Journal of Computational and Applied Mathematics 30 (1990) 117–126
North-Holland

Letter Section

A nonlinear multigrid method for one-dimensional semiconductor device simulation: results for the diode

P.W. HEMKER
Centrum voor Wiskunde en Informatica, P.O. Box 4079, 1009 AB Amsterdam, Netherlands

Received 18 September 1989
Revised 16 October 1989

Abstract: This paper studies a multigrid method for the solution of the semiconductor device simulation problem. Although the real impact of multigrid will always be in two or more dimensions, here the possibility of the method is investigated for the one-dimensional case. The essential difficulty for multigrid for the semiconductor problem is the possible adverse effect of very coarse grids on the convergence rate of the method, and the difficult computation of a sufficiently close initial approximation for the nonlinear iterative solver on the coarse grids.

For the solution on the coarsest grid, continuation is applied together with Newton iteration. The latter is stabilized by the possible insertion of collective symmetric Gauss–Seidel relaxation sweeps for smoothing the iterands.

The multigrid method makes use of a box method and a dedicated nonlinear prolongation that is adapted to the Scharfetter–Gummel discretisation.

A standard diode-model problem, both with forward and with reversed bias, is used to show that true multigrid convergence can be obtained indeed, even with very coarse meshes on the coarse grids. However, V-cycles are not always sufficient and W-cycles may be required.

Keywords: Semiconductor device simulation, multigrid method.

1. Introduction

We study the solution of the nonlinear system of equations that is obtained by discretisation of the 1-D semiconductor device modelling equations, by means of nonlinear FAS iteration as described in [1], and we restrict ourselves to one of the simplest possible cases: the 1-D diode.

For the discretisation of the 1-D equations, the interval of definition is partitioned into cells. A box scheme is used and the flux at the cell boundaries is computed by the Scharfetter–Gummel scheme. This scheme is derived by assuming constant fluxes on dual cells. Based on this assumption, a nonlinear interpolation was introduced in [1]. With this particular interpolation

and with a straightforward finite-volume restriction, on a set of nested cell partitionings, the Scharfetter–Gummel discretisations form—in a sense—a nested set of discretisations. This is the motivation to use these prolongation and restriction operators in the FAS multigrid method.

To start the multigrid iteration, initial estimates are computed by Full Multi Grid. A combination of a continuation process, a nonlinear relaxation method and Newton's method is used to solve the discrete problems on the coarsest grids.

In Section 4 we describe the computation of the initial estimate, in Section 5 we describe the diode problem that is solved, and in Section 6 we give the results obtained with the multigrid cycling procedure. Finally, we summarise some conclusions.

2. The equations

The partial differential equations describing the behaviour of the semiconductor device are given by (cf., e.g., [2]):

$$-\operatorname{div}(\epsilon \operatorname{grad} \psi) = q(p - n + D), \tag{2.1a}$$

$$-\operatorname{div}\big(\mu_n(\operatorname{grad} n - n \operatorname{grad}(\alpha\psi + \log n_i))\big) = -R, \tag{2.1b}$$

$$-\operatorname{div}\big(\mu_p(\operatorname{grad} p + p \operatorname{grad}(\alpha\psi - \log n_i))\big) = -R, \tag{2.1c}$$

on $\Omega \subset \mathbb{R}^2$. The variables ψ, n and p represent the electric potential and the electron and hole densities respectively; ϵ, q and α are constant values. The doping profile D is a given (nonsmooth) function of the independent variable x. The parameters μ_n, μ_p as well as the right-hand side R generally are functions of x, ψ, n and p, and n_i is a mild function of x. For simplicity, we consider only $R = 0$ and constant μ_n, μ_p and n_i. With these assumptions, (2.1) reduces to

$$-\operatorname{div}(\epsilon \operatorname{grad} \psi) = q n_i(\bar{p} - \bar{n} + \bar{D}), \tag{2.2a}$$

$$-\operatorname{div}\big(\mu_n(\operatorname{grad} \bar{n} - \bar{n} \operatorname{grad}(\alpha\psi))\big) = 0, \tag{2.2b}$$

$$-\operatorname{div}\big(\mu_p(\operatorname{grad} \bar{p} + \bar{p} \operatorname{grad}(\alpha\psi))\big) = 0, \tag{2.2c}$$

where $\bar{n} = n/n_i$, $\bar{p} = p/n_i$ and $\bar{D} = D/n_i$. Usual boundary conditions are either of Dirichlet type (at the contacts $\overline{pn} = 1$, $\bar{p} - \bar{n} + \bar{D} = 0$, ψ prescribed) or of Neumann type (cf. [2]).

