of high voltages is shown by the simulation of a reverse biased diode. This also gives an illustrative example of the automatic grid adaption. The results are compared to classical methods displaying the actual rate of savings in computer time. Secondly, we present solutions of the semiconductor equations for a thyristor operating in the forward mode. The typical properties of such devices are high current densities. Our investigations concentrate not only on silicon but also on GaAs devices. The simulation results of a recessed-gate GaAs MESFET conclude this section.

A. Diode

Fig. 7 shows the geometry and the location of the p–n junction of a high-voltage diode. The concentration of the donors is described by a Gaussian profile with a maximum of 10^{19} cm^{-3} and a projected range of 4 μm; the acceptor concentration is homogenous $N_A = 10^{14}$ cm^{-3}. A voltage of 100 V is applied at the cathode. Just that voltage drop and the charge neutral approach are used for the initial guess. We do not increment the contact potential from a much lower level as it has been suggested in the literature, e.g., [15], [16], thus saving computer time. The application of mathematically well-founded discretization strategies [3]–[5] guarantees global convergence. Figs. 8–10 show the calculated electron and hole densities and the electric field (logarithmically plotted), respectively. We start the calculation at a 135- point initial mesh (Fig. 11). Nine grid updates are performed. The final grid contains 916 points (Fig. 12). The number of grid points *GP* and Newton steps

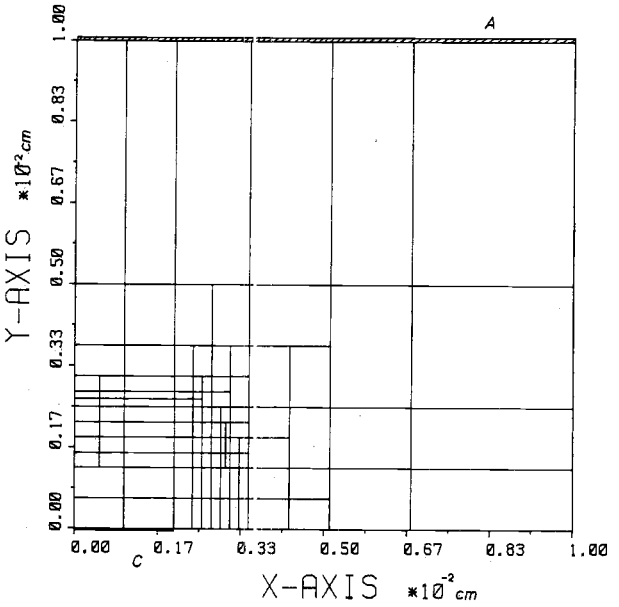


Fig. 11. Initial grid of the diode.

