SIAM J. SCI. STAT. COMPUT.
Vol. 13, No. 2, pp. 512-530, March 1992

005

# NONLINEAR MULTIGRID APPLIED TO A ONE-DIMENSIONAL STATIONARY SEMICONDUCTOR MODEL*

P. M. DE ZEEUW†

**Abstract.** The nonlinear multigrid method is applied to a transistor problem in one dimension. A weak spot in the linearization of the well-known Scharfetter-Gummel discretization scheme is reported. Further, it is shown that both the residual transfer and the solution transfer from a fine to a coarse grid need special requirements due to the rapidly varying problem coefficients. Some modifications are proposed which make the multigrid algorithm perform well for the hard example problem.

**Key words.** semiconductor equations, multigrid methods

**AMS(MOS) subject classifications.** 65N20, 65H10

**1. Introduction.** There is a great demand for a proper numerical simulation of semiconductors in order to reduce the costs of constructing expensive prototypes. The search for a fast and robust algorithm has proven to be a challenge. So far only a few papers have considered the multigrid solution of the discrete semiconductor equations (e.g., see [1], [2], [6], [9], [13]) and therefore extensive further research is required.

In this paper we restrict ourselves on purpose to one space dimension as a preparatory study for the case of more space dimensions. We study a particular example problem which has been put forward by Schilders (Philips, the Netherlands). This problem models a transistor and turns out to be much harder to solve than the forward or reversed biased diode problem. We apply the nonlinear multigrid method and encounter a serious difficulty due to the nonlinearity of the problem. Some modifications are proposed which significantly increase the robustness of the nonlinear multigrid method and which look promising also for the higher-dimensional case.

**2. The problem.** The behavior of a steady semiconductor device can be described by the following set of equations (cf. [10]):

$$\nabla(-\varepsilon\nabla\psi) = q(p - n + D), \tag{2.1a}$$

$$\nabla J_n = +qR, \tag{2.1b}$$

$$\nabla J_p = -qR, \tag{2.1c}$$

where $J_n$ and $J_p$ are defined by

$$J_n = q\mu_n\left(\frac{1}{\alpha}\nabla n - n\nabla\psi\right), \tag{2.2a}$$

$$J_p = -q\mu_p\left(\frac{1}{\alpha}\nabla p + p\nabla\psi\right). \tag{2.2b}$$

Substitution of (2.2) into (2.1) results in a system of three nonlinear partial differential equations for $\psi$, $n$, and $p$. In (2.1) $\psi$ represents the electrostatic potential, $p$ and $n$ describe the concentration of holes and electrons, respectively. Equations (2.1b) and (2.1c) are called the continuity equations, $J_n$ is the electron current density, $J_p$ is the hole current density, and $R$ is the recombination-generation rate, a function of $n$ and $p$. The doping profile $D$ is a function of the space variable $x$. The quantities $\varepsilon$, $q$, $\alpha$,

* Received by the editors June 21, 1989; accepted for publication (in revised form) November 14, 1990.
† Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, the Netherlands.

$\mu_n$, $\mu_p$ represent the permittivity, the elementary charge, the inverse of the thermal voltage, and the electron and hole mobility, respectively.

In this paper we consider the case of only one space dimension and assume $\varepsilon$, $\alpha$, $\mu_n$ and $\mu_p$ to be constant. It is common practice to replace the variables $n$ and $p$ by the hole and electron quasi-Fermi potentials $\phi_n$ and $\phi_p$ defined by the relations

$$n = n_i e^{\alpha(\psi-\phi_n)}, \tag{2.3a}$$

$$p = n_i e^{\alpha(\phi_p-\psi)}. \tag{2.3b}$$

On the one hand, by this change of variables, the nonlinearity of the problem is strongly increased, on the other hand the values assumed by $(\psi, \phi_n, \phi_p)$ are in a much more moderate range. For extensive discussions on the choice of variables see [9], [10]. Using (2.3) the equations (2.1) are transformed into

$$-\nabla J_\psi = n_i q(e^{\alpha(\phi_p-\psi)} - e^{\alpha(\psi-\phi_n)}) + qD, \tag{2.4a}$$

$$-\nabla J_n = +qR, \tag{2.4b}$$

$$-\nabla J_p = -qR, \tag{2.4c}$$

where $J_\psi$ is defined by

$$J_\psi = \varepsilon \nabla \psi, \tag{2.5a}$$

and $J_n$, $J_p$ are now defined by

$$J_n = \bar{\mu}_n e^{\alpha(\psi-\phi_n)} \nabla(\alpha\phi_n), \tag{2.5b}$$

$$J_p = \bar{\mu}_p e^{\alpha(\phi_p-\psi)} \nabla(\alpha\phi_p), \tag{2.5c}$$

with

$$\bar{\mu}_n = \frac{n_i q \mu_n}{\alpha}, \qquad \bar{\mu}_p = \frac{n_i q \mu_p}{\alpha}. \tag{2.5d}$$

In this paper we adhere to the formulation (2.4)-(2.5).

**2.1. A particular one-dimensional model problem.** We will focus our attention to a particular (hard) one-dimensional model problem which has been supplied by Schilders [14]. Here the problem constants are

$$\begin{aligned} &\varepsilon = 1.035918_{10}^{-12}\ \text{As V}^{-1}\ \text{cm}^{-1}, \qquad q = 1.6021_{10}^{-19}\ \text{As}, \\ &\mu_n = \mu_p = 500\ \text{V}^{-1}\ \text{s}^{-1}\ \text{cm}^2, \qquad n_i = 1.22_{10}{}^{10}\ \text{cm}^{-3}, \\ &k = 1.38054_{10}^{-23}\ \text{V As K}^{-1}, \qquad T = 300\ \text{K}, \qquad \alpha = q/kT. \end{aligned} \tag{2.6}$$

The function $R$ is given by

$$R = \frac{pn - n_i^2}{\tau(p+n+2n_i)}, \qquad \tau = 10^{-6}\ s.$$

The doping function $D$ (in $\text{cm}^{-3}$) is given by

$$\begin{aligned} D(x) = 6_{10}{}^{15} + 6_{10}{}^{19} \exp(-(x/7.1_{10}^{-5})^2) - 2.15_{10}{}^{18} \exp(-(x/1.15_{10}^{-4})^2) \\ + 1.1_{10}{}^{19} \exp(-((x-8_{10}^{-4})/1.3_{10}^{-4})^2). \end{aligned}$$

The equations (2.4) are defined on the domain $\Omega = [0, 8_{10}^{-4}](\text{cm})$. We have three contacts to our semiconductor device (the one-dimensional model of a transistor): the emitter ($E$), the basis ($B$) and the collector ($C$) (see Fig. 1).

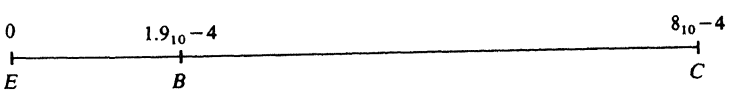


FIG. 1. *The contacts in the one-dimensional transistor problem.*

In Fig. 2 the doping function $D(x)$ is shown after the transformation $D \to sign(D)^{10} \log(1+|D|)$. Boundary conditions at the emitter $E$ are:

(2.7a) $p - n + D = 0$ (i.e., vanishing space charge),

(2.7b) $\phi_n = V_E$,

(2.7c) $J_p = 0$.

Boundary conditions at the basis $B$:

(2.8) $\phi_p = V_B = 0$.

Boundary conditions at the collector $C$:

(2.9a) $p - n + D = 0$,

(2.9b) $\phi_n = V_C$,

(2.9c) $\phi_p = V_C$.

For fifteen different cases, each characterized by a pair of voltages $(V_E, V_C)$, the solution is required (see Table 4.1). Figure 8 shows the solution-component $\psi$ for the subsequent cases.

**3. Discretization.** At the outset of this section we give a short preview of its contents.

In order to abide by the law of conservation we use a finite volume technique based on the piecewise constant approximation of $J_\psi$, $J_n$, and $J_p$. As a consequence we arrive at a cell-centered version of the well-known Scharfetter-Gummel scheme [4], [9], [11]. We examine how the nonlinear discrete operator depends on the discrete solution.