Because of the large range of possible values for $\bar{n}$ and $\bar{p}$, it is convenient to introduce the quasi-Fermi levels as new variables:

$$\phi_n = \psi - \frac{\log \bar{n}}{\alpha} = \psi - \frac{1}{\alpha}\log\left(\frac{n}{n_i}\right), \tag{2.3a}$$

$$\phi_p = \frac{\log \bar{p}}{\alpha} + \psi = \frac{1}{\alpha}\log\left(\frac{p}{n_i}\right) + \psi. \tag{2.3b}$$

In the new set of variables (ψ, ϕ_n, ϕ_p), (2.2) can be rewritten

$$-\operatorname{div}(\lambda^2 \operatorname{grad} \psi) = e^{\alpha(\phi_p - \psi)} - e^{\alpha(\psi - \phi_n)} + \bar{D}, \tag{2.4a}$$

$$-\operatorname{div}\big(\mu_n\, e^{\alpha\psi - \alpha\phi_n} \operatorname{grad}(\alpha\phi_n)\big) = 0, \tag{2.4b}$$

$$-\operatorname{div}\big(\mu_p\, e^{\alpha\phi_p - \alpha\psi} \operatorname{grad}(\alpha\phi_p)\big) = 0, \tag{2.4c}$$

where $\lambda^2 = \epsilon/q n_i$.

Introducing the notation $J_\psi = \lambda^2 \text{grad}\ \psi$, $J_n = \mu_n e^{\alpha\psi - \alpha\phi_n} \text{grad}(\alpha\phi_n)$ and $J_p = \mu_p e^{\alpha\phi_p - \alpha\psi} \text{grad}(\alpha\phi_p)$, we find for arbitrary $\Omega_\alpha \subset \Omega$:

$$-\int_{\Gamma_\alpha} J_\psi \nu \, d\Gamma = \int\int_{\Omega_\alpha} e^{\alpha(\phi_p - \psi)} - e^{\alpha(\psi - \phi_n)} + \bar{D} \, d\Omega, \tag{2.5a}$$

$$-\int_{\Gamma_\alpha} J_n \nu \, d\Gamma = 0, \tag{2.5b}$$

$$-\int_{\Gamma_\alpha} J_p \nu \, d\Gamma = 0, \tag{2.5c}$$

where ν is the outward pointing normal at Γ, the boundary of Ω_α. This system of equations, together with the boundary conditions, is written in symbolic form as

$$N(q) = r(q), \tag{2.6}$$

where N is the nonlinear differential operator in the left-hand side of (2.4) and $r(q)$ is the right-hand side; q denotes the vector of unknown functions $q = (\psi, \phi_n, \phi_p)$.

To preserve the conservation character of the equations, for the discretisation of (2.4) we use a finite-volume technique. We divide the interval $\Omega = (x_0, x_N)$ in disjoint *boxes* (i.e., intervals) $\Omega_i = (x_{i-1}, x_i)$ $(i = 1, \ldots, N)$. Inside each box Ω_i a point $x_{i-1/2}$ is selected and for each box we approximate values of the variables ψ, ϕ_n and ϕ_p. To define a proper sequence of refining meshes as $N \to \infty$, we introduce a monotonously increasing $C^1[0, 1]$-function $\gamma : [0, 1] \to \Omega$ such that, for a fixed N, $x_i = \gamma(i/N)$. Another set of subintervals $\{D_i\}$ is introduced with $D_i = (x_{i-1/2}, x_{i+1/2})$ $(i = 1, \ldots, N-1)$, $x_{i-1/2} = \frac{1}{2}(x_{i-1} + x_i)$ or $x_{i-1/2} = \gamma((i - \frac{1}{2})/N)$, $D_0 = (x_0, x_{1/2})$, $D_N = (x_{N-1/2}, x_N)$. These intervals form the set of *dual boxes*. Thus, for a given function γ, sets $\{\Omega_i\}_{i=1,\ldots,N}$, and $\{D_i\}_{i=0,\ldots,N}$ are defined for an arbitrary $N \in \mathbb{N}$. The different discretisations are parametrised by $h = 1/N$. The set of boxes is denoted by $\Omega_h = \{\Omega_i \mid i = 1, 2, \ldots, N\}$.