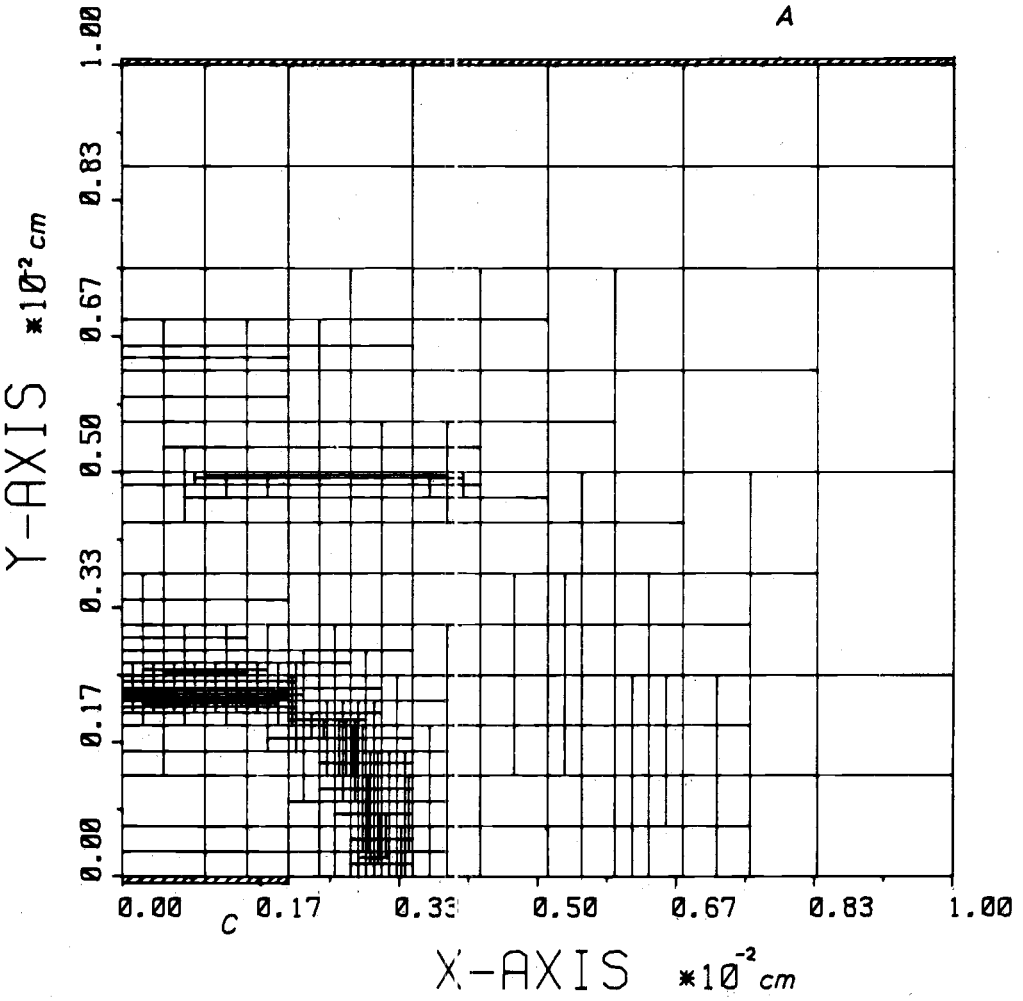


Fig. 12. Final grid of the diode.

TABLE I

	GP	NN	WF	CPT	GP_C	NN_C	WF_C	CPT_C
1	135	7	32.610	1.6	376	8	45.511	7.1
2	172	5	11.026	3.8	525	3	10.857	14.7
3	200	5	3.8149	7.0	941	5	2.782	45.5
4	252	4	2.6598	10.5				
5	288	4	3.2061	14.3				
6	351	9	2.4362	24.4				
7	438	6	1.7605	36.2				
8	550	6	1.5648	52.3				
9	763	4	.71152	74.6				
10	916	4	.09530	100.0				

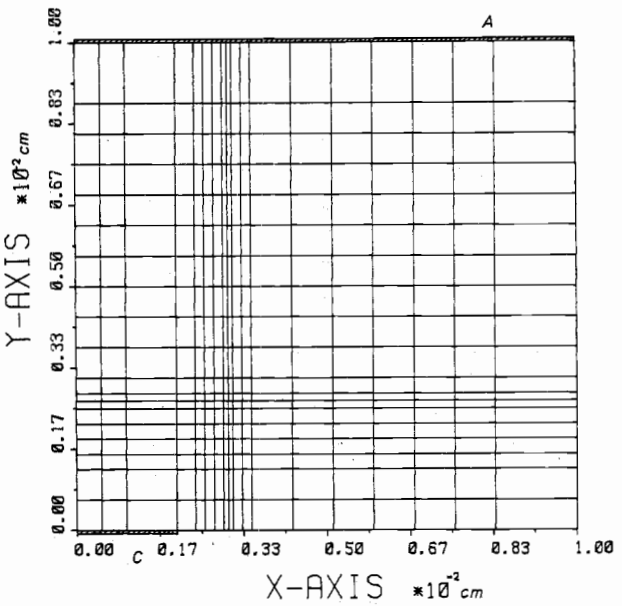


Fig. 13. Initial classical grid of the diode.