**3.1. Box integration.** The interval $\Omega = (x_0, x_N)$ is split up into disjoint *boxes* $B_j = (x_{j-1}, x_j)$, $j = 1(1)N$. A point $x_j$ is called a *wall*, a point $x_{j-1/2} = (x_{j-1} + x_j)/2$ is called a *center*. The basis $B$ is at the partition-wall between two boxes. Another set of

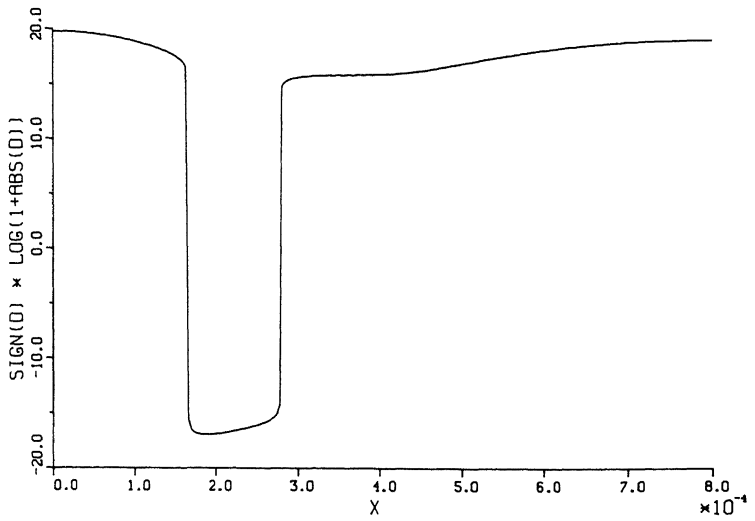


FIG. 2. *The doping profile.*

subintervals $\{D_j\}$ is defined by

$$D_0 = (x_0, x_{1/2}),$$
$$D_j = (x_{j-1/2}, x_{j+1/2}), \quad j = 1\,(1)\,N-1,$$
$$D_N = (x_{N-1/2}, x_N).$$

This set is called the set of *dual boxes.*

Now, by applying the Gauss divergence theorem in one dimension to (2.4) on the domains $B_j$ we find

$$\begin{aligned} &-J_\psi|_{x_{j-1}}^{x_j} - n_i q \int_{B_j} (e^{\alpha(\phi_p-\psi)} - e^{\alpha(\psi-\phi_n)})\, d\Omega = q \int_{B_j} D\, d\Omega, \\ &-J_n|_{x_{j-1}}^{x_j} - q \int_{B_j} R\, d\Omega = 0, \\ &-J_p|_{x_{j-1}}^{x_j} + q \int_{B_j} R\, d\Omega = 0, \quad j = 1\,(1)\,N. \end{aligned} \tag{3.1}$$

We can write (3.1) in symbolic form as

$$\mathcal{M}(\boldsymbol{q}) = f \tag{3.2}$$

where $\boldsymbol{q}$ denotes the vector $(\psi, \phi_n, \phi_p)^T$, $\mathcal{M}$ the nonlinear operator in the left-hand side of (3.1) and $f$ the right-hand side of (3.1).

**3.2. Box discretization.** We introduce the variables $(\psi_j, \phi_{n,j}, \phi_{p,j})^T$, $j = 1\,(1)\,N$, which are associated with the centers $x_{j-1/2}$ of the boxes $B_j$. Let $\approx$ denote approximation by midpoint quadrature. We then define

$$\begin{aligned} S_j &\approx n_i q \int_{B_j} (\exp(\alpha(\phi_p - \psi)) - \exp(\alpha(\psi - \phi_n)))\, d\Omega, \\ F_j &\approx q \int_{B_j} D\, d\Omega, \\ R_j &\approx q \int_{B_j} R\, d\Omega, \\ j &= 1\,(1)\,N. \end{aligned} \tag{3.3}$$

We make the assumption that $J_\psi$, $J_n$ and $J_p$ are piecewise constant on the dual set $\{D_j\}$ (see [4], [9], [11]) and correspondingly we use the notation $J_{\psi,j}$, $J_{n,j}$, $J_{p,j}$. By this assumption and applying (3.3) we arrive at the following discrete equations:

$$-J_{\psi,j} + J_{\psi,j-1} - S_j = F_j, \tag{3.4a}$$

$$-J_{n,j} + J_{n,j-1} - R_j = 0, \tag{3.4b}$$

$$-J_{p,j} + J_{p,j-1} + R_j = 0, \tag{3.4c}$$

with

$$J_{\psi,j} = \varepsilon \frac{\psi_{j+1} - \psi_j}{x_{j+1/2} - x_{j-1/2}}, \tag{3.5a}$$

$$J_{n,j} = \bar{\mu}_n \frac{\exp(-\alpha\phi_{n,j+1}) - \exp(-\alpha\phi_{n,j})}{\exp(-\alpha\psi_{j+1}) - \exp(-\alpha\psi_j)} \cdot \frac{\alpha\psi_{j+1} - \alpha\psi_j}{x_{j+1/2} - x_{j-1/2}}, \tag{3.5b}$$

$$J_{p,j} = \bar{\mu}_p \frac{\exp(\alpha\phi_{p,j+1}) - \exp(\alpha\phi_{p,j})}{\exp(\alpha\psi_{j+1}) - \exp(\alpha\psi_j)} \cdot \frac{\alpha\psi_{j+1} - \alpha\psi_j}{x_{j+1/2} - x_{j-1/2}}. \tag{3.5c}$$

At the emitter $E$, basis $B$, and collector $C$ similar equations are obtained; for full details see [16]. Thus we have obtained the cell-centered version of the well-known Scharfetter-Gummel scheme (see [4], [9], [11]). Summarizing, we have discretized (2.4), together with the boundary conditions (2.7)-(2.9), into a set of $3N$ nonlinear equations (3.4) with the $3N$ variables $\psi_j$, $\phi_{n,j}$, $\phi_{p,j}$; $j=1\ (1)N$. We can write (3.4) in symbolic form as

$$\mathcal{M}_h(q_h)=f_h \tag{3.6}$$

where $\mathcal{M}_h$ denotes the nonlinear difference operator and $f_h$ the right-hand side.

**3.3. Properties of the discretized operator.** In this subsection we study how the Jacobian of the nonlinear discrete operator $\mathcal{M}_h$ depends on the discrete solution. We assume the recombination term to be zero and confine ourselves to the dependency on $\phi_p$. Results for $\phi_n$ can be derived analogously. We freeze the solution components $\psi$ and $\phi_n$ and consider the $\phi_p$-stencil, at box $B_j$, defined by the triplet

$$[stp(j,-1),\ stp(j,0),\ stp(j,+1)] \tag{3.7a}$$

with

$$stp(j,k)=\frac{\partial(-J_{p,j}+J_{p,j-1})}{\partial\phi_{p,j+k}}, \qquad k=-1,0,1. \tag{3.7b}$$

We introduce the notation

$$\Delta_j x \equiv x_{j+1/2}-x_{j-1/2},$$

$$\Delta_j \psi \equiv \psi_{j+1}-\psi_j$$

and the function $s(z):\mathbb{R}\to\mathbb{R}$ by

$$s(z)\equiv\frac{z}{e^z-1}. \tag{3.8}$$

By straightforward computation it can be verified that the following equalities hold:

$$stp(j,-1)=-\alpha\bar{\mu}_p \exp(\alpha(\phi_{p,j-1}-\psi_j))\frac{s(-\alpha\Delta_{j-1}\psi)}{\Delta_{j-1}x}, \tag{3.9a}$$

$$stp(j,0)=\alpha\bar{\mu}_p \exp(\alpha(\phi_{p,j}-\psi_j))\left\{\frac{s(-\alpha\Delta_{j-1}\psi)}{\Delta_{j-1}x}+\frac{s(\alpha\Delta_j\psi)}{\Delta_j x}\right\}, \tag{3.9b}$$

$$stp(j,+1)=-\alpha\bar{\mu}_p \exp(\alpha(\phi_{p,j+1}-\psi_j))\frac{s(\alpha\Delta_j\psi)}{\Delta_j x}, \tag{3.9c}$$

and

$$stp(j,0)=-(stp(j,-1)+stp(j,+1))+\alpha(-J_{p,j}+J_{p,j-1}). \tag{3.9d}$$

Because $s(z)>0$ for all $z$, it follows that

$$stp(j,-1)<0, \qquad stp(j,0)>0, \qquad stp(j,+1)<0, \tag{3.10}$$

so the $\phi_p$-stencils correspond with an $\mathcal{L}$-matrix. Further, at the exact discrete solution, i.e., when $-J_{p,j}+J_{p,j-1}=0$ is satisfied, it follows from (3.9d) that

$$stp(j,0)=-(stp(j,-1)+stp(j,+1)),$$

so then the $\mathscr{L}$-matrix also possesses weak diagonal dominance (provided there is at least one stencil corresponding with a Dirichlet boundary condition, see [15]). However, in the middle of some iterative process to determine the solution, we may well have negative residuals so that (3.9d) implies the loss of diagonal dominance. Therefore ill-conditioning and numerical difficulties can be expected.