A discrete representation $q_h \in S_h$ of the state of the semiconductor is given by the $3N$-dimensional vector $q_h = \{q_i\}_{i=1,\ldots,N} = \{(\psi_i, \phi_{n,i}, \phi_{p,i})\}_{i=1,\ldots,N}$. Notice that q_i is associated with the box Ω_i and can also be associated with $x_{i-1/2}$.

The discretisation we use is based on the piecewise constant approximation of J_ψ, J_n and J_p on the dual mesh $\{D_i\}$. These piecewise constant functions are derived from q_h by

$$J_{\psi,i} = \lambda^2 \frac{\psi_{i+1} - \psi_i}{x_{i+1/2} - x_{i-1/2}}, \tag{2.7}$$

and

$$J_{n,i} = \mu_n \frac{\exp(-\alpha\phi_{n,i+1}) - \exp(-\alpha\phi_{n,i})}{\exp(-\alpha\psi_{i+1}) - \exp(-\alpha\psi_i)} \cdot \frac{\alpha\psi_{i+1} - \alpha\psi_i}{x_{i+1/2} - x_{i-1/2}} \tag{2.8}$$

and a similar expression for $J_{p,i}$.

The discretisation of (2.1) is simply based on (2.5) with $\Omega_\alpha = \Omega_i$ $(i = 1, \ldots, N)$. The integral $\int\int_{\Omega_i}$ in the right-hand side of (2.5a) is approximated by the one-point quadrature formula and the

left-hand side is computed as

$$-\int_{\Omega_i} J_\psi \nu \, d\Gamma = J_{\psi,i-1} - J_{\psi,i}, \tag{2.9a}$$

$$-\int_{\Omega_i} J_n \nu \, d\Gamma = J_{n,i-1} - J_{n,i}, \tag{2.9b}$$

$$-\int_{\Omega_i} J_p \nu \, d\Gamma = J_{p,i-1} - J_{p,i}. \tag{2.9c}$$

Analogous to (2.6), the discrete equations are written in symbolic form as

$$N_h(q_h) = r_h(q_h). \tag{2.10}$$

In fact, by the above construction we derive a cell-centered version of the well-known Scharfetter–Gummel scheme. In [1] we construct a $P_{h,2h}$ such that

$$\bar{R}_{2h,h} N_h(P_{h,2h} q_{2h}) = N_{2h}(q_{2h}) \tag{2.11}$$

for all coarse-grid functions q_{2h}.

3. The multigrid method

So solve the nonlinear system

$$M_h(q_h) := N_h(q_h) - r_h(q_h) = f_h, \tag{3.1}$$

we use a nonlinear multigrid (FAS) method [1]. For a vanishing right-hand side f_h this system is the system of equations (2.10). The FAS method to solve (3.1) is an iterative process, in which each cycle consists of: (i) a number of p nonlinear relaxation sweeps; (ii) a coarse-grid correction; (iii) another q nonlinear relaxation sweeps.

As a *relaxation* procedure we use a nonlinear Collective Symmetric Gauss–Seidel (CSGS) relaxation. The *coarse-grid correction* consists of the following steps:

$$M_{2h}(\tilde{q}_{2h}) = M_{2h}(q_{2h}) + \bar{R}_{2h,h}\left(f_h - M_h\left(q_h^{(n)}\right)\right), \tag{3.2a}$$

$$q_h^{(n+1)} = q_h^{(n)} + P_{h,2h}\tilde{q}_{2h} - P_{h,2h} q_{2h}. \tag{3.2b}$$

Here q_{2h} is an (arbitrary) approximation to the solution on the grid Ω_{2h}. The value $\tilde{q}_{2h}$ may be either computed from the nonlinear system (3.2a), or it may be approximated by a number of σ multigrid cycles for the solution of (3.2a), applied to the initial approximation q_{2h}. For details on the FAS iteration we refer to [1]. In the present paper we concentrate on the computation of the initial estimate and on the observed rate of convergence for a model diode.

