A Novel Finite-Element Approach to Device Modeling

JÖRG MACHEK AND SIEGFRIED SELBERHERR, MEMBER, IEEE

Abstract—**A finite-element device-simulation program is presented. An adaptive grid-refinement procedure is used to minimize the number of nodes. Two different kinds of elements are generated (triangles and rectangles) thus enabling the use of an irregular mesh. Different shape functions have been developed for the three variables; they are linear/bilinear for the electric potential and linear/bilinear in Bernoulli-like functions for the quasi-Fermi potentials. Numerical examples are presented.**

I. Introduction

SINCE ITS conception in the 1940's, the finite-element method has become a powerful tool for solving virtually any system of partial differential equations. Its roots are to be found in the field of structural engineering, where analogies were developed between actual discrete elements and the continuum. Much of the terminology stems from these days (e.g., shape functions, stiffness matrix, etc.). Only in recent years has the impact of this method on the analysis of semiconductor devices been felt.

The characteristics of a semiconductor device are modeled by three coupled nonlinear partial differential equations. They consist of the Poisson equation and the equations of hole and electron current continuity. There exist very effective and general packages which solve the semiconductor equations using the finite element approach [8], [1], [2]. In the present analysis, the three equations are tackled using a finite-element method in its residual formulation [7], [12], [16]. Therefore the weighted residual integrals over the whole domain are set to zero. The unknowns $\psi, \varphi_n, \varphi_p$, have to be approximated in every subdomain (element) by a linear combination of basis functions.

A critical step in the application of the finite-element method is the design of the mesh and, therefore, the selection of proper elements [4], [8], [9]. To allow variable grid spacings, an adaptive mesh-refinement process has been developed which generates two different kinds of elements (triangles and rectangles). This allows a mesh generation/refinement where broad grid spacings can alternate with regions where a fine mesh is imperative without creating an excessive number of nodes. This will be considered in detail in Section II.

Since the shape functions are designed to approximate the solution in the diverse elements, they too play an important role in the flexibility and accuracy of the final simulation. If they are constructed to anticipate the correct solution, convergence will be accelerated [2]. Shape functions which are linear in a triangular element cause oscillations of the solution [1], [3] and are, therefore, not feasible. This led to the development of a novel kind of shape functions for the exponentials of the quasi-Fermi levels which describe an exponential distribution of the carrier densities exactly. They are linear/bilinear for the electric potential and linear/bilinear in Bernoulli-like functions for the exponentials of the quasi-Fermi levels. This will be the topic of Section III.

In Section IV the finite-element analysis of the semiconductor equations and their numerical implementation is dealt with. It shows that the Poisson equation and the continuity equations are calculated differently. The consistent treatment of ordinary rectangular elements and the 5-node transition element is demonstrated.

As a descriptive bench mark for these finite-element studies, a bipolar structure was used.

II. The Mesh

In the past, much attention has been focused on the use of triangular elements for the discretization of the simulation domain. The geometric drawbacks this procedure imposes are obvious. A general mapping from a triangular parent mesh to a complex geometry can be found, but local refinement of the grid is rather difficult and loosening of the grid spacing is almost impossible. This line of thought, coupled with physical insight and investigations of the local discretization error, made it evident that the global solution area has to be made up of subregions which require a fine resolution and, alternately, regions where a broad grid spacing is sufficient.

Considerations of interelemental current continuity and the intuitive assumption of a linear electric potential ψ led directly to a mainly rectangular subdivision of the simulation area. To locally refine or coarsen the grid, a 5-node transition band had to be defined which, for simplicity's sake, was assumed to be generated by three triangles (Fig. 1).

The discretization error of a triangularized mesh is proportional to $1/\sin\alpha$ where α is the smallest angle of the triangle. One tries, therefore, as far as possible to avoid obtuse triangles. Another plausible explanation for the rejection of such triangles is demonstrated in the finite-element analysis of the FIELDAY program [1], [2], [15]. Such

Manuscript received November 2, 1982; revised April 6, 1983. This work was sponsored by Siemens AG, Munich, Germany.

The authors are with Abteilung für Physikalische Elektronik, Institut für Allgemeine Elektrotechnik und Elektronik, Technical University of Vienna, Gusshausstrasse 27, A-1040 Vienna, Austria.

0018-9383/83/0900-1083$01.00 ©1983 IEEE

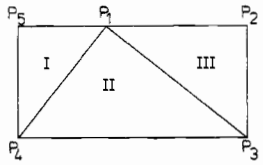

Fig. 1. Transition band.

an element exhibits the unphysical result of a negative flux cross section.

Owing to the very special shape of the 5-node element, it can be shown that as long as the aspect ratio of the rectangular structure is between 0.5 and 2, no obtuse triangle is formed. The larger the aspect ratio becomes, the more the midside node is pushed to one of its neighbors in order to avoid obtuse triangles. The first aim of computations involving a mesh construction and/or refinement is to find a mesh of smallest possible number of elements (nodes) such that the error is below a given tolerance. Computational costs of a refinement step can be rather high, and it certainly increases with the number of nodes. One has to find a balance between optimal accuracy and a given computational cost range [8], [9].

An adaptive grid-generation and grid-refinement process has been developed [30]. The initial grid is established using only information on the device geometry and the doping profile. Boundary zones and interior regions are each marked with different attributes. The "preinitial" grid consists of the absolutely essential elements which contain nodes on every boundary and in every interior region. Thus it is guaranteed that every subdomain is included in the simulation. This very first grid is locally refined in regions where the doping gradient is high, utilizing the physical knowledge that in such areas the local discretization error (which depends on the space charge) [28] is rather large. With this initial grid the discrete nonlinear system is solved to the desired accuracy, and the grid is updated again using the elemental space charge as indication for grid refinement. This seemingly heuristic characterization of an optimal mesh is consistent with the local computation of an error estimate after Babuska [4]. It is inherent in *a posteriori* error estimates that the global convergence condition can only be guaranteed after at least one additional computation with the refined grid.

After every grid update a regularization procedure is called. The only restriction is that refinement has to be locally confined to one direction—the transition band is to be a 5-node macroelement.

The boundary is approximated by a piecewise linear function which implies that at a slanted boundary purely triangular elements exist (Fig. 2).

Fig. 3 shows the simulation grid of a diode with a curved p-n junction. The local refinement near the anticipated p-n junction is noticeable. This is the simulation grid for a rather low bias (1 V), and no appreciable space-charge layer has yet developed. The fine grid is, therefore, confined to the metallurgical junction.