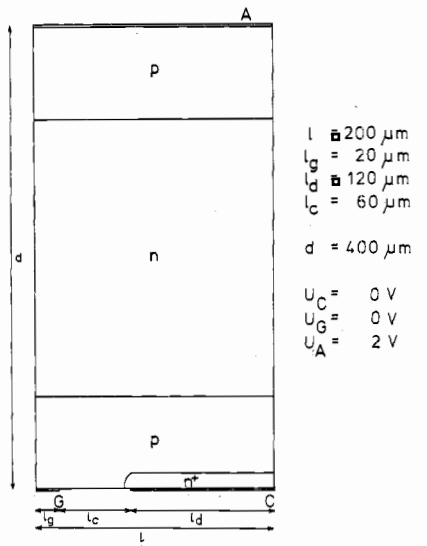


Fig. 14. Geometry thyristor.

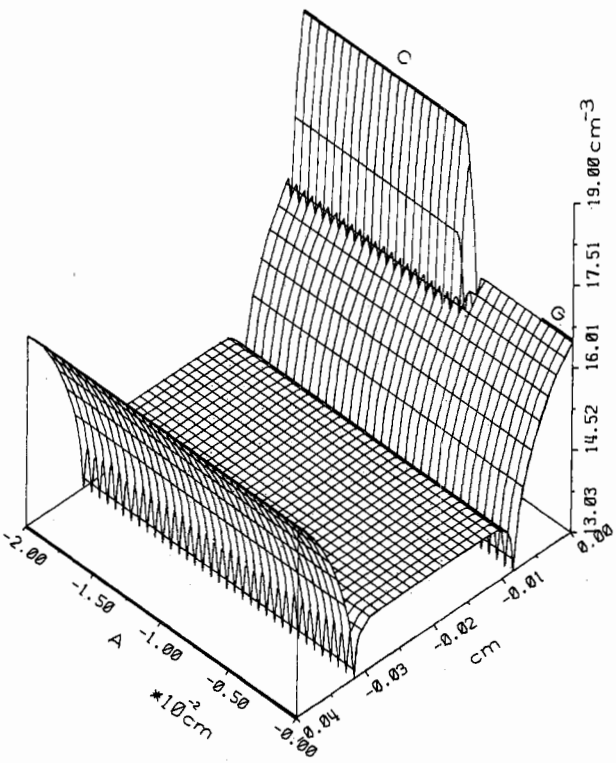


Fig. 15. Doping profile of the thyristor.

NN for calculation and the used CPU time *CPT* in percent of the total jobtime at each grid is given in Table I.

The maximum value of the weight function *WF* proportional to the error functional declines, and the final global discretization error does not exceed 0.1 percent of the applied voltage. We have performed the same calculation on a regular grid without terminating lines to compare the "finite boxes" approach with the classical method. The corresponding initial grid (Fig. 13) consists of 376 points. The number of Newton steps for the solution at each grid, the maximum value of the weight function, and the CPU time in percent of the total time for the "finite boxes" calculation are shown in Table I indicated by subscript *c*. After only three updates, 941 points have been generated and 45.5 percent of the computation time has been used. The "finite boxes" approach needs only 10–24 percent for the same accuracy (maximum value of the weight function).

B. Thyristor

The geometry of the simulated power thyristor is outlined in Fig. 14 and the assumed doping profile is shown in Fig. 15. The demonstrated point of operation is on the ON characteristic. Since the cathode as well as the gate contact are set to 0 V, an emitter shortcut is simulated this way. The final grid-point allocation (Fig. 16) gives an idea of the saving when using "finite boxes" instead of a classical regular mesh. At the applied anode voltage of 2 V, the electron and the hole concentration both exceed by far the impurity concentration of the substrate. The carrier densities differ from each other only in a region near the ohmic contacts. Therefore only the electron concentration is plotted (Fig. 17). The current through the device is strongly influenced by the models of the physical parameters. The generation/recombination model includes a Shockley–Read–Hall term as well as the Auger process [17]. The model of the carrier mobilities takes into account the carrier-carrier and impurity scattering [18], [19]. The resulting electron current density is shown in Fig. 18. The peak