4. The computation of an initial estimate

To start the multigrid solution procedure, an initial estimate for the solution is needed. A natural and generally efficient way to obtain this estimate is by the use of *Full Multi Grid*, i.e., by the computation first of an estimate on a coarser grid, and subsequent interpolation of this

nate to the next finer one. The estimate on the coarser grid is obtained from one on a still ser grid, etc. This is a good and efficient procedure when a useful approximation on the y) coarsest grid can be found. Thus we are left with the need to compute such a coarse oximation (if it exists). The problem to construct a method for finding such an approxima- should be considered as a problem much different from the search for an accurate solution iod on a fine mesh. The reason is that most asymptotic arguments for $h \to 0$, to justify the ss of a discrete solution (method), lose their meaning on these coarse grids as they are based ocal properties of the solution. For very coarse grids we can only rely on global properties as ided by conservation laws and maximum principles (if available).

he difficulty that comes first to mind, viz. that the representation of the solution on a very se grid may be only a very poor approximation of the true solution is of minor importance. problems arise, e.g., with simple elliptic problems where the solution minimises a functional. — in general— such functionals do not exist, and it will be hard to verify the existence and ueness of the coarse discrete solutions in some—relatively large—neighbourhood of an ing and unique continuous solution.

hen working in large neighbourhoods around solutions, an additional difficulty, inherent to semiconductor problem, is the range of the exponential and/or the domain of the logarith- function. To handle the nonlinearity in the equations (2.4), they are best expressed in the ables ψ, n and p. However, computation in the variables n and p suffers from the restricted esentation of large real numbers in a computer. Using the variables ψ, ϕ_n and ϕ_p, and the ection transformation as described in [1,2] we may find large corrections $d\psi$, $d\phi_n$ and $d\phi_p$, h may lead to negative arguments for the logarithmic functions. That is, in the Newton ess iterands may appear that lie outside the domain of definition of the differential ations. This means that we cannot rely on Newton type methods only.

e use the following techniques to find the solutions on the coarse grids: (1) continuation respect to a parameter (v_B, the applied voltage at a contact), and (2) local relaxation.

he primary approach is to use a continuation process, starting from an applied voltage which ls an "easy" problem, and moving stepwise to the applied voltage for which the solution is ted. The steps in the continuation process are made by a combination of a Newton and a xation procedure. The Newton procedure is used because of its quadratic convergence in the hbourhood of the solution, and it is supplemented with local relaxation (the nonlinear ective Symmetric Gauss–Seidel relaxation as described in [1]) when—in the Newton process roblems are encountered with the domain of the logarithmic function. This combination of vton iteration and CSGS relaxation is called *CSGS–Newton iteration*. It appears that rnating the Newton steps with the CSGS steps improves the robustness of the iteration ess considerably.

he CSGS–Newton iteration starts with a nonlinear Collective Symmetric Gauss–Seidel xation sweep. Then a Newton step is made, using the correction transformation [1, eqs. (5.2)]. ase damping ([1, eq. (5.4)] with $s < s_0$) was necessary in the application of [1, (5.2)], the vton step is followed by another CSGS sweep. After this possible relaxation, the process tinues with a new Newton step, possibly followed by a another CSGS relaxation, etc. In a verging process, the final corrections will be sufficiently small and damping is no longer essary. Thus, the later steps in the CSGS–Newton iteration are only Newton steps. The uracy requirement in the solution of the small nonlinear systems in the CSGS relaxation is the sup-norm of the Newton correction should be less than 0.02.

If, in the CSGS–Newton iteration, after a certain number of steps (i.c. 19) the Newton corrections are still larger than required (i.c. $1.0 \cdot 10^{-10}$), or if some correction becomes "unreasonably" large ($\Delta\psi$, $\Delta\phi_n$ or $\Delta\phi_p$ larger than $2\Delta v_B$), then the process is considered as nonconverging.