**4. The Newton method and expedients.** An obvious way of solving the set of nonlinear equations (3.6) is application of the Newton method. Because the Newton method is not globally convergent and the operator $\mathscr{M}_h$ is strongly nonlinear in the variables $(\psi, \phi_n, \phi_p)$ we use two additional tools which are considered subsequently in this section:

(1) Correction transformation.

(2) Smoothing of the Newton-iterates.

It turns out that these expedients make the Newton method well applicable. Other modifications of the Newton method including inexact line searches and related techniques have been found to be reliable elsewhere (see [4]).

In two or more space dimensions direct application of the Newton method to (3.6) would involve large storage requirements and the solution of large linear systems. If well designed, a nonlinear multigrid algorithm holds out a prospect of both a computational complexity which is linear in the number of gridpoints and low storage requirements even for the case of two or more space dimensions. Therefore we want to apply the Newton method only for very coarse grids and we restrict the use of the Newton method as a *coarsest grid solver* for multigrid methods (§ 5).

**4.1. Correction transformation.** The correction transformation introduced by Schilders [10] is a device to transform the Newton-correction $(d\psi, d\phi_n, d\phi_p)$, computed by linearisation with respect to $(\psi, \phi_n, \phi_p)$, into the correction for these very variables that would be obtained if linearisation were applied with respect to $(\psi, n, p)$. Because the system in terms of $(\psi, n, p)$ is much less nonlinear, a much better convergence behavior of the Newton method can be expected. By performing the calculations in terms of $(\psi, \phi_n, \phi_p)$ and applying a transformation afterwards, we avoid complications due to the extremely wide range of values of $n$ and $p$. In this way we take advantage of the benefits of both variable sets [9], [10], [14].

**4.2. Smoothing.** In § 4.1 we pointed out a technique to improve the global convergence behavior of the Newton method. Even yet difficulties are encountered when we apply the improved Newton method. As an example consider Fig. 3 which shows subsequent Newton iterates for case 12 starting from the solution for case 11.

The dips in the iterates are attended with very small pivot numbers while solving the linear systems. Section 3.3 explains the ill-conditioning whenever there is a large residual somewhere. Artificially increasing the main diagonal of the Jacobian turned out to be not efficient. Simply cutting off the correction at certain points is hardly justifiable because of lack of a more or less general criterion to do so. A more appropriate way of handling the phenomenon sketched above is to apply relaxation or smoothing sweeps at the beginning of the Newton process [9]. As a smoother the collective symmetric Gauss–Seidel relaxation (CSGS) can be used. It is called collective because at each box we solve collectively the three nonlinear equations which arise (employing Newton's method).

We will present here some numerical results to show the effect of smoothing. The grid is more or less uniform and satisfies $x_{N/4} = B$. The set of voltages $\{(V_E, V_C)\}$ for which a solution is required is defined in Table 4.1. For each case $> 0$ the solution of the previous case serves as a starting solution; in case 0 we start with $\phi_{n,j} = \phi_{p,j} = 0$,

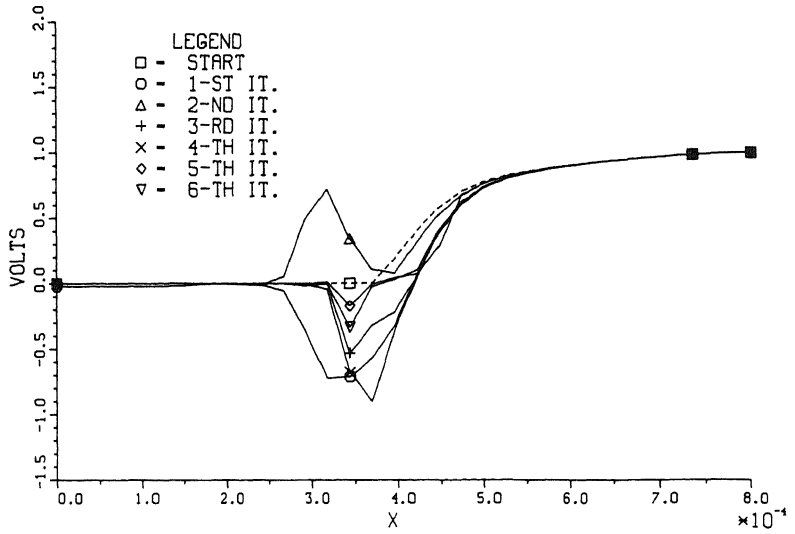


FIG. 3. *Subsequent Newton iterates of* $\phi_p$.

TABLE 4.1
*Subsequent voltages at the emitter and collector for which a solution is required.*

| Case | $V_E$ | $V_C$ |
|---|---|---|
| 0 | 0 | 0.0 |
| 1 | 0 | 0.2 |
| 2 | 0 | 0.4 |
| 3 | 0 | 0.6 |
| 4 | 0 | 0.8 |
| 5 | 0 | 1 |
| 6 | −0.2 | 1 |
| 7 | −0.4 | 1 |
| 8 | −0.6 | 1 |
| 9 | −0.7 | 1 |
| 10 | −0.8 | 1 |
| 11 | −0.85 | 1 |
| 12 | −0.9 | 1 |
| 13 | −0.95 | 1 |
| 14 | −1 | 1 |

for all $j$, and $\psi$ is determined by assuming space charge neutrality. We use the correction transformation. For the solution of the linear systems we apply rowscaling followed by rowpivotting. Table 4.2 shows the number of Newton sweeps required to reach a correction with absnorm $< 10^{-12}$, and the smallest pivot number encountered during the solution process. Table 4.2 also contains the results for the case when in addition a CSGS sweep is applied each time after a Newton-correction for which the infinity norm of the correction was larger than 0.1. This method will henceforth be referred to as Newton-CSGS. We observe that in the difficult cases 11–14 the application of smoothing sweeps has a positive effect on the efficiency and robustness of the Newton method. When smoothing is applied the smallest pivot numbers encountered keep a substantial distance from zero which shows that then the Jacobians generated within

TABLE 4.2
*Number of Newton and* CSGS *sweeps used, and smallest pivot numbers*; $N = 32$.

| | No smoothing applied | | Smoothing applied | | |
|---|---|---|---|---|---|
| Case | Newton sweeps | Smallest pivot number | Newton sweeps | CSGS sweeps | Smallest pivot number |
| 0 | 6 | $3_{10}-3$ | 4 | 1 | $5_{10}-1$ |
| 1 | 5 | $2_{10}-4$ | 5 | 2 | $5_{10}-1$ |
| 2 | 6 | $1_{10}-9$ | 5 | 2 | $5_{10}-1$ |
| 3 | 5 | $2_{10}-7$ | 5 | 2 | $5_{10}-1$ |
| 4 | 5 | $1_{10}-7$ | 5 | 2 | $3_{10}-1$ |
| 5 | 5 | $4_{10}-5$ | 5 | 2 | $3_{10}-1$ |
| 6 | 5 | $3_{10}-3$ | 5 | 2 | $2_{10}-1$ |
| 7 | 5 | $3_{10}-3$ | 5 | 2 | $2_{10}-1$ |
| 8 | 5 | $2_{10}-4$ | 5 | 2 | $2_{10}-1$ |
| 9 | 5 | $4_{10}-7$ | 5 | 2 | $3_{10}-1$ |
| 10 | 6 | $1_{10}-7$ | 6 | 3 | $3_{10}-1$ |
| 11 | 9 | $4_{10}-9$ | 7 | 3 | $2_{10}-1$ |
| 12 | 15 | $5_{10}-13$ | 8 | 4 | $4_{10}-2$ |
| 13 | 13 | $2_{10}-11$ | 7 | 3 | $1_{10}-1$ |
| 14 | 10 | $5_{10}-10$ | 7 | 3 | $1_{10}-1$ |

the Newton method are far from being singular (e.g., compare to case 12 in Table 4.2) and therefore no large dips in the Newton-corrections do occur. Experiments for $N = 16, 64, 128$ show results similar to Table 4.2.

**5. The multigrid method.** More advanced ways of solving a set of nonlinear equations are the full approximation scheme (FAS) [5], and the nonlinear multigrid method (NMGM) [7]. Both *multigrid methods* are very similar although the NMGM is more general. The multigrid method has already found many specific applications in the fields of elliptic, parabolic, and hyperbolic equations and integral equations as well. Recently, also in the field of semiconductor equations research on multigrid methods has been initiated ([1], [2], [6], [9], [13]). If well applied, a multigrid method can be optimal in the sense that the rate of convergence is independent of the meshsize. An important advantage of the FAS/NMGM method is that no large linear systems need to be stored and solved. The subsequent stages of a usual FAS method, applied to (3.6), are

(1) Apply $p$ nonlinear relaxation sweeps; thus we get an approximation $q_h$ of the solution which has a *smooth* residual $d_h \equiv f_h - \mathcal{M}_h(q_h)$.