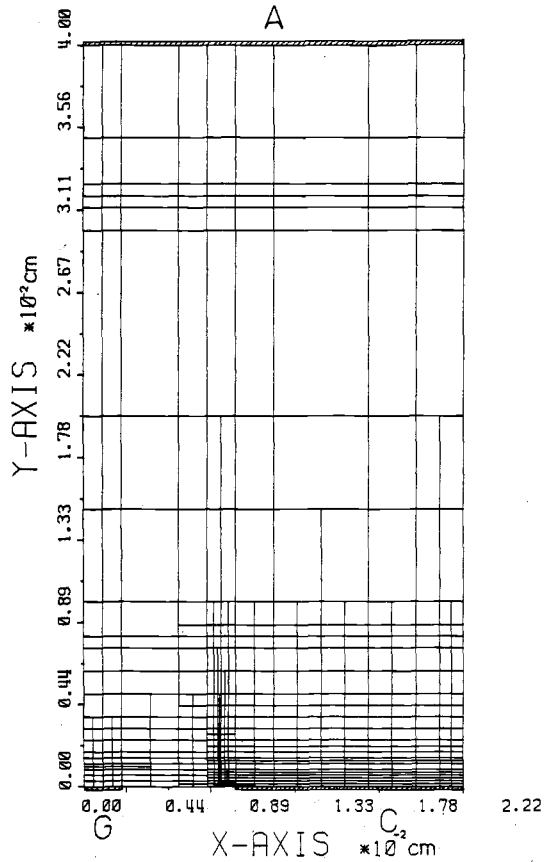


Fig. 16. Final grid of the thyristor.

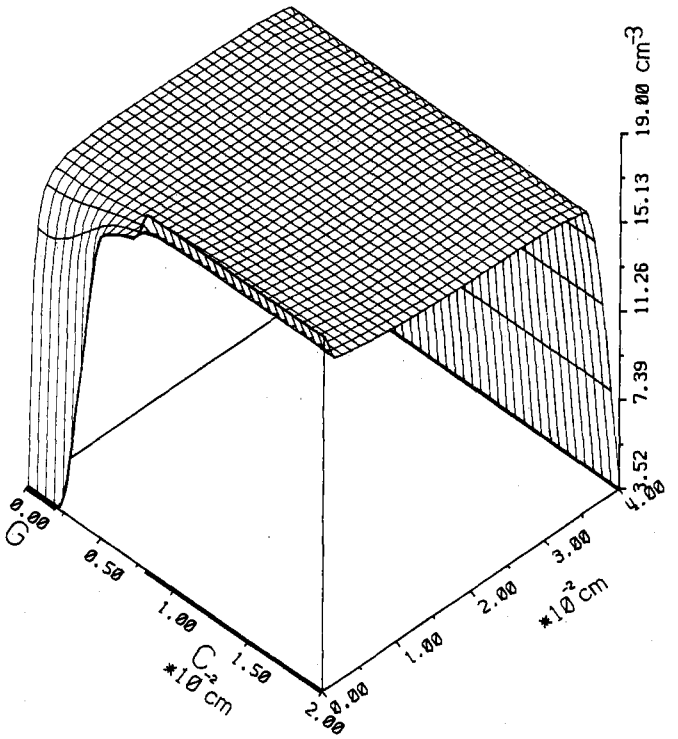


Fig. 17. Electron distribution of the thyristor.

at the edge of the cathode can be interpreted as follows. In the homogenous n-doped drift region in the center of the device, the electron current density is constant along any cross section. The electrons cannot reach the gate contact, and therefore the transfer current must pass through a smaller area resulting in a peak at the edge of the current flow.

C. GaAs MESFET

As an example of a compound semiconductor (GaAs) device, we present the simulations of a GaAs MESFET. Its size is much smaller than that of the two devices described earlier. Fig. 19 outlines the geometry. The Schottky contact (gate) is recessed into the active layer and diminishes the cross section of the conducting channel. The chosen operating point lies within the saturation region. Voltages of 4 V at the drain and -0.8 V at the gate are applied with respect to source. Fig. 20 shows the final mesh. The discretization in the region between the gate and the drain electrode as well as in the boundary layer between the active epilayer and the substrate is very fine. Now we present results for two different models of the electron mobility. Using a constant value for μ gives the electron concentration of Fig. 21. A space-charge region is built up between the gate and drain, and the electron concentration does not exceed the concentration of dopants anywhere. Secondly we have investigated a field-dependent mobility model including velocity saturation and negative differential mobility [20]. In the conducting channel underneath the region of the gate towards the drain, a stationary dipole is formed by an excess electron density and positive donor charge (Fig. 22). The corresponding potential distribution (Fig. 23) exhibits a locally increased electric field there.