5. The problems

As a test problem we consider the simple one-dimensional standard diode as described in [2]. It is given by the equation (2.2) in one space dimension, on the interval (0.0, 0.001); with $\epsilon = 1.0359 \cdot 10^{-12}$, $q = 1.6021 \cdot 10^{-19}$, $n_i = 1.22 \cdot 10^{10}$, $\alpha = 38.68293$. The dope function is given by

$$D(x) = 1.0 \cdot 10^{18} \text{sign}(x - 0.0005),$$

$R = 0$ and μ_n and μ_p are constants. At the boundary charge neutrality is required: $p - n + D = 0$. Further, at $x = 0$ the boundary conditions are $\phi_n = \phi_p = 0$; at $x = 0.001$ the applied voltage v_B is given: $\phi_n = \phi_p = v_B$.

Computations are made for $v_B = 5.0$ (the standard case) and further for $v_B = 100.0$ (reverse bias) and $v_B = -1.0$ (forward bias).

For the discretisation two kinds of mesh are used: (1) a uniform, and (2) a nonuniform mesh. Both meshes were used with $N = 2^L$ ($L = 0, 1, \ldots, 8$ cells). The nonuniform mesh was defined by the mapping $x_i = f(\xi_i)$, where f is a differentiable and monotonously increasing function, and $\{\xi_i\}_{i=0,\ldots,N}$ form a uniform partition of [0, 1]. This function f is chosen such that a reasonable resolution of the layer at $x = 0.0005$ can be obtained for the case $v_B = 5.0$. At first sight it seems unreasonable to try uniform meshes for these problems, because it is known that the solution is rapidly varying near the depletion layer. Nevertheless, we are interested in the behaviour of the numerical methods for these cases, because we want to know how the numerical methods behave for not well-adapted coarse meshes.

For all v_B, the problem is first solved (on the coarsest grid) by continuation with respect to v_B, starting from the problem with $v_B = 0$. The initial estimate at $v_B = 0$ is the solution of the reduced problem ((2.2) with $\epsilon = 0$), i.e.

$$\phi_n(x) = \phi_p(x) = 0,$$
$$\psi(x) = \psi(0.0) \qquad \text{for } x < 0.0005,$$
$$\psi(x) = \psi(0.001) \qquad \text{for } x > 0.0005.$$

The continuation process is straightforward: first, the problem is solved for $v_B = 0$. The first v_B-step immediately tries v_B equal to its final value. In each continuation step the problem is solved by CSGS–Newton iteration. If the CSGS–Newton process does not converge, the process is restarted after halving the continuation step. If a continuation step is successful, it is accepted and the next step is taken 1.5 times the previous one. This process is continued until the final v_B is reached.

6. Results

To test the multigrid FAS iteration, we studied the convergence rate of the maximum norm of the residual for different meshes and different values of the parameters. The standard FAS