(2) Transfer $q_h$ and $d_h$ from $\Omega_h$ to a coarser grid $\Omega_H$ by means of the respective restriction operators $R_H$ and $\bar{R}_H$.

(3) Solve (approximately) on $\Omega_H$ the equation $\mathcal{M}_H(q_H) = \mathcal{M}_H(R_H q_h) + \bar{R}_H d_h$.

(4) Interpolate the correction, computed on $\Omega_H$, onto $\Omega_h$ and add the correction to $q_h$.

(5) Apply $q$ nonlinear relaxation sweeps.

The combination of stages 2, 3, and 4 is called the coarse grid correction (CGC). Stage 3 may be obtained by applying a number of $\sigma$ FAS cycles on the coarser grid. In this way a recursive procedure is obtained in which a sequence of increasingly coarser grids is used. In this paper we use $p = q = \sigma = 1$ throughout. In the subsections to come we will define precisely the coarse grid correction and the grid transfer operators

involved. In § 6 a significant improvement of the CGC will be introduced. It consists of a solution-dependent adjustment of the restriction of the residual $d_h$.

**5.1. Nested boxes.** Let a coarse grid $\Omega_H$, a discretization of $\Omega$, be given by the set of boxes $\{B_{H,j}\}_{j=1(1)N}$. From $\Omega_H$ we construct the next finer grid $\Omega_h=\{B_{h,j}\}_{j=1(1)2N}$ by division of each $B_{H,j}$ into two disjoint boxes $B_{h,2j-1}$ and $B_{h,2j}$. By repetition we obtain thus a sequence of increasingly finer grids. By definition all boxes are nested. Of course, the corresponding dual boxes are not nested. For all our numerical experiments in this paper we assume in addition that $B_{h,2j-1}$ and $B_{h,2j}$ have equal size.

**5.2. Restriction operators.** For the problem (3.2) on $\Omega$, let $S$ denote the domain and $V$ the range of nonlinear operator $\mathcal{M}$. For each discretization on $\Omega_h$, we have the spaces $S_h$ and $V_h$, the discrete analogues of $S$ and $V$.

Let the restriction operator for right-hand side functions

$$\bar{R}_h : V \to V_h \tag{5.1a}$$

be defined by

$$\bar{R}_h f = f_h, \tag{5.1b}$$

$$f_{h,j} = \int_{B_{h,j}} f \, d\Omega, \qquad \forall j \text{ at } \Omega_h. \tag{5.1c}$$

It follows for the next coarser grid that

$$(\bar{R}_H f)_j = (\bar{R}_h f)_{2j-1} + (\bar{R}_h f)_{2j}, \qquad \forall j \text{ at } \Omega_H. \tag{5.2}$$

By this equality, $\bar{R}_H$ can be defined also on $V_h$:

$$\bar{R}_H : V_h \to V_H, \tag{5.3a}$$

$$(\bar{R}_H f_h)_j = f_{h,2j-1} + f_{h,2j}, \qquad \forall j \text{ at } \Omega_H. \tag{5.3b}$$

The restriction operator for solutions

$$R_h : S \to S_h \tag{5.4}$$

may be defined by the well-known full weighting operator [5].

**5.3. Prolongation/interpolation.** A prolongation transfers a solution from a coarse grid to a finer one:

$$P_h : S_H \to S_h. \tag{5.5}$$

A common and simple choice for the prolongation should be linear interpolation. However, two objections against this choice do arise. Firstly, by the use of linear interpolation it is implicitly assumed that the solution behaves like a smooth function on $\Omega_H$. Because of the exponential behavior of the solution in some areas, this is only true on an unfeasibly fine grid. Secondly, linear interpolation does not satisfy here the so-called Galerkin condition

$$\bar{R}_H \mathcal{M}_h(P_h s_H) = \mathcal{M}_H(s_H), \tag{5.6}$$

which is a condition that ascertains the reduction of low frequency components in the residual after a CGC. Hemker [9] has introduced a prolongation which is based on the assumption of smoothness of fluxes, and which satisfies (5.6) for the simplified case that all $S_j$ and $R_j$ are zero, see (3.4). Here, we use the same assumption but we choose a short and convenient formulation in order to handle also the situation near the inner boundary point $B$. Figure 4 depicts how the dual box $[L, R]$ is divided into the boxes $[L, M]$ and $[M, R]$.

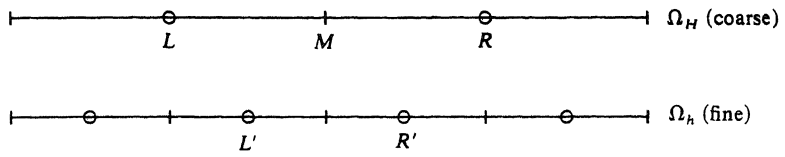


FIG. 4. *Staggering of a coarse and fine grid.*

The assumption reads that $J_\psi$, $J_n$, $J_p$ are constant on $[L, R]$. Given the values of the variables $(\psi, \phi_n, \phi_p)$ at $L$ and $R$ we wish to compute the values at $L'$ and $R'$. From (3.5a) it follows that $\psi|_{L'}$ and $\psi|_{R'}$ can be computed by linear interpolation. For $\phi_p$ we first determine the value at the wall $M$. If we write

$$\Delta\psi = \psi|_R - \psi|_L, \qquad \Delta\phi_p = \phi_p|_R - \phi_p|_L \tag{5.7}$$

then we derive that

$$\begin{aligned}
&\phi_p|_M = \\
&\quad \textbf{if } \Delta\psi < 0 \\
&\quad \textbf{then if } \frac{\Delta\psi}{2} - \Delta\phi_p < 0 \\
&\qquad \textbf{then } \phi_p|_R + z\left(\frac{\Delta\psi}{2}, \frac{\Delta\psi}{2} - \Delta\phi_p\right) \\
&\qquad \textbf{else } \phi_p|_L + \frac{\Delta\psi}{2} + z\left(\frac{\Delta\psi}{2}, -\frac{\Delta\psi}{2} + \Delta\phi_p\right) \\
&\quad \textbf{end if} \\
&\quad \textbf{else if } \frac{\Delta\psi}{2} - \Delta\phi_p < 0 \\
&\qquad \textbf{then } \phi_p|_R - \frac{\Delta\psi}{2} + z\left(-\frac{\Delta\psi}{2}, \frac{\Delta\psi}{2} - \Delta\phi_p\right) \\
&\qquad \textbf{else } \phi_p|_L + z\left(-\frac{\Delta\psi}{2}, -\frac{\Delta\psi}{2} + \Delta\phi_p\right) \\
&\qquad \textbf{end if} \\
&\quad \textbf{end if}
\end{aligned} \tag{5.8}$$

where the function $z: \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$z(u, v) = \frac{1}{\alpha} \log\left(\frac{\exp(\alpha v) + 1}{\exp(\alpha u) + 1}\right). \tag{5.9}$$

Note that in (5.8) the function $z$ is used with only nonpositive arguments and

$$|z(u, v)| \leqq \frac{\log(2)}{\alpha} \quad \text{for } u \leqq 0,\ v \leqq 0. \tag{5.10}$$

By repeating the interpolation procedure, we can compute $\phi_p|_{L'}$ from $\phi_p|_L$ and $\phi_p|_M$, and $\phi_p|_{R'}$ from $\phi_p|_M$ and $\phi_p|_R$. In the particular case that the wall $M$ is the basis $B$, we do not first determine $\phi_p|_M$ by interpolation, but simply state that

$$\phi_p|_M \equiv \phi_p|_B = V_B. \tag{5.11}$$

Analogously, we can derive a formula for $\phi_n|_M$.

**5.4. Coarse grid correction.** Let $q_h^{old}$ and $q_H^{old}$ be given approximations to the solution on $\Omega_h$ and $\Omega_H$, respectively. The CGC is defined by

$$\text{compute } d_H = \bar{R}_H(f_h - \mathcal{M}_h(q_h^{old})), \tag{5.12a}$$

$$\text{solve } \mathcal{M}_H(q_H^{new}) = \mathcal{M}_H(q_H^{old}) + d_H, \tag{5.12b}$$

$$\text{compute } q_h^{new} = q_h^{old} + (P_h q_H^{new} - P_h q_H^{old}), \tag{5.12c}$$

where $\bar{R}_H$ and $P_h$ are the grid transfer operators defined in the previous subsections. Note that $q_H^{new}$ in (5.12b) may be approximated by applying a number of $\sigma$ FAS cycles on the grid $\Omega_H$ with $q_H^{old}$ as an initial approximation. The approximation $q_H^{old}$ may be given by means of full weighting:

$$q_H^{old} = R_H q_h^{old}, \tag{5.13a}$$

$$q_{H,j}^{old} = \tfrac{1}{2} q_{h,2j-1}^{old} + \tfrac{1}{2} q_{h,2j}^{old}, \qquad \forall j \text{ at } \Omega_H. \tag{5.13b}$$

Another possibility is to take $q_H^{old}$ equal to $q_H^{new}$ obtained from the last of previous CGCs. The solution efficiency of many nonlinear problems is not influenced by either choice of $q_H^{old}$. In our case however it is (see § 7).