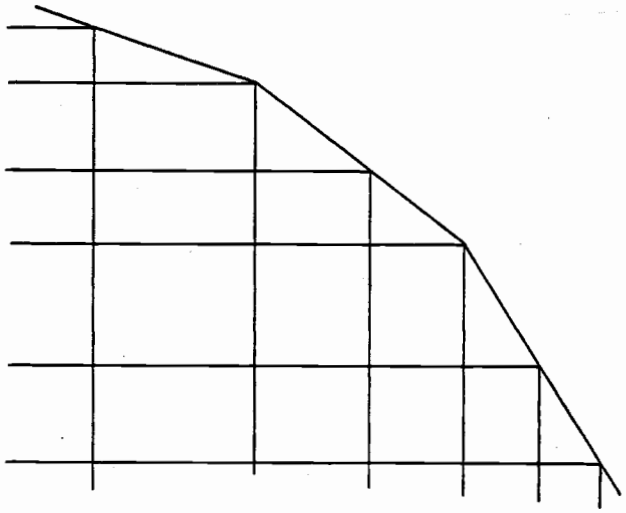

Fig. 2. Slanted boundary.

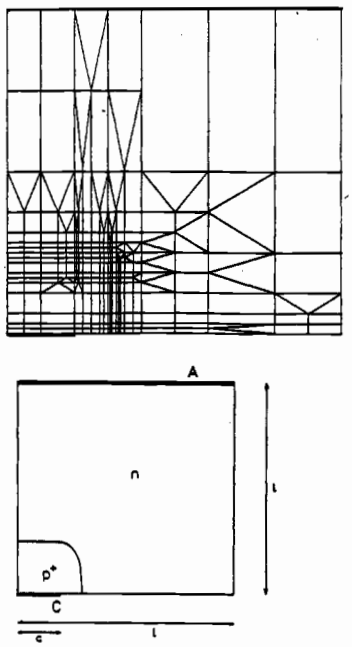

Fig. 3. Simulation grid of diode with curved p-n junction. $l = 100\ \mu$m, $c = 20\ \mu$m, $U_A = 0$ V, $U_C = 1$ V.

III. The Shape Functions

A central task in the numerical solution of the semiconductor equations is the preservation of current continuity parallel to element edges and the preconditioning of the system in a way that the exponential variation of carrier densities will be represented exactly by the discrete system.

A. Rectangle

It is well known that an ordinary differential equation of second order with an exponential nonlinearity or with solutions exhibiting exponential layer behavior can be discretized using exponential fitting [24], [25]. The logical generalization to an elliptic system in two dimensions with

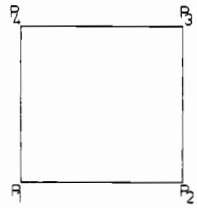

Fig. 4. Unit square.

a similar nonlinearity seems to be the concise solution of the boundary-value problem over one element (Fig. 4).

The current continuity equation without recombination gives (all quantities suitably normalized and all calculations carried out on the unit square)

$$\operatorname{div}(J_n) = 0 \tag{1}$$

where

$$J_n = \exp(\psi)\operatorname{grad}(\zeta) \tag{2}$$

with

$$\zeta = \exp(-\varphi_n).$$

Substituting (2) into (1) yields

$$\begin{aligned} 0 = \operatorname{div}(J_n) &= \operatorname{grad}(\exp(\psi))\operatorname{grad}(\zeta) \\ &\quad + \exp(\psi)\operatorname{div}(\operatorname{grad}(\zeta)) \\ &= \exp(\psi)[\operatorname{grad}(\psi)\operatorname{grad}(\zeta) + \operatorname{div}(\operatorname{grad}(\zeta))] \end{aligned} \tag{3}$$

or, fully expanded

$$\frac{\partial\psi}{\partial x}\frac{\partial\zeta}{\partial x} + \frac{\partial^2\zeta}{\partial x^2} = -\frac{\partial\psi}{\partial y}\frac{\partial\zeta}{\partial y} - \frac{\partial^2\zeta}{\partial y^2}. \tag{4}$$

This differential equation can be solved for ζ assuming a known electric potential ψ and known boundary values for ζ. It has so far only been possible to find an analytical solution under the additional assumption of a constant field, i.e., linearity of ψ in x and y. Although this is not totally consistent with bilinear shape functions for ψ, the resulting solution is still worth close consideration. It comes about by separation of the variables

$$\zeta = \Xi(x)\mathrm{H}(y).$$

This leads to two independent ordinary differential equations of second order.

$$\begin{aligned} \Xi'' + \frac{\partial\psi}{\partial x}\Xi' - \lambda\Xi = 0 \\ \mathrm{H}'' + \frac{\partial\psi}{\partial y}\mathrm{H}' + \lambda\mathrm{H} = 0 \end{aligned} \tag{5}$$

with the solution for $\lambda = 0$.

$$\begin{aligned} \Xi &= C_1 + C_2\exp\left(-\frac{\partial\psi}{\partial x}x\right) \\ \mathrm{H} &= K_1 + K_2\exp\left(-\frac{\partial\psi}{\partial y}y\right). \end{aligned} \tag{6}$$

Calculating the product and using the boundary values to evaluate the constants yields

$$\begin{aligned} \zeta(x,y) = \zeta_1\alpha_1(x,y) + \zeta_2\alpha_2(x,y) \\ + \zeta_3\alpha_3(x,y) + \zeta_4\alpha_4(x,y). \end{aligned} \tag{7}$$

The $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ are the values of the function at the 4 nodes, and the α_i are the basis functions.

$$\alpha_1(x,y) = [1-\varphi_1(x,y)][1-\varphi_2(x,y)] \tag{8}$$

$$\alpha_2(x,y) = \varphi_1(x,y)[1-\varphi_2(x,y)] \tag{9}$$

$$\alpha_3(x,y) = \varphi_1(x,y)\varphi_2(x,y) \tag{10}$$

$$\alpha_4(x,y) = [1-\varphi_1(x,y)]\varphi_2(x,y) \tag{11}$$

for $\exp(-\varphi_n)$, and

$$\beta_1(x,y) = [1-\sigma_1(x,y)][1-\sigma_2(x,y)] \tag{12}$$

$$\beta_2(x,y) = \sigma_1(x,y)[1-\sigma_2(x,y)] \tag{13}$$

$$\beta_3(x,y) = \sigma_1(x,y)\sigma_2(x,y) \tag{14}$$

$$\beta_4(x,y) = [1-\sigma_1(x,y)]\sigma_2(x,y) \tag{15}$$

for $\exp(\varphi_p)$, where

$$\begin{aligned} \varphi_1(x,y) &= f\left(x - \frac{\partial\psi}{\partial x}\right) \\ \varphi_2(x,y) &= f\left(y - \frac{\partial\psi}{\partial y}\right) \\ \sigma_1(x,y) &= f\left(x, \frac{\partial\psi}{\partial x}\right) \\ \sigma_2(x,y) &= f\left(y, \frac{\partial\psi}{\partial y}\right) \end{aligned}$$

with

$$f(x,a) = (\exp(ax)-1)/(\exp(a)-1).$$

With the appropriate scaling factors, this derivation holds for any rectangle. This implies that physically plausible shape functions can only be derived from a solution of the Dirichlet problem. It proved that formulation of shape functions in closed form were only possible if a constant field was assumed in each rectangle.