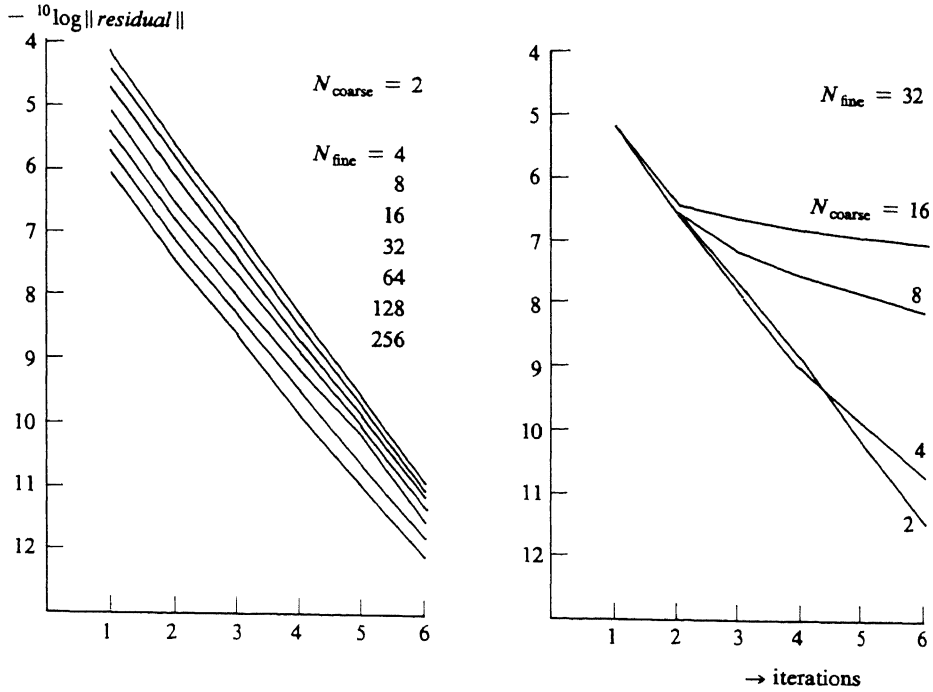

Fig. 1. (a) MG convergence for the diode with $v_B = -1.0$, for different values of N_{fine}; (b) As Fig. 1(a) but $N_{fine} = 32$, and different values for N_{coarse}.

method used (unless mentioned otherwise) consists of a V-cycle on a uniform mesh, with one CSGS relaxation sweep before and one after the coarse-grid correction (i.e., $p = q = 1$). The approximate solution procedure on the coarsest grid consists of two CSGS sweeps.

Application of FAS iteration to the forward biased diode ($v_B = -1.0$) shows a straightforward MG convergence behaviour: a constant decrease of the residual norm by a factor ≈ 0.06 is observed, independent of the number of cells in the mesh (Fig. 1(a)). It is also seen that the coarse grids have a positive effect on the convergence rate. In Fig. 1(b) we see the convergence behaviour on the finest mesh ($N_{fine} = 32$), in case of different levels for the coarsest grid (viz. $N_{coarse} = 2, 4, 8, 16$).

We notice that in this case the solution consists of a smooth component, an interior layer and two boundary layers. The layers are very sharp and cannot be resolved on the uniform mesh with $N_{fine} = 256$.

Application of FAS to the reverse biased problems ($v_B = 5.0$ or $v_B = 100.0$) shows a similar behaviour as the forward biased problem (Figs. 2(a) and 3). However, the convergence is less regular. A distinction should be made between the meshes on which the shape of the boundary layer can be resolved ($N > 64$ for $v_B = 100$, $N \geqslant 256$ for $v_B = 5$) and the coarser meshes. On the meshes that are fine enough, and on the very coarse meshes, an almost constant convergence rate of 0.06 is observed. A more irregular behaviour is seen for the intermediate meshes. On these meshes no asymptotic behaviour can be assumed for the discretisation. As in the forward biased case, it is seen that the coarse meshes still have an advantageous influence on the convergence. Figure 2(b) shows that a 2-level FAS method has a slower convergence than a more-level method.