**5.5. Full multigrid.** The full multigrid (FMG) algorithm provides the efficient construction of an initial approximation to the solution on a fine grid, once a solution on a coarse grid has been computed [5], [7]. Let $\Omega_{coarse}$ be the coarsest grid and $\Omega_{fine}$ be the finest one. Intermediate grids are denoted by $l \in \mathbb{N}$. Operators and grid functions now have $l$ as a subscript instead of $h$ or $H$. Here we introduce an improvement of the usual FMG in a quasi-Algol description.

(5.14)

```
procedure BOX-FMG ('M_fine(q_fine) = f_fine', input: f_fine, output: q_fine)
begin
 (1) for l from fine − 1 by −1 to coarse
 (2) do f_l = R̄_l f_{l+1}
 (3) end do
 (4) SOLVE ('M_coarse(q_coarse) = f_coarse', input: f_coarse, output: q_coarse)
 (5) for l from coarse + 1 to fine
 (6) do q_l = P_l q_{-1}
 (7)     to γ
 (8)     do FAS ('M_l(q_l) = f_l', input: f_l, in/output: q_l)
 (9)     end do
(10) end do
end procedure
```

where $\bar{R}_l$ is defined by (5.3). The improvement is in the lines (1)–(3) of the procedure. The grid function $f_l$ is independent of $q_l$; the components represent

$$f_{l,j} = \int_{B_j} D \, d\Omega, \tag{5.15}$$

i.e., the dope function integrated over box $B_j$. By means of (1)–(3) we compute the integral as a Riemann sum over a larger number of subintervals. This is more accurate because $D$ is a rapidly varying function. In the numerical experiments to come we use $\gamma = 1$ throughout. For SOLVE ( ) we use the techniques of § 4.

**6. Adaptation of the coarse grid correction.** Hemker [9] successfully applied box centered multigrid FAS iteration to the forward and the reverse biased diode problem. A key feature in his application is the prolongation based on locally constant fluxes.

This prolongation has been reformulated and made suitable for the transistor problem in § 5. Application of the same MG algorithm to the transistor problem gives rise to a complication in the CGC due to drastically varying problem coefficients. This complication and possible remedies are the topics of this section.

**6.1. Improper solution transfer.** The first attempt of applying multigrid to our specific problem was done by employing BOX-FMG with only two grids. The coarse grid problem (5.12b), within the CGC of FAS, was to be solved up to machine accuracy by means of Newton-CSGS. For several cases of our test problem it turned out that the two-grid algorithm gets stuck precisely at stage (5.12b) of the CGC. This is remarkable because Newton-CSGS was shown in § 4.2 to be successful for $\mathcal{M}_H(q_H) = f_H$ even for rather coarse grids. Apparently $f_H$ is within an appropriate range of $\mathcal{M}_H$ while the right-hand side of (5.12b) may be outside such a proper range of $\mathcal{M}_H$. The computational difficulty occurs in CSGS on the coarse grid exactly where one or more of the three solution components depicts a steep gradient. Consider two adjacent boxes $B_L^h$ and $B_R^h$ on the fine grid which together constitute a box $B^H$ on the coarse grid. Because of the steep gradient it may well occur that the problem coefficients, i.e., the entries of the Jacobian of $\mathcal{M}_h$, show a quite different order of magnitude on $B_L^h$ and $B_R^h$, respectively. Grid function $d_H$, the restriction of the residual, is dominated by the fine grid box with the large coefficients. On the other hand, the operator $\mathcal{M}_H$ is generated by the particular choice of $q_H^{old}$. This particular choice may be full weighting applied to $q_h^{old}$, or the last $q_H$ available, etc. Because of the steep gradient in $q_h^{old}$ there is a large range of possible values for $q_H^{old}$ at $B^H$. Depending on the choice of $q_H^{old}$ the operator $\mathcal{M}_H$ may have either large or small coefficients at box $B^H$ due to the exponential behavior of the entries in the Jacobian as a function of the solution. In the case of small coefficients, the right-hand side of (5.12b) may become out of the appropriate range for $\mathcal{M}_H$ ($d_H$ does not depend on the particular choice of $q_H^{old}$) and the two-grid algorithm gets stuck.

We will now confirm the foregoing by considering our discretized problem in more detail. Consider the center of the $\phi_p$-stencil given by (3.9b) and let us suppose that $\psi$ is monotonous on $[x_{j-3/2}, x_{j-1/2}]$; then either $s(-\alpha\Delta_{j-1}\psi) \geqq 1$ or $s(\alpha\Delta_j\psi) \geqq 1$. If both $|\Delta_{j-1}\psi|$ and $|\Delta_j\psi|$ are sufficiently small then $stp(j, 0)$ is approximated by

$$stp(j,0) \approx \alpha\bar{\mu}_p \exp(\alpha(\phi_{p,j} - \psi_j)) \cdot \left(\frac{1}{\Delta_{j-1}x} + \frac{1}{\Delta_j x}\right).$$

If both $|\Delta_{j-1}\psi|$ and $|\Delta_j\psi|$ are sufficiently large then $stp(j, 0)$ is approximated by

$$stp(j,0) \approx \alpha\bar{\mu}_p \exp(\alpha(\phi_{p,j} - \psi_j)) \cdot \begin{cases} \alpha\dfrac{\Delta_{j-1}\psi}{\Delta_{j-1}x} & \text{if} \quad \Delta_{j-1}\psi \geqq 0, \\ -\alpha\dfrac{\Delta_j\psi}{\Delta_j x} & \text{if} \quad \Delta_j\psi \leqq 0. \end{cases}$$

These approximations show that indeed the $\phi_p$-stencil is extremely sensitive to the difference $(\phi_{p,j} - \psi_j)$. Hence the $\phi_p$-stencil on the coarse grid is sensitive to how $\phi_{p,j}$ and $\psi_j$ on the coarse grid are determined from their counterparts on the fine grid. If $q_H^{old}$ is determined by applying full weighting (linear interpolation) to $q_h^{old}$ then

$$stp(j/2, 0) \approx \exp\left(-\frac{\alpha}{2}|\Delta_j(\phi_p - \psi)|\right) \cdot \max\{stp(j-1, 0), stp(j, 0)\}$$

where $stp(j-1, 0)$, $stp(j, 0)$ ($j$ even) are defined at the fine grid $\Omega_h$ and $stp(j/2)$ at the coarse grid $\Omega_H$. If $(\phi_p - \psi)$ shows a steep gradient then indeed

$$stp(j/2) \ll \max\{stp(j-1, 0), stp(j, 0)\}.$$

*Note.* The possible occurrence of the above sketched phenomenon has already been noted (for general nonlinear problems) by Brandt [5, p. 279], where he discusses how the transferred solution (i.e., $q_H^{old}$) implicitly determines the problem coefficients on the coarse grid.

**6.2. Possible remedies.** Let $L$ and $R$ be the centers of the two adjacent boxes $B_L^h$ and $B_R^h$ on the fine grid $\Omega_h$ which together constitute a coarse grid box $B_M^H$ with center $M$ on the coarse grid $\Omega_H$ (see Fig. 5).