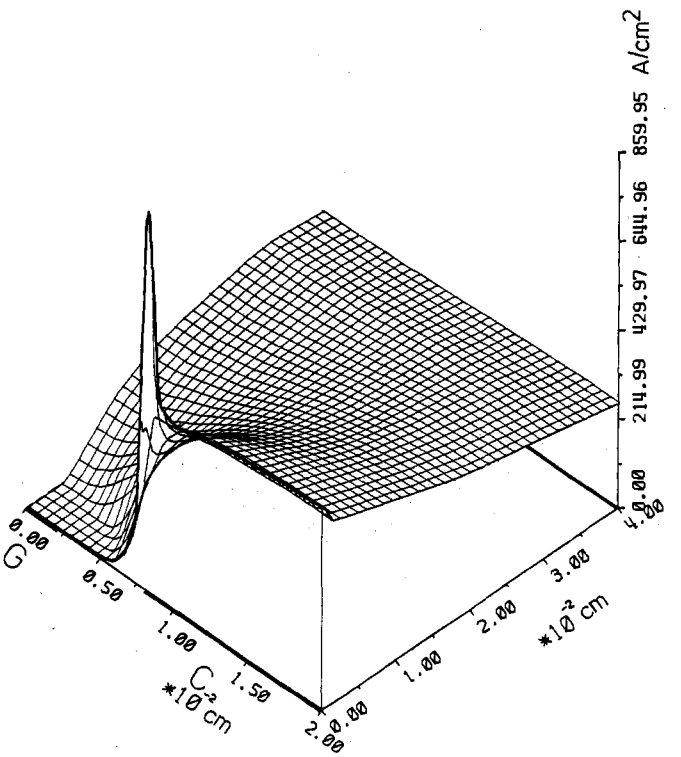
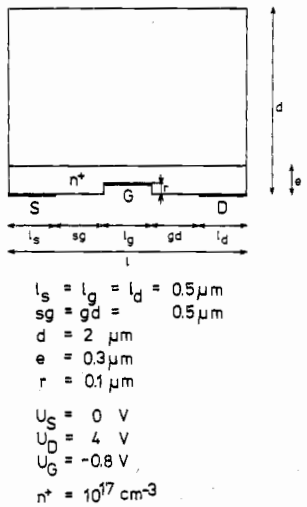


Fig. 18. Modulus of the electron current of the thyristor ($U_{AC} = 2$ V, $U_{GC} = 0$ V).



$l_S = l_G = l_D = 0.5\,\mu m$
$sg = gd = 0.5\,\mu m$
$d = 2\ \mu m$
$e = 0.3\,\mu m$
$r = 0.1\,\mu m$

$U_S = 0\ V$
$U_D = 4\ V$
$U_G = -0.8\ V$
$n^+ = 10^{17}\ cm^{-3}$

Fig. 19. Geometry MESFET.

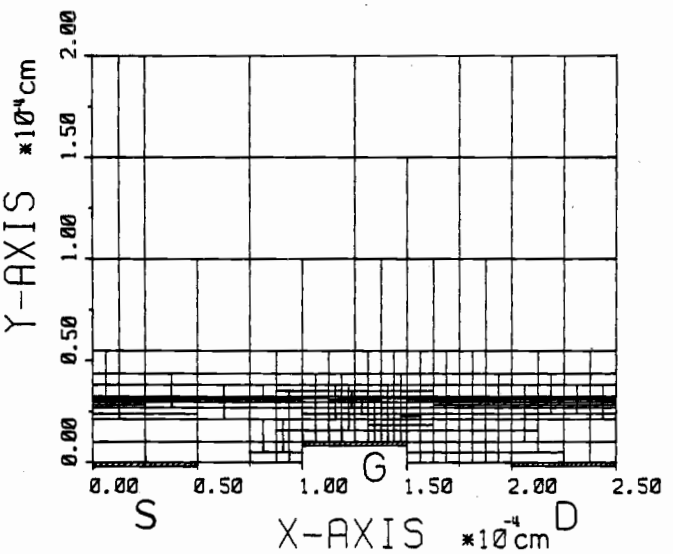


Fig. 20. Final grid of the MESFET.