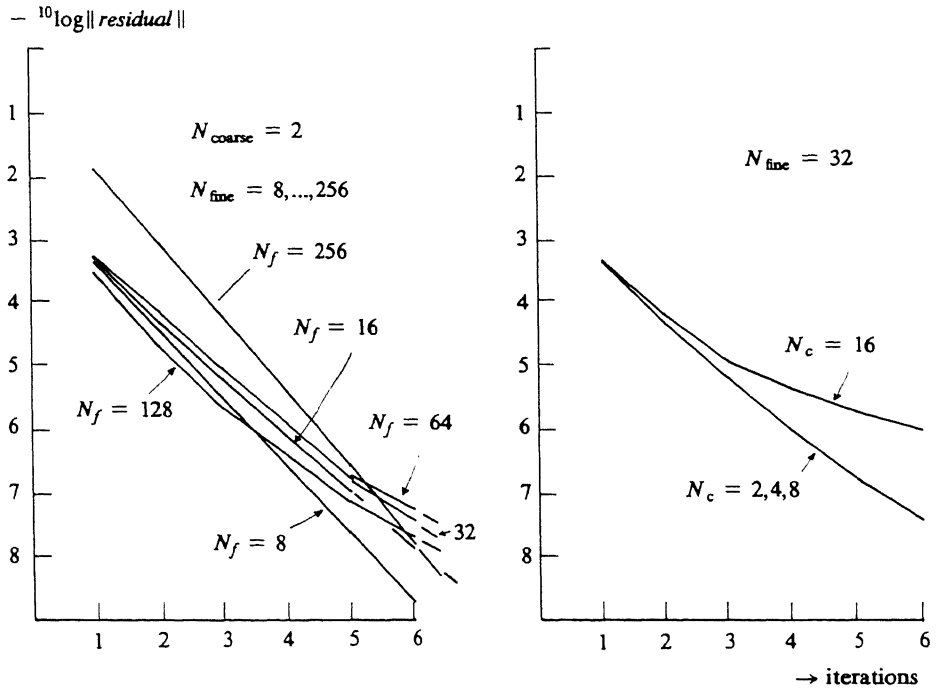

Fig. 2. (a) MG convergence for the diode with $v_B = 5.0$, for different values of N_{fine}; (b) As Fig. 2(a) but $N_{fine} = 32$, and different values for N_{coarse}.

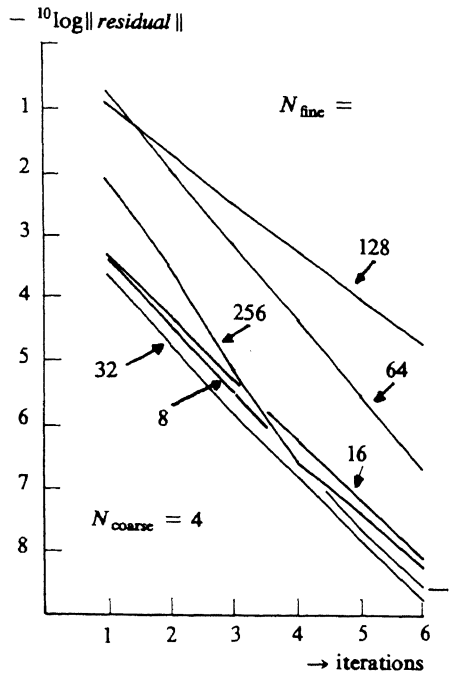

Fig. 3. MG convergence for the diode with $v_B = 100.0$, for different values of N_{fine}.