With these basis functions, one can derive expressions in the unit square for the electron concentration

$$n(x,y) = n_1\beta_1 + n_2\beta_2 + n_3\beta_3 + n_4\beta_4 \tag{16}$$

and for the electron current density

$$\begin{aligned} J_{nx}(x,y) &= (1-\sigma_2)\left[B\left(\frac{\partial\psi}{\partial x}\right)n_2 - B\left(-\frac{\partial\psi}{\partial x}\right)n_1\right] \\ &\quad + \sigma_2\left[B\left(\frac{\partial\psi}{\partial x}\right)n_3 - B\left(-\frac{\partial\psi}{\partial x}\right)n_4\right] \end{aligned} \tag{17}$$

$$\begin{aligned} J_{ny}(x,y) &= (1-\sigma_1)\left[B\left(\frac{\partial\psi}{\partial y}\right)n_4 - B\left(-\frac{\partial\psi}{\partial y}\right)n_1\right] \\ &\quad + \sigma_1\left[B\left(\frac{\partial\psi}{\partial y}\right)n_3 - B\left(-\frac{\partial\psi}{\partial y}\right)n_2\right] \end{aligned} \tag{18}$$

and similarly for the hole concentration

$$p(x,y) = p_1\alpha_1 + p_2\alpha_2 + p_3\alpha_3 + p_4\alpha_4 \tag{19}$$

and for the hole current density

$$J_{px}(x,y) = (1-\varphi_2)\left[B\left(\frac{\partial\psi}{\partial x}\right)p_2 - B\left(-\frac{\partial\psi}{\partial x}\right)p_1\right] + \varphi_2\left[B\left(\frac{\partial\psi}{\partial x}\right)p_3 - B\left(-\frac{\partial\psi}{\partial x}\right)p_4\right] \tag{20}$$

$$J_{py}(x,y) = (1-\varphi_1)\left[B\left(\frac{\partial\psi}{\partial y}\right)p_4 - B\left(-\frac{\partial\psi}{\partial y}\right)p_1\right] + \varphi_1\left[B\left(\frac{\partial\psi}{\partial y}\right)p_3 - B\left(-\frac{\partial\psi}{\partial y}\right)p_2\right] \tag{21}$$

where $B(\alpha)$ is the Bernoulli function

$$B(\alpha) = \alpha/(\exp(\alpha)-1). \tag{22}$$

It is interesting that in these derived quantities the basis functions of the electron Fermi levels determine the hole distribution and *vice versa*. Another though not surprising property which can be deduced for the current relation is that for a high electric field the current degenerates into a pure field current

$$J_{nx} = \frac{\partial\psi}{\partial x}[n_1(1-y) + n_4 y] \tag{23}$$

and, in the case of no field, the current only consists of diffusion current

$$J_{ny} = (1-\sigma_2)(n_2 - n_1) + \sigma_2(n_3 - n_1) \tag{24}$$

for $\frac{\partial\psi}{\partial x} \gg 0$ and $\frac{\partial\psi}{\partial y} = 0$.

B. Transition Band

This macroelement is partitioned into three triangles. In each of these triangles we take the current density to be constant. Thus the component of the current density parallel to the edge which connects vertices i and j becomes

$$J_{ij} = \left[n_i B(\psi_i - \psi_j) - n_j B(\psi_j - \psi_i)\right]/l_{ij} \tag{25}$$

where l_{ij} is the length of that edge [1], [2].

Because of the nonlinearity of the Bernoulli function, the current expressions are no longer additive

$$J_{ij} \neq J_{ik} + J_{kj}. \tag{26}$$

Since the main point of interest is continuity of the current component parallel to the edges, one has to give preference to the components in the x and y direction. This is indeed simple for triangles I and III. For rectangular elements it is also obviously trivial. Difficulties arise in nonrectangular triangles. Although any arbitrary triangle can be linearly transformed into a rectangular triangle, the currents would no longer transform correctly. There exist a number of possibilities to calculate the y component of the current in triangle II. The most promising appeared to us to be the superposition of J_{14} and J_{13}.

Note that this treatment of the transition band is consistent with rectangular elements with bilinear basis functions as well as with triangular elements with linear shape functions for the electric potential ψ as used on slanted boundaries. The elements are numbered counterclockwise.

Fig. 5 shows the rectangular shape functions in the unit square for the degenerate case. One can clearly see the bilinear behavior on that isoline plot, since for every cut parallel to the coordinate axes, the lines of equal level are equidistant. In Fig. 6 a moderate field parallel to the y axis has been applied. The functions have been compressed in the direction of the field, emphasizing the points with higher potential and almost suppressing the points with lower potential. The functions are still linear perpendicular to the field since the isolines are equidistant in x direction. This trend exhibits a behavior very similar to the "upwind"-weighting functions for flow problems mentioned in [7]. In Fig. 7 a field diagonal to the square has been applied, and again one notices the compression of the isolines in the direction of the field. Remarkable is the fact that in the case of the diagonal field, the node with the highest potential is very prominent whereas the one with the lowest potential is almost neglected. At the nodes with the highest potentials, the shape functions tend to become very steep and decrease rapidly, but at the nodes with lower potential the functions are rather flat, giving these nodes a greater weight in the approximation of the solution.

IV. The Solution

In this section we describe the finite-element analysis of the semiconductor equations and their computer implementation. The intention is to solve a system of three quasi-harmonic differential equations [26]

$$\operatorname{div}\left(P^i \operatorname{grad}(u^i)\right) = F^i, \quad i = 1,2,3 \tag{27}$$

where P is a constant (the dielectric permittivity) for the Poisson equation and, essentially, the majority carrier distributions for the two continuity equations (i.e., electron density for the electron current and hole density for the hole current equation). F is the space charge $(n - p - N_D + N_A)$ for the Poisson equation and the recombination-generation term for the continuity equations. The variable u is the electric potential ψ for the Poisson equation and the quasi-Fermi levels φ_n, φ_p for the electron and hole current continuity equation, respectively.