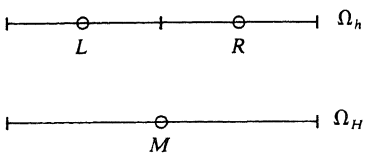


FIG. 5. *Nested boxes.*

Let us assume that $\partial\psi/\partial x = c$ is constant on $B_L^h \cup B_R^h$. The centers of the $\phi_p$-stencils at $L, R$ are then determined by the coefficients $a_L^h \equiv \exp(\alpha(\phi_p - \psi)|_L)\bar{c}$, $a_R^h \equiv \exp(\alpha(\phi_p - \psi)|_R)\bar{c}$, respectively, and the center of the $\phi_p$-stencil at $M$ by $a_M^H \equiv \exp(\alpha(\phi_p - \psi)|_M)\bar{c}$ with $\bar{c} = \alpha^2\bar{\mu}_p|c|$ (see § 6.1). Let $\Delta_M(\phi_p - \psi)$ denote the variation $\Delta_M(\phi_p - \psi) = (\phi_p - \psi)|_R - (\phi_p - \psi)|_L$. The solution at $M$ on the coarse grid somehow relates to the solution at $L$ and $R$ on the fine grid (for instance by means of the full weighting restriction). If $\Delta_M(\phi_p - \psi)$ is small (a smooth solution) then obviously $a_M^H$ does not differ much from either $a_L^h$ or $a_R^h$. If $\Delta_M(\phi_p - \psi)$ is large (a steep gradient in the solution) then $a_M^H$ may differ orders of magnitude from both $a_L^h$ and $a_R^h$, and therefore the MG algorithm may get stuck as was pointed out in the previous subsection. A radical remedy to meet this situation is to prevent $\Delta_M(\phi_p - \psi)$ from getting large, i.e., to introduce local refinement of the mesh just where the solution has a large variation $\Delta_M(\phi_p - \psi)$, e.g., by means of equidistributing the variation. However, we want to be able to find solutions without much refinement, in order to apply coarse grids in our MG algorithm. Besides, a solution without much resolution can serve as a guide for where a local mesh refinement should take place. For these reasons we resort to another remedy. Let us consider the CGC (5.12). Let $d_h(L)$, $d_h(R)$ be the residuals at $L, R$ (e.g., for the third equation (3.4c) only). At $M$ the difference between $q_H^{new}$ and $q_H^{old}$ may have the order of magnitude $d_H(M)/a_M^H$ with $d_H(M) = d_h(L) + d_h(R)$. Because of (5.12c) at either $L$, $R$, or both $L$ and $R$ a correction with order of magnitude $d_H(M)/a_M^H$ is added to the solution $q_h^{old}$. Assume that (because of a steep gradient in the solution) the inequality

$$a_M^H \ll \max\{a_L^h, a_R^h\}$$

holds. Therefore

$$d_H(M)/a_M^H \gg (d_h(L) + d_h(R))/\max\{a_L^h, a_R^h\},$$

which implies that the correction that will be transferred to the fine grid becomes far too large and the solution $q_h^{old}$ gets spoiled. A way to prevent this situation is to multiply the restricted residual $d_H$ with

$$\theta_M^H \equiv \frac{a_M^H}{\max\{a_L^h, a_R^h\}}, \qquad 0 < \theta_M^H \leqq 1, \tag{6.1}$$

at each center $M$. For a smooth part of the solution this fraction will be near one, for

a rapidly varying part of the solution it will be near zero, so that the solution $q_h^{old}$ will be preserved. The foregoing is the motivation for the following modification of the FAS algorithm (MFAS) using the notation of § 5.5:

(6.2)

**Procedure** MFAS ('$\mathcal{M}_l(q_l)=f_l$', *input*: $f_l$, *in/output*: $q_l$)
**begin**
(1) **If** $l=$ coarse
(2) **then** SOLVE ('$\mathcal{M}_{coarse}(q_{coarse})=f_{coarse}$', *input* : $f_{coarse}$, *in/output*: $q_{coarse}$)
(3) **else** RELAX ('$\mathcal{M}_l(q_l)=f_l$', *input*: $f_l$, *in/output*: $q_l$)
(4) $d_{l-1} := \bar{R}_{l-1}(f_l - \mathcal{M}_l(q_l))$
(5) $q_{l-1} := R_{l-1}q_l$ (optional !)
(6) $d_{l-1} := \Theta_{l-1}(\mathcal{M}_{l-1}, \mathcal{M}_l)d_{l-1}$
(7) $d_{l-1} := d_{l-1} + \mathcal{M}_{l-1}(q_{l-1})$
(8) $s_{l-1} := q_{l-1}$
(9) **to** $\sigma$
(10) **do** MFAS ('$\mathcal{M}_{l-1}(q_{l-1}) = d_{l-1}$', *input*: $d_{l-1}$, *in/output*: $q_{l-1}$)
(11) **end do**
(12) $q_l := q_l + P_l q_{l-1} - P_l s_{l-1}$
(13) RELAX ('$\mathcal{M}_l(q_l)=f_l$', *input*: $f_l$, *in/output*: $q_l$)
(14) **end if**
**end procedure**

The modification is in line (6). Here $\Theta_{l-1}$ represents a diagonal matrix $\mathbb{R}^{3N(\Omega_{l-1})} \to \mathbb{R}^{3N(\Omega_{l-1})}$ ($N(\Omega_{l-1})$ denotes the number of boxes at $\Omega_{l-1}$). It is defined by

$$\Theta_{l-1}d_{l-1} = (\boldsymbol{\theta}_{l-1,1}d_{l-1,1}, \cdots, \boldsymbol{\theta}_{l-1,j}d_{l-1,j}, \cdots, \boldsymbol{\theta}_{l-1,N(\Omega_{l-1})}d_{l-1,N(\Omega_{l-1})})^T$$

with $d_{l-1,j} \in \mathbb{R}^3$, $\theta_{l-1,j}: \mathbb{R}^3 \to \mathbb{R}^3$ and

$$\boldsymbol{\theta}_{l-1,j} \equiv \begin{pmatrix} \theta_{l-1,j,1} & 0 & 0 \\ 0 & \theta_{l-1,j,2} & 0 \\ 0 & 0 & \theta_{l-1,j,3} \end{pmatrix}$$

$$(\theta_{l-1,j,k} \in \mathbb{R},\ k=1,2,3).$$

For our particular semiconductor problem the $\theta_{l-1,j,k}$ are defined by:

$$\theta_{l-1,j,1} = 1, \tag{6.3a1}$$

$$\theta_{l-1,j,2} = \min\{2\eta_{l-1,j}, 1\}, \qquad \eta_{l-1,j} \in \mathbb{R}, \tag{6.3a2}$$

$$\theta_{l-1,j,3} = \min\{2\xi_{l-1,j}, 1\}, \qquad \xi_{l-1,j} \in \mathbb{R}, \tag{6.3a3}$$

where

$$\eta_{l-1,j} \equiv \exp(\alpha(\psi_j^{l-1} - \phi_{n,j}^{l-1})) \Big/ \max_{i=0,-1} \exp(\alpha(\psi_{2j+i}^{l} - \phi_{n,2j+i}^{l})),$$

$$\xi_{l-1,j} \equiv \exp(\alpha(\phi_{p,j}^{l-1} - \psi_j^{l-1})) \Big/ \max_{i=0,-1} \exp(\alpha(\phi_{p,2j+i}^{l} - \psi_{2j+i}^{l})). \tag{6.3b}$$

The superscripts $l-1$, $l$ refer to $\Omega_{l-1}, \Omega_l$, respectively.

The first component of the restricted residual does not need to be adjusted. The definition originates from the evaluation of expression (6.1). By means of (6.3a2)-(6.3a3) the numbers $\theta_{l-1,j,2}$ and $\theta_{l-1,j,3}$ are rounded off upwards to 1 when $\eta_{l-1,j}$, $\xi_{l-1,j}$ are $\geqq \frac{1}{2}$. Summarizing, we observe the following from (6.2)-(6.3):

(i) Where $q_l$ is smooth, $d_{l-1}$ will not be suppressed.

(ii) Where $q_l$ depicts a steep gradient, $d_{l-1}$ may be strongly suppressed.

(iii) Let $\Omega_0$ be some fixed grid, then, for $l \to \infty$, the matrix $\Theta_{l-1}$ becomes asymptotically the identity matrix.

(iv) By a proper local mesh refinement the suppression of $d_{l-1}$ will decrease. The performance of the modified FAS algorithm will be shown and discussed in § 7.

In the nonlinear multigrid algorithm as proposed by Hackbusch [7, p. 187], the restricted residual $d_{l-1}$ is divided by a global parameter $s \geqq 1$ and the resulting coarse grid correction is multiplied by $s$. The division by an appropiate $s$ ensures that the right-hand side of the coarse grid equation is within an appropriate range of the coarse grid operator $\mathcal{M}_{l-1}$. There are two main differences with our approach. First, the same number $s$ is used at each different box. Second, within our class of problems we have to omit the multiplication of the correction by $s$. Such a multiplication would result in a far too large correction and thereby a dip or peak in the fine grid solution. In recent work of Hackbusch and Reusken [8] a global parameter $\psi$ is proposed by which the coarse grid correction should be damped. For a limited class of problems an appropriate $\psi$ can be computed. Important differences with our approach are the following:

(i) $\psi$ is a damping parameter for the correction, instead of the residual.

(ii) $\psi$ is a global parameter, i.e., the same $\psi$ is used at each different box.