The transfer current with the field-dependent mobility only slightly exceeds 10 percent of the value resulting from constant mobility. A vector plot (Fig. 24) shows the direction of the electron flux.

V. Conclusion

A two-dimensional numerical-device simulation system has been presented. We generalized the classical finite-difference method developing the so called "finite boxes" approach which allows an optimal grid point allocation and can be applied to nonrectangular devices. Some regularity conditions for the mesh have been investigated. Some typical "finite boxes" grids have been compared to classical regular grids demonstrating the advantages of the former. The setup of the initial mesh has been explained. The automatic grid refinement is performed by equidistributing the local discretization error.

The linear system of equations obtained by the "finite boxes" discretization is solved by a modified Newton method. Two algorithms have been implemented, a simultaneous solution of the coupled equations and a block Newton SOR method. To achieve high flexibility, the physical parameters can be defined by user-supplied models. This requires numerical calculation of parts of the coefficients of the Jacobian. The calculation of the iteration matrix is rather time consuming because of the numerical evaluation of the external functions. Therefore supplementary algorithms have been implemented. An efficient ordering of the grid points based on the "minimum-

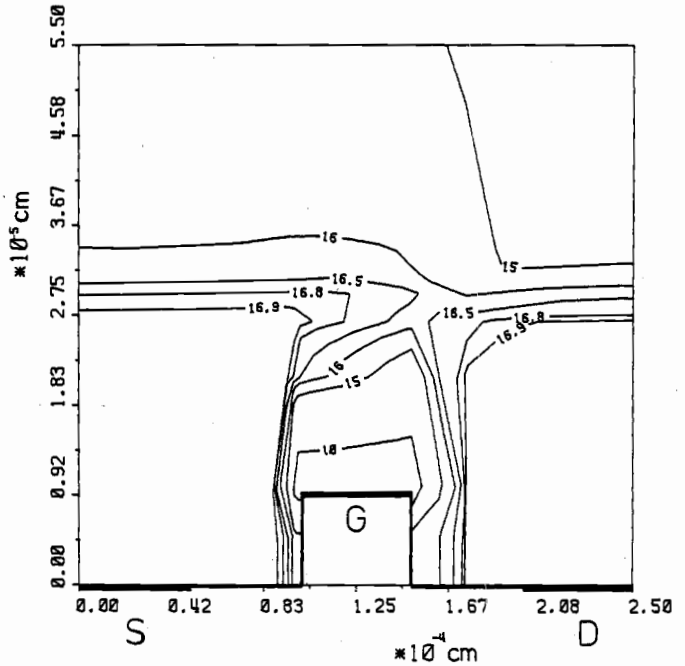


Fig. 21. Electron distribution of the MESFET with constant mobility.