Since the equations (27) are of a type of partial differential equations which do not necessarily derive from a variational principle (because of the convective terms in the continuity equations), a weighted residual method is being used.

A. The Poisson Equation

The electric potential ψ is approximated in terms of basis functions b_j.

$$\psi = \sum_{j=1}^{N} \psi_j b_j \tag{28}$$

where N is the degree of freedom ($N = 3$ for a triangle with linear basis functions; $N = 4$ for a rectangle with bilinear

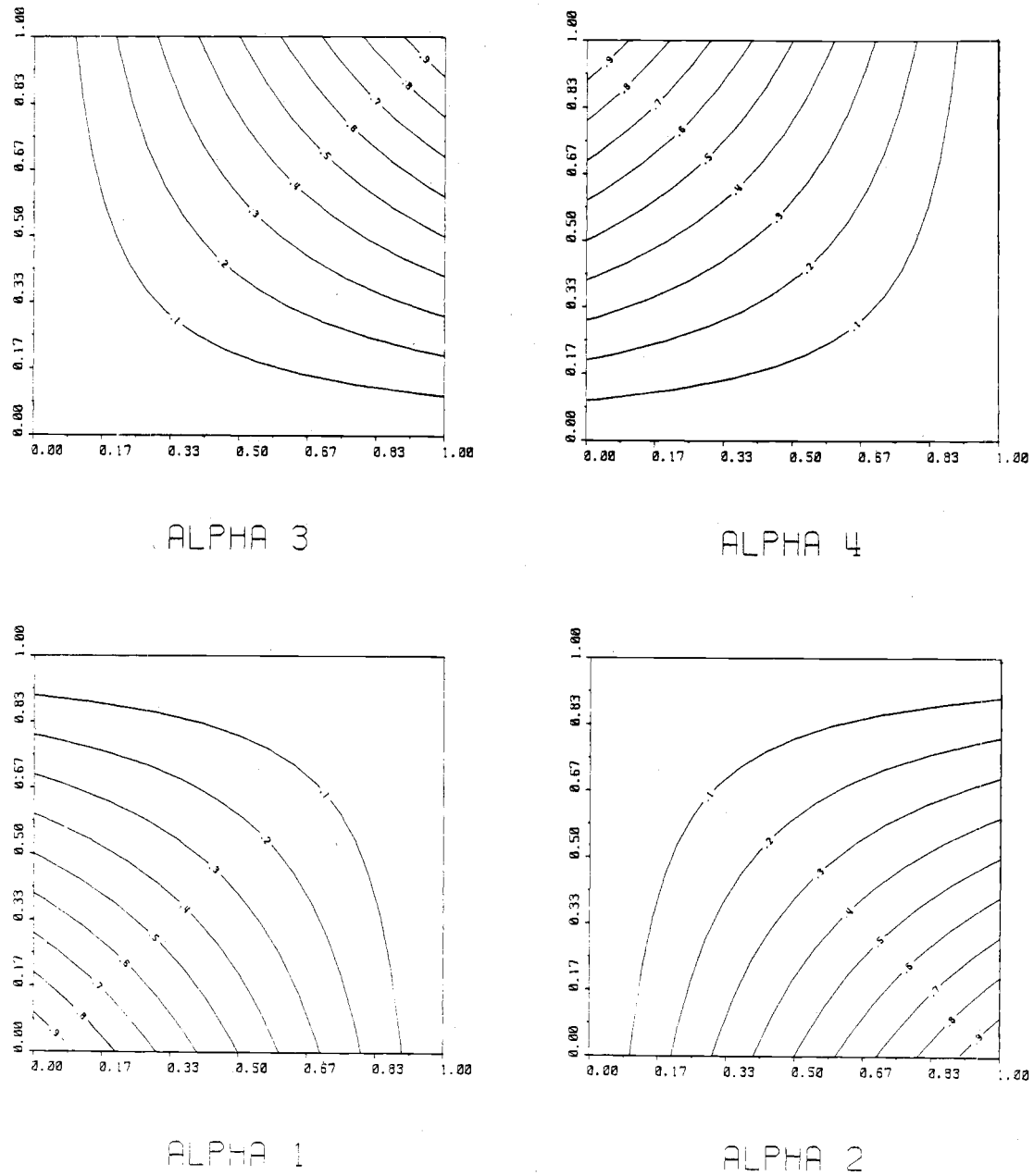

Fig. 5. Shape functions in rectangle for no field.

basis functions). The b_j are the standard bilinear/linear shape functions which can be found in any finite-element textbook [7], [12], [13], [16], [29]. The generalized coordinates (= value of the unknown function at the nodes) ψ_j are determined by the weighted residual conditions

$$0 = \int_\Omega \int \operatorname{div}(\epsilon \operatorname{grad}(\psi)) w_i \, d\Omega - \int_\Omega \int F w_i \, d\Omega$$

$$= \oint_{\partial\Omega} (w_i P \operatorname{grad}(\psi)) \vec{n} \, d\partial\Omega$$

$$- \int_\Omega \int P \operatorname{grad}(\psi) \operatorname{grad}(w_i) \, d\Omega - \int_\Omega \int F w_i \, d\Omega \quad (29)$$

where the w_i are the weighting functions which, in the case of the Poisson equation, are equal to the basis functions (then the procedure is called a Galerkin formulation of the problem). The first integral vanishes on all boundaries where $\vec{n}\operatorname{grad}(\psi) = 0$ (natural Neumann boundary) and where $w_i = b_i = 0$ (forced on all Dirichlet boundaries).

We split the domain Ω into L finite elements whose union is Ω itself. The b_i are different polynomials in each element and are continuous in adjacent elements. Therefore, we can express (11) as a summation over all elements, which allows us to treat one element at a time.

$$r_i = \sum_{l=1}^{L} \left[\int\int_{\Omega^1} \left(\epsilon \operatorname{grad}(\psi) \operatorname{grad}(b_i) \, d\Omega^1 + \int\int_{\Omega^1} F w_i \, d\Omega^1 \right] = 0. \quad (30)$$