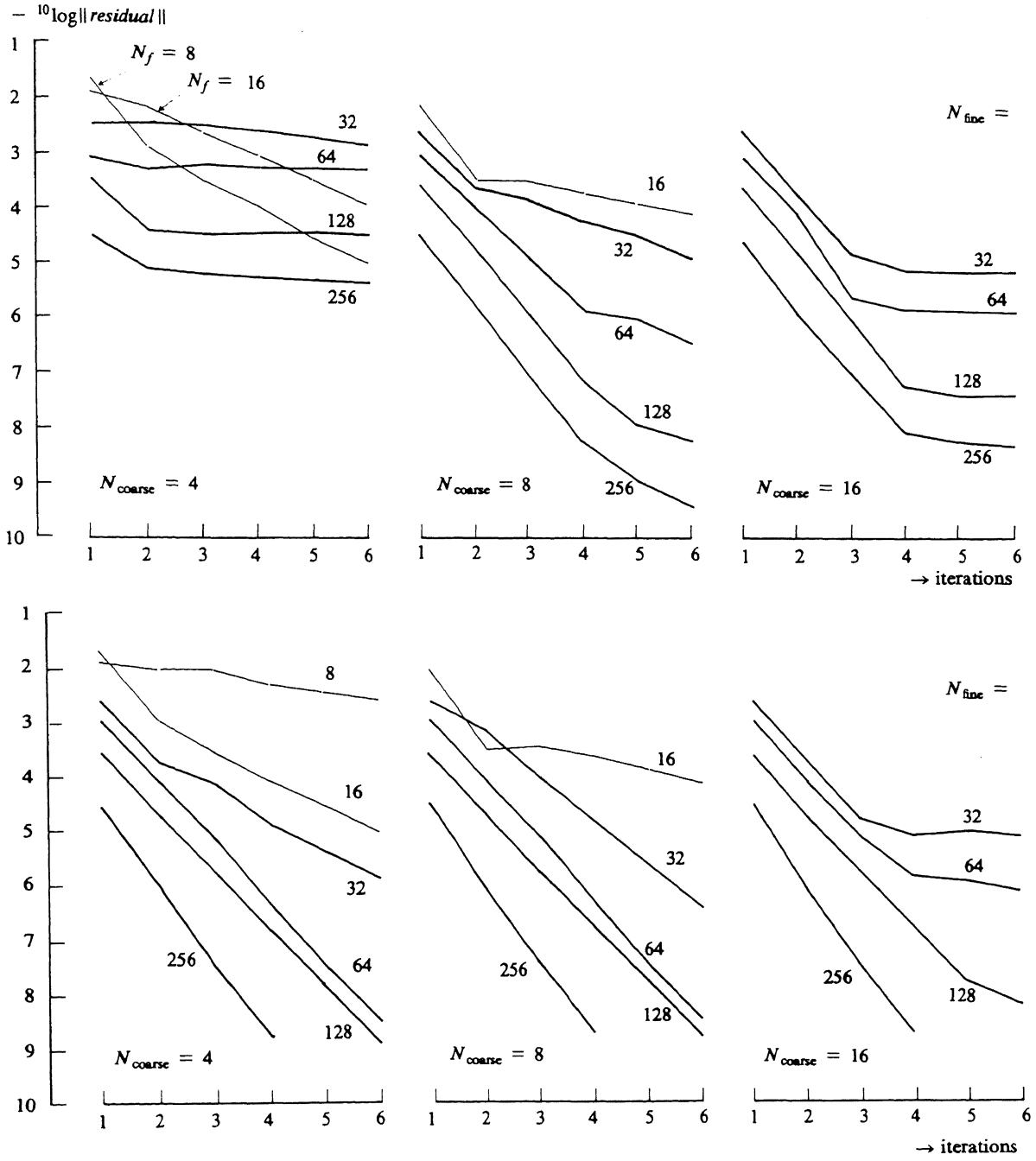

Fig. 4. MG convergence for for $v_B = 5.0$, on a nonuniform grid; The V-cycle (Figs. 4(a, b, c)) and the W-cycle (Figs. 4(d, e, f)).

Computation of the truncation error for the solution on the different meshes shows that in all cases 1 or 2 iteration steps are sufficient to drive the residual below the truncation error.

Application of the FAS V-cycle to the forward biased problem ($v_B = 5.0$) with a nonuniform mesh, shows that the convergence rate decreases after a few iteration steps (Figs. 4(a, b, c)). In particular, if the coarsest mesh contains only 4 cells, it was observed that the approximate solution on the coarse(st) grid caused the trouble. A more accurate solution method on the coarsest grid (e.g., the use of Newton's method instead of relaxation) improved the multigrid convergence. With relaxation as the approximate solver in the coarsest grid, convergence independent of N_{fine} was also seen if a W-cycle was used instead of a V-cycle (Figs. 4(d, e, f)). We observe a convergence rate of ≈ 0.05 if the number of meshes is sufficiently large.

7. Conclusion

We find that, for the 1-D diode as a model problem, typical multigrid convergence can be obtained for the discrete semiconductor device equations. A convergence factor is found that is essentially independent of the meshwidth. By embedding in a "Full Multigrid" algorithm, 1 or 2 iteration steps are sufficient to reduce the iteration error below truncation error. For this purpose, in the nonlinear FAS procedure the prolongation had to be adapted to the method of discretisation (Scharfetter–Gummel). Discrete operators on extremely coarse meshes still enhance the convergence behaviour.

References

[1] P.W. Hemker, A nonlinear multigrid method for one-dimensional semiconductor device simulation, in: Guo Ben-Yu and J.J.H. Miller, Eds., *BAIL V, Proc. Fifth Internat. Conf. on Boundary and Interior Layers, Shanghai* (Boole, Dublin, 1988).

[2] S.J. Polak, C. Den Heijer, W.H.A. Schilders and P. Marcowich, Semiconductor device modelling from the numerical point of view, *Internat. J. Numer. Methods Engrg.* **24** (1987) 763–838.