(iii) After sufficient FAS sweeps the damping parameter $\psi$ converges to 1, the parameters $\theta_{l-1,j,2}$, $\theta_{l-1,j,3}$ do not and should not converge to 1.

(iv) The $\psi$ parameter is meant to enlarge the domain of guaranteed convergence on the analogy of the damping parameter in the Newton method; the $\Theta_{l-1}$ operator is meant to deal with discrepancies between the operators $\mathcal{M}_{l-1}$ and $\mathcal{M}_l$ due to rapidly varying problem coefficients.

**7. Numerical results.** In this section we investigate the performance of our nonlinear multigrid algorithm. We focus our attention on the effects of local suppressing of the restricted residual and the choice of the coarse grid solution. The residual norm ($\|\cdot\|_{res}$) that we use is the maximum norm of the scaled residual. At level $l$ the said scaling is done by multiplying the residual at each box with the inverse of the $3 \times 3$-matrix

$$\left( \frac{\partial \mathcal{M}_l|_{x_{j-1/2}}}{\partial (\psi_j^l, \phi_{n,j}^l, \phi_{p,j}^l)^T} \right).$$

The performance of the MFAS algorithm is shown in Table 7.1. In the heading of the table we use the following abbreviations:

| | |
|---|---|
| case: | see Table 4.1. |
| $q_H^{old}$: defined by ...: | see § 5.4. |
| without $\Theta$: | No local suppression of the restricted residual is applied. |
| with $\Theta$: | Local suppression of the restricted residual is applied on all coarser grids. |
| #MFAS, $10^{-1}$ red: | The average number of MFAS sweeps necessary to obtain an additional reduction factor $10^{-1}$ of the residual norm after the application of BOX-FMG. |
| after FMG: | The last column shows the scaled norm of the residual, after application of BOX-FMG (see (5.14), $\gamma = 1$). In each case $> 0$ we obtain a starting approximation of the solution on the coarsest grid by means of continuation and application of Newton-CSGS (see § 4.2). |

For Table 7.1 the multigrid procedures are applied with 3 grids, with $N = 16$, 32, 64, respectively. In the event of no convergence the symbol * is written.

We observe that the use of the $\Theta$ operator, combined with a proper choice of the coarse grid solution, gives convergence for all cases. In the cases 3–6 the use of the $\Theta$ operator is essential for convergence. The use of the $\Theta$ operator does not slow down

TABLE 7.1
*Performance of* MFAS; *use of* 3 *grids*: $N = 16, 32, 64$, *respectively.*

| | $q_H^{old}$: Defined by full weighting | | $q_H^{old}$: Defined by $q_H^{new}$ | | |
|---|---|---|---|---|---|
| Case | Without $\Theta$ #MFAS, $10^{-1}$ red. | With $\Theta$ #MFAS, $10^{-1}$ red. | Without $\Theta$ #MFAS, $10^{-1}$ red. | With $\Theta$ #MFAS, $10^{-1}$ red. | With $\Theta$ After FMG |
| 0 | 0.80 | 0.80 | 0.66 | 0.66 | $4.1_{10}-4$ |
| 1 | 1.07 | 1.07 | 0.88 | 0.73 | $8.8_{10}-4$ |
| 2 | * | 1.15 | * | 0.85 | $1.2_{10}-3$ |
| 3 | * | * | * | 1.03 | $1.2_{10}-3$ |
| 4 | * | * | * | 1.06 | $7.1_{10}-4$ |
| 5 | * | * | * | 0.89 | $5.7_{10}-4$ |
| 6 | * | * | * | 0.89 | $5.7_{10}-4$ |
| 7 | * | 1.38 | * | 0.89 | $5.7_{10}-4$ |
| 8 | 1.40 | 1.37 | 0.89 | 0.89 | $5.6_{10}-4$ |
| 9 | 1.54 | 1.37 | 0.87 | 0.90 | $5.7_{10}-4$ |
| 10 | 2.59 | 2.52 | 0.88 | 0.82 | $1.1_{10}-3$ |
| 11 | 1.22 | 1.22 | 1.25 | 1.25 | $1.3_{10}-3$ |
| 12 | 1.75 | 1.74 | 2.01 | 2.01 | $9.1_{10}-4$ |
| 13 | 4.66 | 4.66 | 2.19 | 2.19 | $1.9_{10}-3$ |
| 14 | 2.44 | 2.43 | 1.76 | 1.77 | $2.5_{10}-3$ |

convergence in the cases where it is not needed (cases 0–2 and 7–14). We observe further that apparently the full weighting approximation of the fine grid solution on the coarse grid may be a poor one. Experiments for more and finer grids showed almost identical results for Table 7.1. Further we observe that mere application of BOX-FMG, without further MFAS sweeps, already gives fairly accurate results which may be good enough for practical purposes.

For two typical cases, case 4 and case 12, we investigate the grid-dependence of the multigrid convergence. In Fig. 6 we show the 10-logarithm of the scaled residual norms after subsequent FAS sweeps, starting from the result obtained by BOX-FMG. The coarsest grid contains 16 boxes; for the finest grid we take 32, 64, 128, and 256 boxes, respectively; $\Theta$ is applied (without application of $\Theta$ case 4 persistently depicts divergence). We observe that the multigrid convergence becomes grid-independent when the meshsize of the finest grid decreases. This indicates that the semiconductor problem has been treated correctly at each multigrid stage. Hereby it is shown that even for the strongly nonlinear (and particularly hard) problem it is possible to compose a multigrid method with optimal multigrid efficiency. Of course, considered in one space dimension only, competitive methods are available. However, the multigrid method as developed in this paper offers several clues for the foundation of an MG algorithm which solves the semiconductor problem also in more space dimensions with a computational complexity that is linear in the number of gridpoints.

In order to give some insight into the behavior of the $\Theta$ operator, we show in Fig. 7 the solution components $\psi$ and $\phi_n$ for case 4 on a 64-grid and a graph of $\theta_{2,j,2}$ (see (6.3a2)).

We observe the typical behavior that $\theta_{2,j,2}$ equals 1 almost everywhere, except for some isolated points.

Fig. 8 shows the electrostatic potential $\psi$ as computed on a grid of 128 boxes.

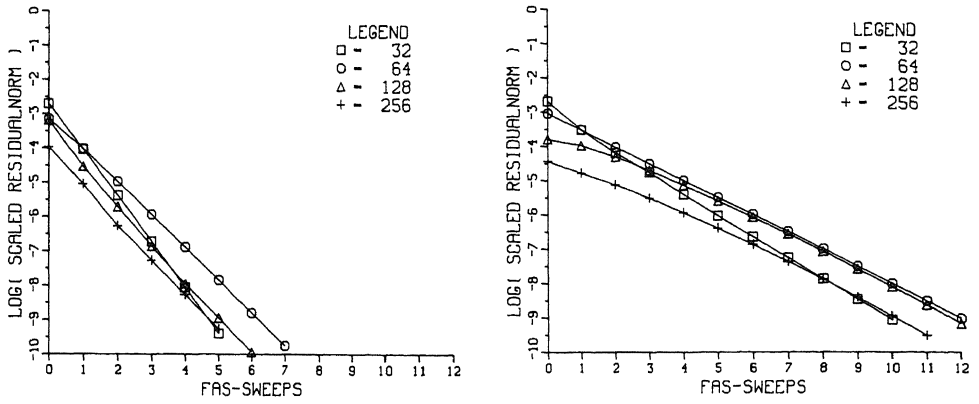


FIG. 6. *Multigrid convergence histories; the coarsest grid numbers* 16 *boxes*: (a) *case* 4; (b) *case* 12.

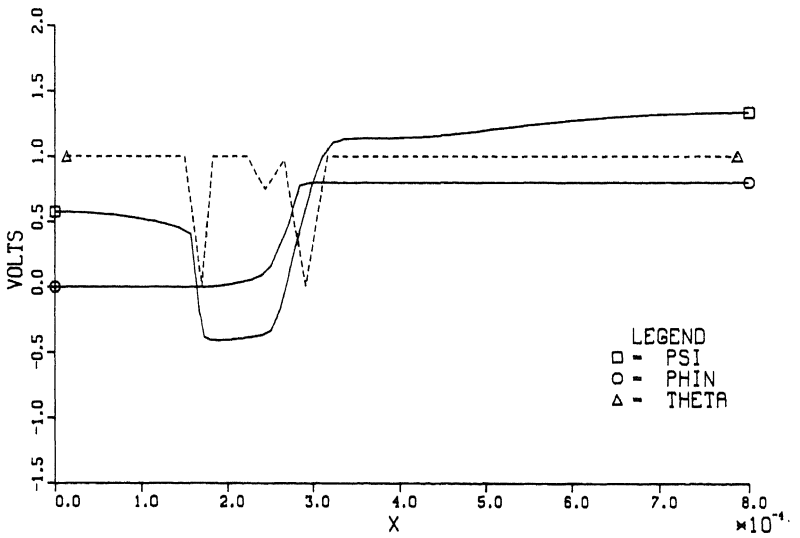


FIG. 7. *The components* $\psi$ *and* $\phi_n$ *of the solution for case* 4 *and the corresponding* $\theta$.