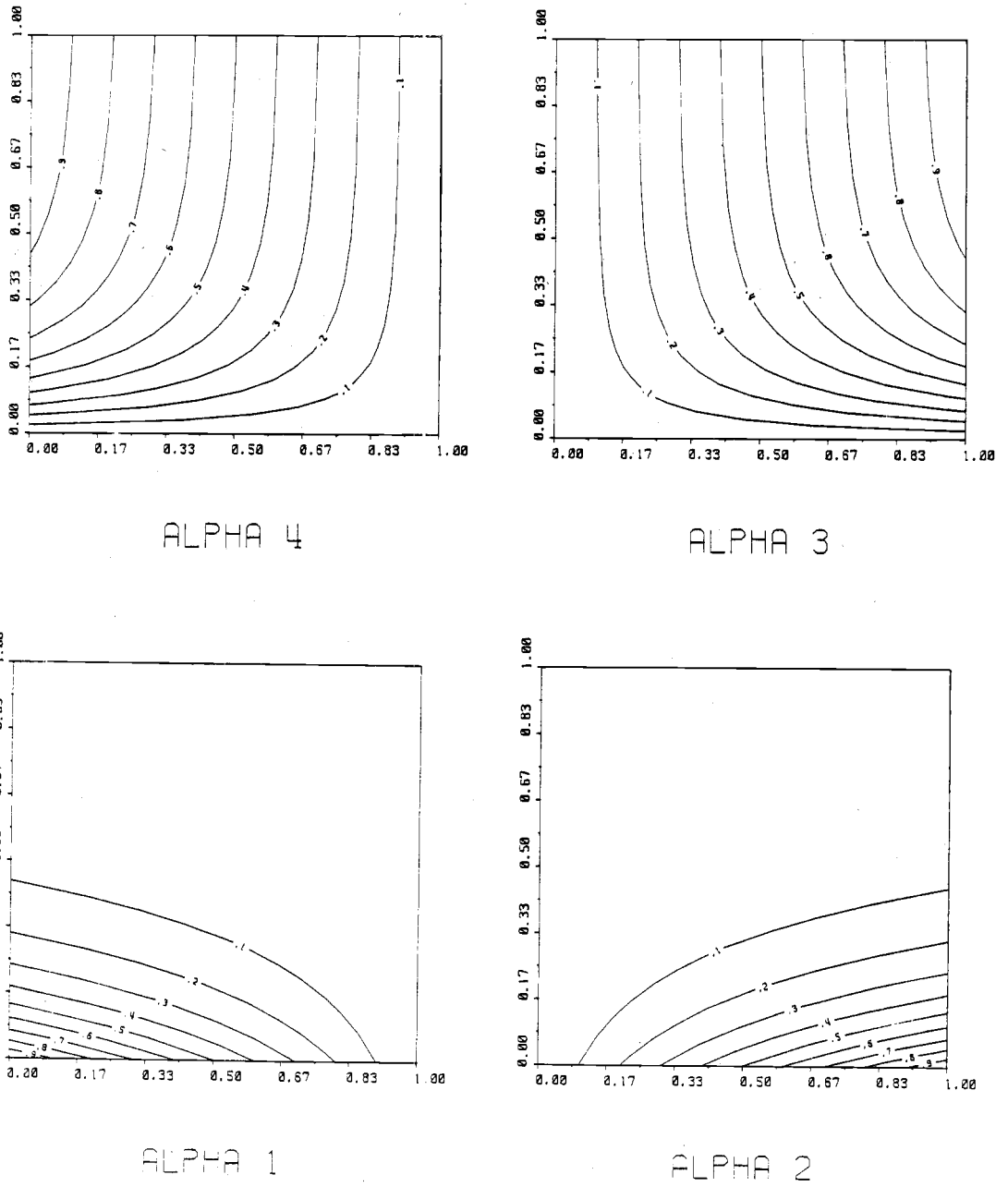

Fig. 6. Shape functions in rectangle for field parallel to the y-axis.

Using the approximation for ψ and substituting in (29) yields

$$r_i = \sum_{l=1}^{L} \sum_{j=1}^{N} \left[\psi_j \int\int_{\Omega^1} \left(\epsilon \operatorname{grad}(b_j) \operatorname{grad}(b_i)\, d\Omega^1 + \int\int_{\Omega^1} F w_i\, d\Omega^1 \right] = 0. \quad (31)$$

The first integral is computed analytically, consisting only of known bilinear functions. As a simple reflection shows, this integral is bilinearly dependant on the mesh size. To simulate a similar behavior for the space charge, the integration of the second term is carried out numerically with a double trapezoidal rule.

B. Current Continuity Equations

In the previous section it has been shown that a close representation of the quasi-Fermi levels (or even their exponentials) is not consistent with the solution of the Dirichlet problem which has been the starting point of our investigations. The procedure to find a discrete residual vector is slightly different to the one employed to discretize the Poisson equation. In the case of the current—continuity equations the $P\operatorname{grad}(u)$ term of the quasi-harmonic

