A NONLINEAR MULTIGRID METHOD FOR ONE-DIMENSIONAL SEMICONDUCTOR DEVICE SIMULATION.

P.W.Hemker
CWI, Centre for Mathematics and Computer Science
P.O.Box 4079, 1009 AB Amsterdam, The Netherlands

ABSTRACT

A multigrid method is introduced to solve the nonlinear system of equations that arises from the discretisation of the steady semiconductor device simulation equations. A special non-linear interpolation procedure is used to accommodate the multigrid (FAS) procedure to the Scharfetter- Gummel discretisation of the continuity equations.

1. INTRODUCTION

In this paper we study the solution of the nonlinear system of equations that is obtained by discretisation of the 1-D semiconductor device modelling equations. A few papers consider the multigrid solution of the discrete semiconductor equations [1, 2, 5, 10-13] . It seems, however, that the results do not show the full advantage of multigrid in the sense (i) that faster convergence could be expected and (ii) that possibly the coarsest grids in the process can be selected coarser. Therefore, we study non-linear FAS iteration. In contrast to the papers mentioned, we don't consider 2-D problems. Also in experiments, we restrict ourselves to one of the simplest possible cases: the 1-D diode.

In this paper, first we make an observation on the character of the PDEs. We see that the principle part of the differential operator is a product of the Laplace operator and a singularly perturbed operator. In the reduced form the principle part is a 4th order elliptic operator. For non-vanishing small parameter (which is the square of the Debye length), the operator is 6th order elliptic and the principle part contains as factors, besides the Laplacian, two convection diffusion operators: one with + grad ψ and one with − grad ψ for the convection direction.

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 a set of dual cells. Based on the same assumption, a non-linear interpolation is introduced. 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 a FAS multigrid method.

To start the multigrid iteration, initial estimates are computed by "Full Multi Grid": they are obtained by interpolation from a coarser grid. A combination of a nonlinear relaxation method and Newton's method is used to solve the discrete

problems on the coarsest grids. Results obtained with the method will be published elsewhere [8] .

2. The equations

The equations modelling the steady semiconductor device are given by (cf. e.g. [9]) :

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

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

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

on $\Omega \subset \mathbb{R}^2$.

The dependent variables ψ, n and p describe the electric potential and the electron and hole densities respectively; ϵ, the permittivity, and q, the elementary charge, are constant values, as is $\alpha = q/kT$, the inverse of the "thermal voltage". The doping D is a given (non-smooth) function of the independent space variable x. The electron and hole mobilities μ_n and μ_p as well as the net recombination- generation rate R generally are functions of x, ψ, n and p, and the intrinsic concentration n_i is a function of x. For simplicity, in this paper we consider only $R = 0$ and constant μ_n, μ_p and n_i. With these assumptions, (2.1) reduces to

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

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

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

where $\overline{n} = n/n_i$, $\overline{p} = p/n_i$ and $\overline{D} = D/n_i$. Usual boundary conditions are either of Dirichlet type (at the contacts $\overline{p}\,\overline{n} = 1$, $\overline{p} - \overline{n} + \overline{D} = 0$, ψ prescribed) or of Neumann type (cf. [9]).

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

$$\phi_n = \psi - \log(\overline{n})/\alpha = \psi - \log(n/n_i)/\alpha, \tag{2.3a}$$

$$\phi_p = \log(\overline{p})/\alpha + \psi = \log(p/n_i)/\alpha + \psi. \tag{2.3b}$$

In the new set of variables (ψ,ϕ_n,ϕ_p) the quantities all have the same dimension. Now (2.2) can be rewritten as

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

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

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

where $\lambda^2 = \epsilon/qn_i$.

In view of (2.4), we introduce the notation $J_\psi = \lambda^2 \operatorname{grad}\psi$, $J_n = \mu_n e^{\alpha\psi-\alpha\phi_n} \operatorname{grad}(\alpha\phi_n)$, $J_p = \mu_p e^{\alpha\phi_p-\alpha\psi} \operatorname{grad}(\alpha\phi_p)$. Let $S \subset [L(\Omega)]^3$ be the set of all functions (ψ,ϕ_n,ϕ_p) such that $J_\psi, J_n, J_p \in H(\operatorname{div},\Omega)$, and $\overline{n},\overline{p} \in L^2(\Omega)$, then for arbitrary $(\psi,\phi_n,\phi_p) \in S$ and $\Omega_\alpha \subset \Omega$ we find

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

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

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

where $\vec{\nu}$ 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 $N: S \to V = [L^2(\Omega)]^3$ is the nonlinear differential operator in the left hand side of eq.(2.4), $r(q)$ is the right hand side and q denotes the vector of unknown functions $q = (\psi,\phi_n,\phi_p)$.