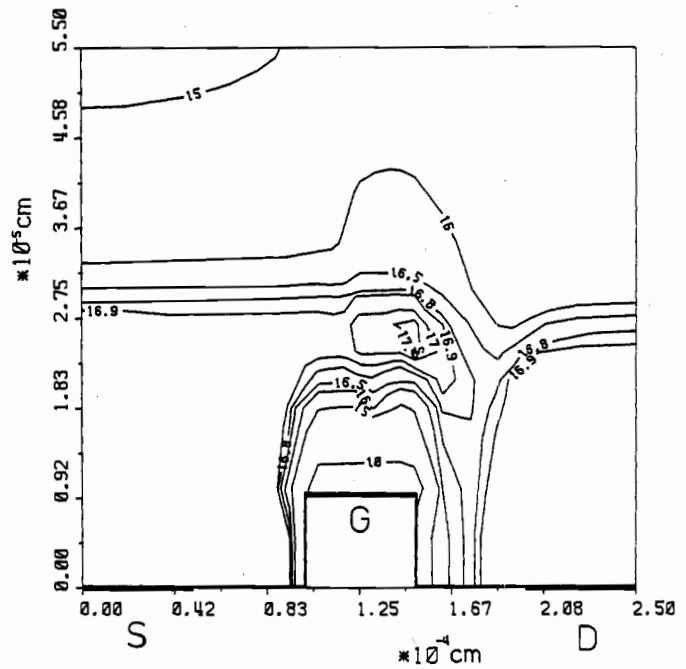


Fig. 22. Electron distribution of the MESFET with field-dependent mobility.

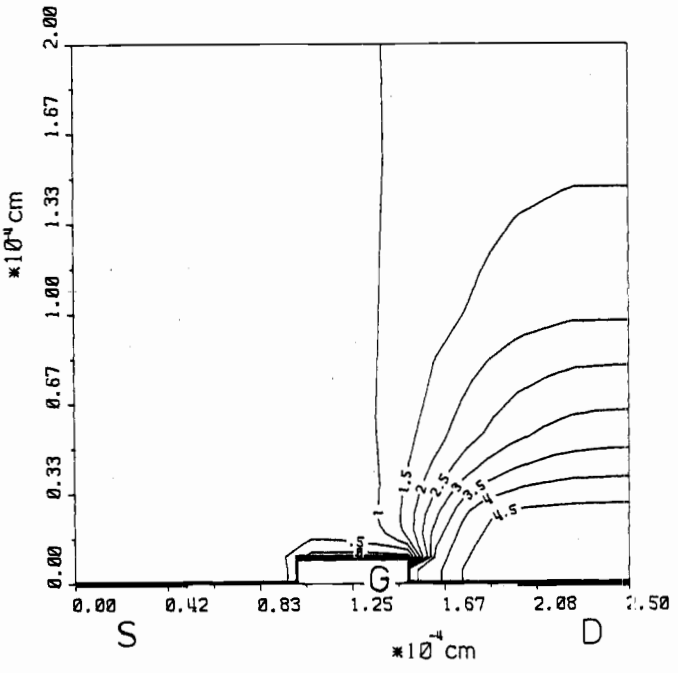


Fig. 23. Electrical potential of the MESFET ($U_{GS} = -0.8$ V, $U_{DS} = 4$ V).

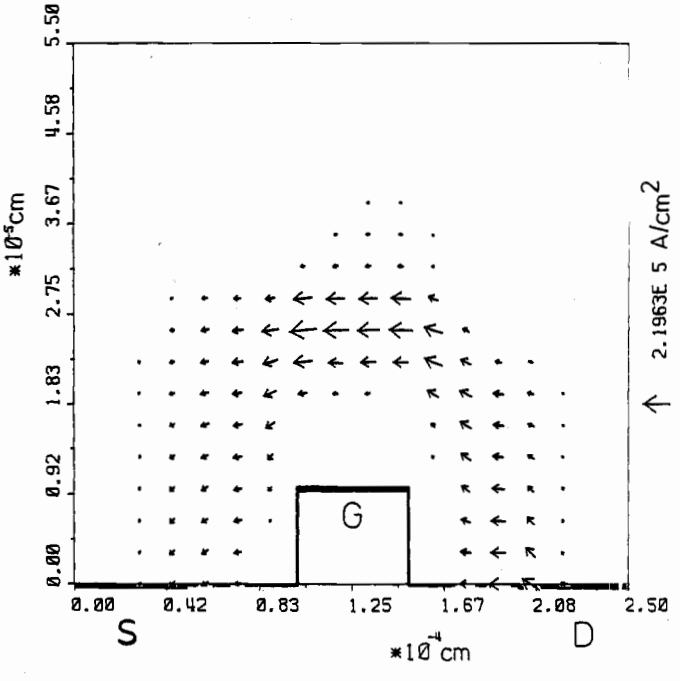


Fig. 24. Electron current flux of the MESFET.

degree" algorithm reduces the memory demand during the solution of the linear system of equations.