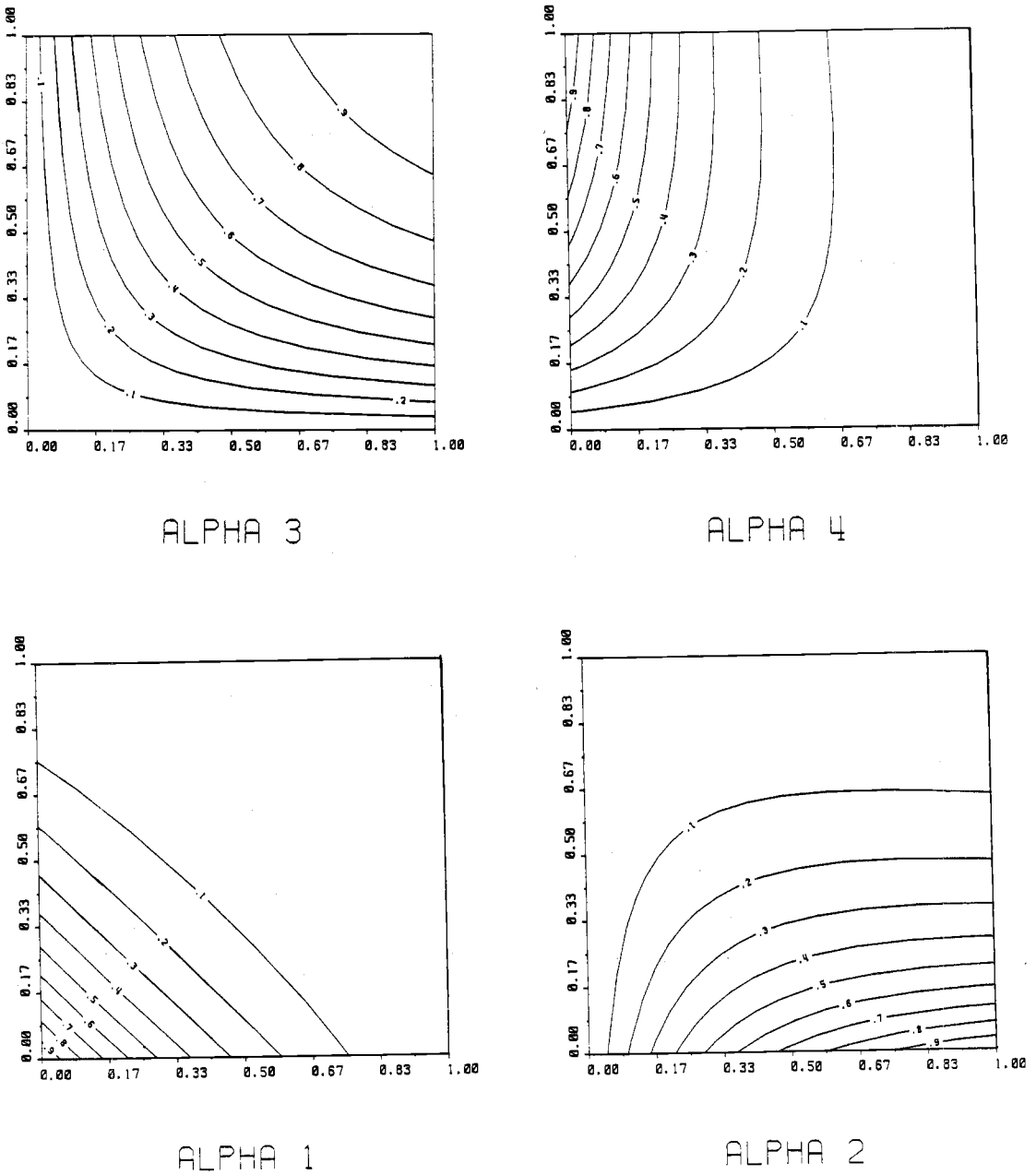

Fig. 7. Shape functions in rectangle for diagonal field.

equation can be interpreted as a current.

$$J_n = -q\mu_n n \operatorname{grad}(\varphi_n)) \tag{32}$$

The continuity equation in its residual formulation looks like

$$r_i = \sum_{l=1}^{L} \left[\int\int_{\Omega^l} \operatorname{div}(J_n) w_i \, d\Omega^l - q \int\int_{\Omega^l} R w_i \, d\Omega^l \right] = 0$$

$$= \sum_{l=1}^{L} \left[\oint_{\partial\Omega^l} J_n w_i \vec{n} \, d\partial\Omega^l - \int\int_{\Omega} J_n \operatorname{grad}(w_i) \, d\Omega^l + q \int\int_{\Omega^l} R w_i \, d\Omega^l \right] = 0. \tag{33}$$

The first (boundary) integral is zero when either the component of J_n normal to the boundary vanishes (holds on Neumann boundaries, because the current leaves the device only at the contacts), or when w_i vanishes, which we force on the Dirichlet boundaries (all contacts).

The numerical integration is slightly different for a rectangle and for a transition band.

In the rectangle the integration is carried out using the trapezoidal rule. This implies that for the numerical evaluation of the integrals only the values of J_n on the element boundaries are needed.

In a transition band, since J_n is the constant

$$\int\int_{\Omega^1} J_n \operatorname{grad}(w_i)\, d\Omega^1 = J_n \int\int_{\Omega^1} \operatorname{grad}(w_i)\, d\Omega^1 = J_n \left[\int\int_{\Omega_I^1} \operatorname{grad}(w_i)\, d\Omega_I^1 + \int\int_{\Omega_{II}^1} \operatorname{grad}(w_i)\, d\Omega_{II}^1 + \int\int_{\Omega_{III}^1} \operatorname{grad}(w_i)\, d\Omega_{III}^1 \right]. \tag{34}$$

Each of these integrals can be calculated analytically, and the contribution of the transition element to current continuity consists of the scalar product of J_n and the sum of the integrals over grad (w_i).

This procedure is justified because the local discretization error due to the linear/bilinear basis functions is in the order of the element size $[0(h)]$. Hence all the integrations and other numerical procedures do not have to be of a higher order. Since integration with the trapezoidal rule has a local discretization error in $0(h)$, its application to the elemental integration is consistent [7].

C. Linearization

The resulting residuals (with the linear/bilinear shape functions of the electric potential as weighting functions) are integrated numerically, and the thus-created system of nonlinear algebraical equations

$$F(x) = 0 \tag{35}$$

is solved iteratively using a Newton-like method [6].

$$A_k(x_{k+1} - x_k) = -F(x_k). \tag{36}$$

The Jacobian matrix is approximated by numerical differentiation. This method is more accurate than [5] and has been shown to hold even for pathological cases [27]. Since this calculation is rather time consuming, an element-wise bypassing algorithm for the update of the entries in the Jacobian matrix has been devised to minimize computer requirements. Convergence is speeded up by the introduction of an acceleration parameter $\lambda \epsilon [0,1]$, so that

$$x_{k+1} = x_k + \lambda(x_{k+1} - x_k) \tag{37}$$

with

$$\|F(x_{k+1})\| < \|F(x_k)\|. \tag{38}$$

The actual matrix inversion is done with a straightforward Gaussian elimination for sparse matrices [23]. To reduce fill-in, the Gibbs–Poole–Stockmeyer ordering algorithm has been used [19], [20], [21], [22].

V. An Example

As an illustrative example, a relatively simple bipolar structure was chosen. In Fig. 3 the simulation grid of the diode with a curved p-n junction is shown. It was chosen because p-n junctions are of paramount importance in modern electronic applications and in the understanding of other semiconductor devices. It is a simple device whose basic theory is well established and closely investigated [28], [31]. It is, therefore, suitable for testing a simulation program and for demonstrating novel approaches. In Fig. 8 the electric potential ψ is plotted for a reverse bias of 1 V. The dimensions of the device are 100 μm × 100 μm, and the space charge layer is not very large. One still notices, however, the potential drop at the junction as one does in

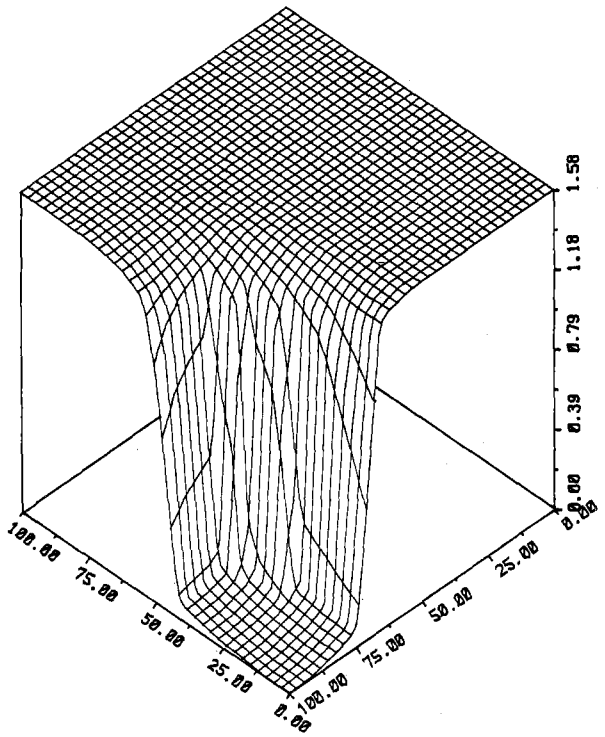

Fig. 8. Potential ψ of diode for a bias of 1 V.