**8. Conclusions.** We find, by deriving explicit expressions for the entries of the Jacobian, that the linearization of the Scharfetter-Gummel discretization scheme contains a weak spot. When applying full multigrid followed by FAS/NMGM-iterations to our one-dimensional transistor problem, we find a serious lack of robustness which is explained by the strong nonlinearity of the discretized problem. This difficulty is met by adaptation of the coarse grid correction, which looks to be equally applicable for the higher-dimensional case. A proper choice of the coarse grid solutions is of importance too, e.g., the full weighting approximation is not satisfactory. Furnished with the improvements as proposed, we obtain a robust multigrid algorithm with a convergence which is independent of the meshsize.

**Acknowledgements.** The author wishes to thank Prof. Dr. P. W. Hemker and Prof. Dr. P. Wesseling for their useful comments and Ms. M. Middelberg for typing the manuscript.

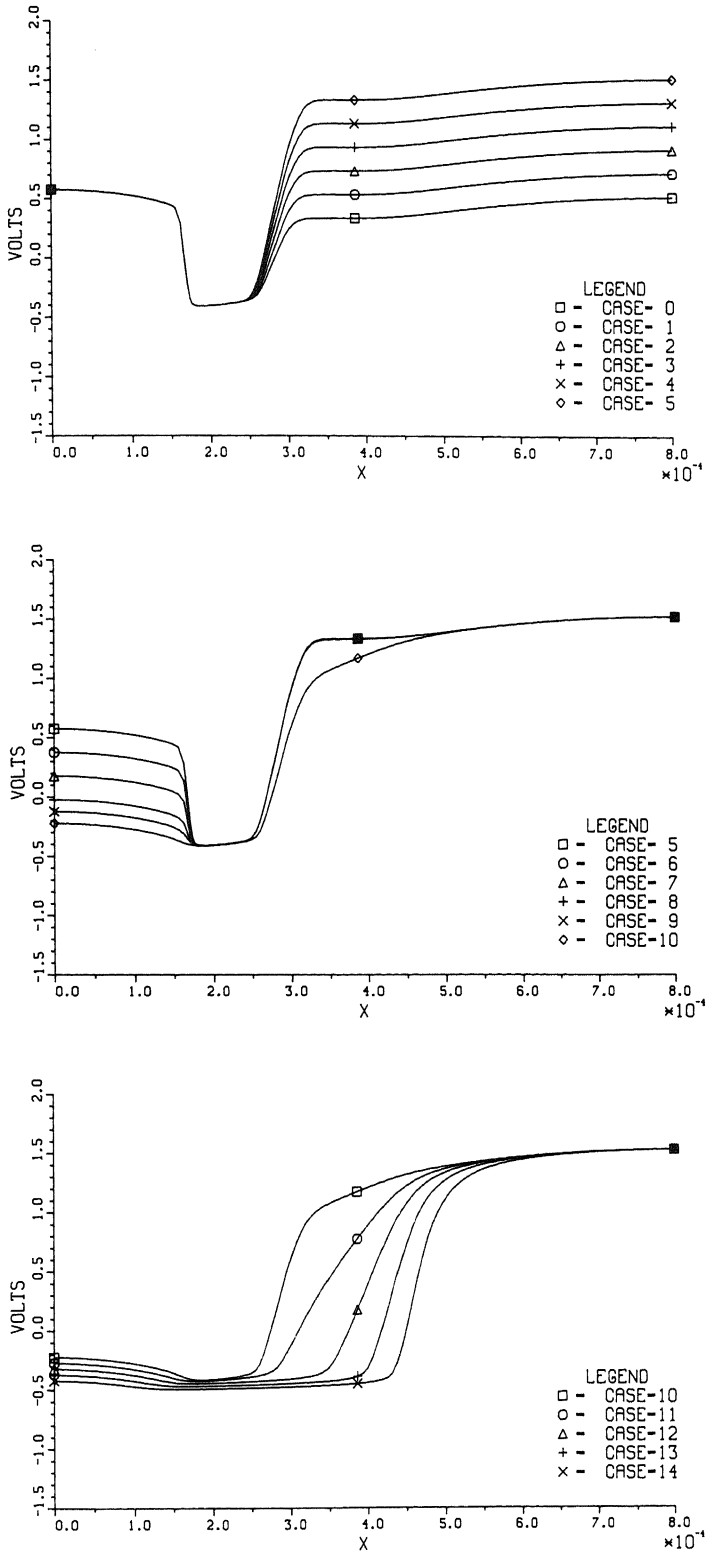


FIG. 8. *The electrostatic potential $\psi$ on a* 128*-grid.*

## REFERENCES

[1] R. E. BANK, J. W. JEROME, AND D. J. ROSE, *Analytical and numerical aspects of semiconductor device modelling*, Computing Methods in Applied Sciences and Engineering, R. Glowinski and J. L. Lions, eds., North-Holland, Amsterdam, 1982, pp. 593–597.

[2] R. E. BANK AND H. D. MITTELMANN, *Continuation and multi-grid for nonlinear elliptic systems*, Multigrid Methods II, W. Hackbusch and U. Trottenberg, eds., Springer-Verlag, Berlin, New York, 1986, pp. 23–37.

[3] R. E. BANK AND D. J. ROSE, *Analysis of a multilevel iterative method for nonlinear finite element equations*, Math. Comp., 39 (1982), pp. 453–465.

[4] R. E. BANK, D. J. ROSE, AND W. FICHTNER, *Numerical methods for semiconductor device simulation*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 416–435.

[5] A. BRANDT, *Guide to multigrid development*, Lecture Notes in Mathematics 960, W. Hackbusch and U. Trottenberg, eds., Springer-Verlag, Berlin, 1981, pp. 220–312.

[6] S. P. GAUR AND A. BRANDT, *Numerical solution of semiconductor transport equations in two dimensions by multi-grid method*, Advances in Computer Methods for Partial Different Equations II, R. Vichnevetsky, ed., IMACS (AICA), New Brunswick, NJ, 1977, pp. 327–329.

[7] W. HACKBUSCH, *Multi-grid methods and applications*, Springer Series in Computational Mathematics 4, Springer-Verlag, Berlin, 1985.

[8] W. HACKBUSCH AND A. REUSKEN, *On global multigrid convergence for nonlinear problems*, Robust Multigrid Methods, Notes on Numerical Fluid Dynamics, W. Hackbusch, ed., Vieweg-Verlag, Braunschweig, 1988.

[9] P. W. HEMKER, *A nonlinear multigrid method for one-dimensional semiconductor device simulation: The diode*, BAIL V Proceedings of the Fifth International Conference on Boundary and Interior Layers, Shanghai, China, 1988.

[10] S. J. POLAK, C. DEN HEIJER, W. H. A. SCHILDERS, AND P. A. MARKOWICH, *Semiconductor device modelling from the numerical point of view*, Internat. J. Numer. Methods Engrg., 24 (1987), pp. 763–838.

[11] D. SCHARFETTER AND H. K. GUMMEL, *Large-signal analysis of a silicon Read diode oscillator*, IEEE Trans. Electron Dev., ED-16, (1969), pp. 64–77.

[12] S. SELBERHERR, *Analysis and simulation of semiconductor devices*, Springer-Verlag, Berlin, New York, 1984.

[13] A. S. L. SHIEH, *Solution of coupled systems of PDEs by the transistorized multi-grid method*, in Proceedings of a Conference on Numerical Solutions of VLSI Devices, Boston, 1984.

[14] B. P. SOMMEIJER, W. H. HUNDSDORFER, C. T. H. EVERAARS, P. J. VAN DER HOUWEN, AND J. G. VERWER, *A numerical study of a 1D stationary semiconductor model*, Note NM-8702, Dept of Numerical Mathematics, Centre for Mathematics and Computer Science, Amsterdam, the Netherlands, 1987.

[15] D. M. YOUNG, *Iterative solution of large linear systems*, Academic Press, New York, London, 1971.

[16] P. M. DE ZEEUW, *Nonlinear multigrid applied to a 1D stationary semiconductor model*, Report NM-R8905, Dept. of Numerical Mathematics, Centre for Mathematics and Computer Science Amsterdam, the Netherlands, 1989.