We demonstrated our concept on three types of devices, namely a high-voltage diode, a power thyristor, and a GaAs MESFET. The presented results gave some outlook of the great field of applications for the "finite boxes" approach.

Acknowledgment

Helpful discussions with Prof. Dr. H. W. Pötzl who also critically read the manuscript and Dr. A. Schütz were highly appreciated. We should also like to thank the "Interuniversitäre Rechenzentrum" for the excellent computer access.

References

[1] M. S. Adler, "A method for achieving and choosing variable density grids in finite difference formulations and the importance of degeneracy and band gap narrowing in device modeling," in *Proc. Nasecode 1 Conf.*, pp. 3–30, 1979.

[2] G. E. Forsythe, and W. R. Wasow, *Finite Difference Methods for Partial Differential Equations*. New York: Wiley, 1960.

[3] D. L. Scharfetter and H. K. Gummel, "Large-signal analysis of a silicon read diode oscillator," *IEEE Trans. Electron Devices*, vol. ED-16, pp. 64–77, 1969.

[4] P. Markowich, C. Ringhofer, S. Selberherr, and M. Lentini, "A singular perturbation approach for the analysis of the fundamental semiconductor equations," MRC Univ. of Wisconsin Rep. 2482, 1982.

[5] P. A. Markowich, C. Ringhofer, S. Selberherr, and E. Langer, "A singularly perturbed boundary value problem modelling a semiconductor device," Univ. of Wisconsin, MRC Rep. 2388, 1982.

[6] G. H. Meyer, "On solving nonlinear equations with a one parameter operator imbedding," *SIAM J. Numer. Anal.*, vol. 5, pp. 739–752, 1968.

[7] R. E. Bank and D. J. Rose, "Parameter selection for Newton-like methods applicable to nonlinear partial differential equations," *SIAM J. Numer. Anal.*, vol. 17, pp. 806–822, 1980.

[8] P. Deuflhard, "A modified Newton method for the solution of ill-conditioned systems of nonlinear equations with application to multiple shooting," *Numer. Math.* vol. 22, pp. 289–315, 1974.

[9] T. Meis and U. Marcowitz, *Numerische Behandlung partieller Differentialgleichungen*. Berlin: Springer, 1978.

[10] G. D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*. Oxford: Clarendon Press, 1978.

[11] A. R. Curtis and J. K. Reid, "The choice of step lengths when using differences to approximate Jacobian matrices," *J. Inst. Math. Its Appl.*, vol. 13, pp. 121–126, 1974.

[12] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic Press, 1970.

[13] A. George and J. W. H. Liu, "A minimal storage implementation of the minimum degree algorithm," *SIAM J. Numer. Anal.*, vol. 17, pp. 282–299, 1980.

[14] A. F. Franz and G. A. Franz, "Ordering of directed graphs: A modified minimum degree method compared to other algorithms," to be published.

[15] V. C. Alwin, D. H. Navon, and L. J. Turgeon, "Time-dependent carrier flow in a transistor structure under nonisothermal conditions," *IEEE Trans. Electron Devices*, vol. ED-24, pp. 1297–1304, Nov. 1977.

[16] W. R. Frensley, "Power-limiting breakdown effects in GaAs MESFET's," *IEEE Trans. Electron Devices*, vol. ED-28, pp. 962–970, Aug. 1981.

[17] J. Dziewior and W. Schmid, "Auger coefficients for highly doped and highly excited silicon," *Appl. Phys. Lett.*, vol. 31, pp. 346–348, 1977.

[18] S. Selberherr, A. Schütz, and H. Pötzl, "Two dimensional MOS transistor modeling," in *Proc. NATO ASI Process Device Simulation MOS-VLSI Circ.*, 1982.

[19] J. M. Dorkel and P. Leturcq, "Carrier mobilities in silicon semi-empirically related to temperature, doping and injection level," *Solid-State Electron.*, vol. 24, pp. 821–825, 1981.

[20] P. Smith *et al.*, "Electron velocity in Si and GaAs at very high electric fields," *Appl. Phys. Lett.*, vol. 37, pp. 797, 1980.