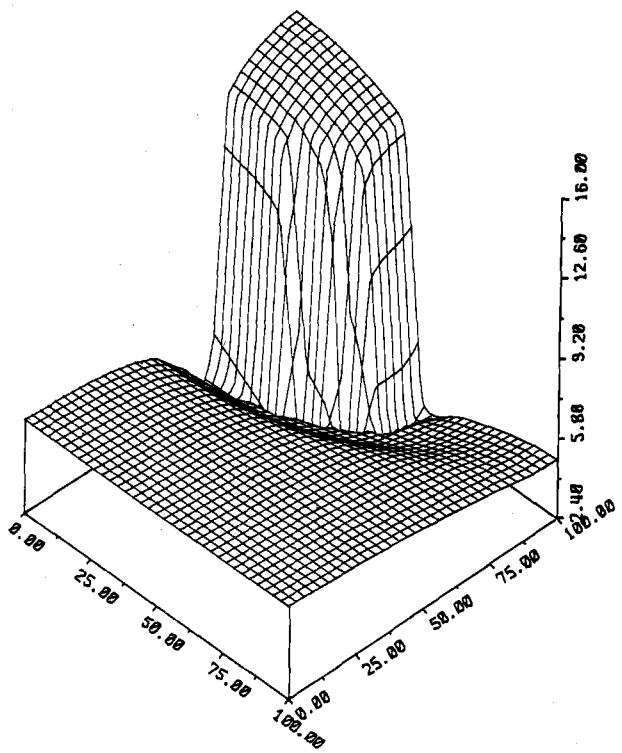

Fig. 9. Hole concentration p of diode for a bias of 1 V.

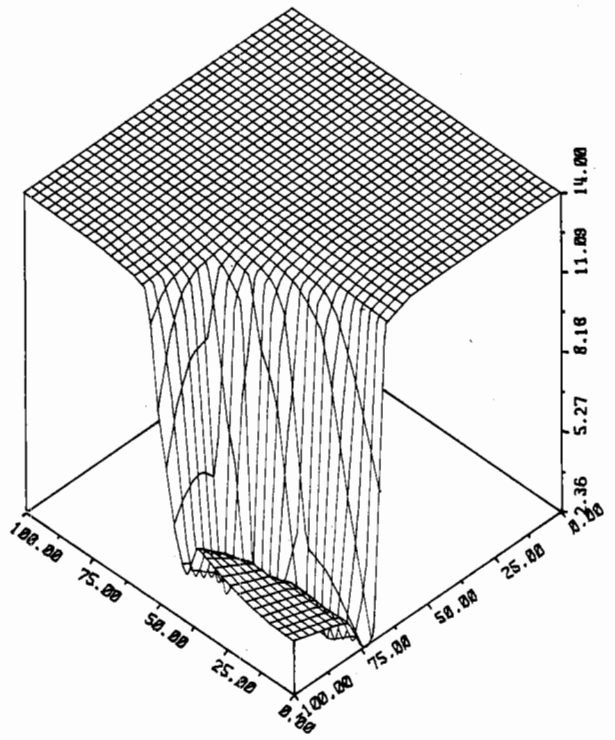

Fig. 10. Electron concentration n of diode for a bias of 1 V.

Fig. 9 where the hole carrier density is drawn. The doping profile is Gaussian in the p-region and constant in the bulk. Not surprisingly the same behavior is found in the plots of carrier densities.

VI. Conclusion

The semiconductor equations have been solved using a finite-element method in its residual formulation. An adaptive mesh-refinement process has been developed which generates two different kinds of elements (triangles and rectangles), thus enabling almost arbitrary subdivision and expansion of grid spacings. This guarantees that only the minimum number of nodes is created.

Solving the Dirichlet problem for one element under certain assumptions led to the development of shape functions for the exponentials of the quasi-Fermi potentials which are bilinear in Bernoulli-like functions. In elements where the carrier concentrations vary exponentially, the discretization error is zero. These functions degenerate into the ordinary bilinear shape functions when the electric potential is constant and into very steep functions when the electric potential is large. The influence of the electric field on the shape functions has been illustrated and discussed. The outlined derivation of the shape functions only holds for rectangular elements. Only a regular rectangular mesh would, therefore, exhibit the discussed features. Since an optimal mesh is not purely rectangular, conditions for a transition element had to be formulated. The generalization to the transition macroelement is achieved by taking the current density in each of the subdividing triangles to be constant. The electric potential ψ itself has consistently been approximated with linear/bilinear functions throughout the device.

This concept of variable and ψ-dependent shape functions led to a set of highly nonlinear equations. As weighting functions, the linear/bilinear basis functions of the electric potential were used. The residuum integration was carried out numerically, and the resulting nonlinear equations were solved using a modified Newton method. The Jacobian matrix was approximated by numerical differentiation. An element-wise bypassing method for the calculation of the entries in the Jacobian matrix has been used to minimize computer requirements and actually enhance throughput. The Newton iterations, necessary for the solution of the nonlinear system, were modified. Specific examples have demonstrated the flexibility and accuracy of the outlined procedure.

Acknowledgment

The authors would like to thank Prof. Dr. H. Pötzl and Dr. P. Markowich for helpful discussions and for critically reading the manuscript. The authors would also like to thank the Interuniversitäre Rechenzentrum for the generous amount of computer resources it provided.

References

[1] E. M. Buturla *et al.*, "Finite element analysis of semiconductor devices: The FIELDAY program," *IBM J. Res. Develop.*, vol. 25, no. 4, pp. 218–231, 1981.

[2] G. D. Hachtel *et al.*, "Semiconductor analysis using finite elements —Part 1: Computational aspects," *IBM J. Res. Develop.*, vol. 25, no. 4, pp. 232–245, 1981.

[3] T. Adachi *et al.*, "Two-dimensional semiconductor analysis using finite-element method," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 1026–1031, 1979.

[4] I. Babuska and W. C. Rheinboldt, "Error estimates for adaptive finite element computations," *J. Inst. Math. Its Appl.*, vol. 13, pp. 121–126, 1974.

[5] 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.

[6] 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.

[7] O. C. Zienkiewicz, *The Finite Element Method.* New York: McGraw-Hill, 1977.

[8] S. E. Laux and R. J. Lomax, "Numerical investigation of mesh size convergence rate of the finite element method in MESFET simulation," *Solid-State Electron.*, vol. 24, pp. 485–493, 1981.

[9] W. C. Rheinboldt, "On a theory of mesh-refinement processes," *SIAM J. Numer. Anal.*, vol. 17, pp. 766–788, 1980.

[10] F. Rahali, "Analyse numerique a 2 dimensions de transistors MOS par la methode des elements finis," EPF-Lausanne Lab. d'Electronique Generale Lausanne, Switzerland, Rep., 1982.

[11] L. Mansfield, "On the solution of nonlinear finite element systems," *SIAM J. Numer. Anal.*, vol. 17, pp. 752–765, 1980.

[12] A. J. Davies, "The finite element method: A first approach" in *Oxford Applied Mathematics and Computing Science Series.* Oxford: Oxford University Press, 1980.

[13] J. T. Oden and J. N. Reddy, *An Introduction to the Mathematical Theory of Finite Elements.* New York: Wiley, 1976.

[14] J. J. Barnes *et al.*, "Finite element simulation of GaAs MESFET's with lateral doping profiles and submicron gates," *IEEE Trans. Electron Devices*, vol. ED-23, pp. 1042–1048, 1976.

[15] P. E. Cottrell and E. M. Buturla, "Two dimensional static and transient simulation of mobile carrier transport in a semiconductor," in *Proc. Nasecode I Conf.*, pp. 31–64, 1979.

[16] H. R. Schwarz, *Methode der finiten Elemente.* Stuttgart: B. G. Teubner, 1980.

[17] Curtis *et al.*, "On the estimation of sparse Jacobian matrices," *J. Inst. Math. Its Appl.*, vol. 13, pp. 117–119, 1974.

[18] L. W. Johnson and R. D., Riess, "An error analysis for numerical differentiation," *J. Inst. Math. Its Appl.*, vol. 11, pp. 115–120, 1973.
[19] N. E. Gibbs *et al.*, "An algorithm for reducing the bandwidth and profile of a sparse matrix," *SIAM J. Numer. Anal.*, vol. 13, pp. 236–250, 1976.
[20] J. G. Lewis, "Implementation of the Gibbs–Poole–Stockmeyer and Gibbs–King algorithms," *ACM Trans. Math. Software*, vol. 8, pp. 180–189, 1982.
[21] J. G. Lewis, "Algorithm 582: The Gibbs–Poole–Stockmeyer and Gibbs–King algorithms for reordering sparse matrices," *ACM Trans. Math. Software*, vol. 8, pp. 190–194, 1982.
[22] Wai-Hang Liu and A. H. Sherman, "Comparative analysis of the Cuthill–McKee and the reverse Cuthill–McKee ordering algorithms for sparse matrices," *SIAM J. Numer. Anal.*, vol. 13, pp. 198–213, 1976.
[23] A. H. Sherman, "Algorithm 533: NSPIV, A Fortran subroutine for sparse Gaussian elimination with partial pivoting," *ACM Trans. Math. Software*, vol. 4, pp. 391–398, 1978.
[24] E. P. Doolan *et al.*, *Uniform Numerical Methods for Problems with Initial and Boundary Layers*. Dublin: Boole Press, 1980.
[25] 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.
[26] S. Selberherr, "Two dimensional MOS transistor modeling," in *Proc. NATO Advanced Study Inst. Conf. Process Device Simulation MOS-VLSI Circ.*, 1982.
[27] S. Selberherr and P. Markowich, "Automatic step size control for numerical differentiation," to be published.
[28] P. A. Markowich *et al.*, "A singularly perturbed boundary value problem modeling a semiconductor device," Univ. of Wisconsin, MRC Tech. Summary Rep. 2388, 1982.
[29] G. Strang and J. G. Fix, *An Analysis of the Finite Element Method*. Englewood Cliffs, NJ, Prentice–Hall, 1973.
[30] A. Franz and G. Franz, private communication.
[31] S. M. Sze, *Physics of Semiconductor Devices*. New York, Wiley, 1969.

Numerical Solution of the Semiconductor Transport Equations with Current Boundary Conditions

BERTRAND M. GROSSMAN AND MICHAEL J. HARGROVE

Abstract — **The semiconductor transport equations are solved by a hybrid finite-element method with current specified as a boundary condition at device contacts. Single carrier or bipolar devices of arbitrary shape, operating under transient or steady-state conditions, can be simulated with current sources or simple circuit elements connected to device terminals. This paper describes the numerical technique and device applications.**

INTRODUCTION

THE transport of electrons and holes in a semiconductor device is simulated by the FIELDAY program [1] with current specified as a boundary condition at device contacts. FIELDAY is an IBM general purpose program used extensively for designing semiconductor devices. The program solves the semiconductor transport equations in one, two or three dimensions, under transient or steady-state conditions, by the finite-element method. Devices of arbitrary shape and doping profile with a variety of different material properties can be modeled. The program's ability to solve the transport equations with current as a boundary condition allows for the simulation of a device within a circuit environment of current sources, and resistive and capacitive loads. This paper specifically focuses on the numerical treatment of current as a boundary condition at device contacts. Two device applications illustrate the use of the model.

Manuscript received November 5, 1982; revised February 10, 1983.

B. M. Grossman is with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598.
M. J. Hargrove is with IBM General Technology Division, Essex Junction, VT 05452.

FORMULATION

Equations (1)—(5) describe the flow of electrons and holes in a semiconductor [2], [3]

$$\nabla^2\psi = \frac{-q}{\epsilon}\left(p - n + N_D - N_A + N_Q\right) \tag{1}$$

$$\nabla \cdot J_n = q\left(\frac{\partial n}{\partial t} + R_n\right) \tag{2}$$

$$\nabla \cdot J_p = -q\left(\frac{\partial p}{\partial t} + R_p\right) \tag{3}$$

$$J_n = q\left[D_n \nabla n - \mu_n n \nabla(\psi + \Delta V_C)\right] \tag{4}$$

$$J_p = -q\left[D_p \nabla p + \mu_p p \nabla(\psi - \Delta V_V)\right]. \tag{5}$$

In these equations, the three unknown quantities are the space-charge potential (ψ), and the electron (n) and hole (p) densities. N_D and N_A are the ionized donor and acceptor impurity densities. N_Q is the density of fixed charged

0018-9383/83/0900-1092$01.00 ©1983 IEEE