To understand the character of these equations we study the principle and sub-principle parts of the operator $(N - r)$. This will tell us what boundary conditions are appropriate and which parts of the operator are locally dominant [4] . This is important for designing the discretisation scheme and, as the multigrid algorithm will be used for the solution of the discrete equations, it gives guidelines for the construction of the relaxation procedure and the grid-transfer operators.

The principle part is the part of the linearised operator which contributes to the highest order term of its determinant. Comparing the equations (2.2) and (2.4) we see that the operator $N - r$ is best linearised with respect to the variables $(\psi,\bar{n},\bar{p})$. The linearised operator is given by

$$(N-r)'(q) = \frac{\partial(N-r)}{\partial(\psi,\phi_n,\phi_p)} = \frac{\partial(N-r)}{\partial(\psi,\bar{n},\bar{p})} \cdot \frac{\partial(\psi,\bar{n},\bar{p})}{\partial(\psi,\phi_n,\phi_p)}, \tag{2.7}$$

where

$$\frac{\partial(\psi,\bar{n},\bar{p})}{\partial(\psi,\phi_n,\phi_p)} = \begin{bmatrix} 1 & 0 & 0 \\ \alpha\bar{n} & -\alpha\bar{n} & 0 \\ -\alpha\bar{p} & 0 & \alpha\bar{p} \end{bmatrix} \tag{2.8}$$

and

$$\frac{\partial(N-r)}{\partial(\psi,\bar{n},\bar{p})} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mu_n & 0 \\ 0 & 0 & \mu_p \end{bmatrix} \cdot \tag{2.9}$$

$$\cdot \begin{bmatrix} -\epsilon\Delta\star & q\star & -q\star \\ \alpha n\Delta\star + \alpha\nabla n\cdot\nabla\star & -\Delta\star + \alpha(\nabla\psi\cdot\nabla\star + \star\Delta\psi) & 0 \\ -\alpha p\Delta\star - \alpha\nabla p\cdot\nabla\star & 0 & -\Delta\star - \alpha(\nabla\psi\cdot\nabla\star + \star\Delta\psi) \end{bmatrix}.$$

Hence,

$$\det (N - r)'(q) =$$

$$= -\alpha^2 \bar{n}\bar{p}\mu_n\mu_p \cdot \{-\Delta\star\}\{\epsilon[(\Delta\star)^2 - \alpha^2(\nabla\psi\cdot\nabla\star)^2] - q\alpha(n+p)\star\Delta\star\} + \text{LOT.}$$

The lower order terms (LOT) contain only derivatives of 3rd and lower order. We see that the principle and sub-principle part in the determinant are characterised by the operator

$$\{\Delta\star\}\{\lambda^2(\Delta\star - \alpha\nabla\psi\cdot\nabla\star)(\Delta\star + \alpha\nabla\psi\cdot\nabla\star) - \alpha(\bar{n}+\bar{p})\star\Delta\star\}. \tag{2.10}$$

The equations are elliptic, and for small λ they are elliptic-elliptic singularly perturbed. For large values of $\|\nabla\psi\|$, strong convection terms appear. This convection acts both in the $\nabla\psi$ and in the $-\nabla\psi$ direction. Both directions act as sub-characteristics. Not only smooth functions but, for large $\|\nabla\psi\|$, also non-smooth functions of the form $e^{\pm\vec{x}\cdot\nabla\psi}$ may contribute to small residuals. We notice that in the matrix $(N-r)'(q)$ the 2nd order differentials are found on the main diagonal exclusively.

It is clear that the equations require a careful discretisation. Although the above considerations may also give guidelines for the discretisation in more dimensions, in the remainder of this paper we shall consider the equation and its discretisation for a single space dimension only.

3. The discretisation

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) $\Omega_i = (x_{i-1}, x_i)$, $i = 1,...,N$. Inside each box Ω_i we select a point $x_{i-1/2}$ 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 $D_i = (x_{i-1/2}, x_{i+1/2})$, $i = 1,...,N-1$, $x_{i-1/2} = (x_{i-1} + x_i)/2$ or $x_{i-1/2} = \gamma((i-1/2)/N)$, $D_0 = (x_0, x_{1/2})$, $D_N = (x_{N-1/2}, x_N)$. These intervals form the set of *dual boxes*. Thus, for a given function γ, sets $\{\Omega_i\}_{i=1,...,N}$, and $\{D_i\}_{i=0,...,N}$ 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 $\Omega_h = \{\Omega_i | i = 1,2,...,N\}$.

A discrete representation $q_h \in S_h$ of the state of the semiconductor is given by the $3N$-dimensional vector $q_h = \{q_i\}_{i=1,...,N} = \{(\psi_i, \phi_{n,i}, \phi_{p,i})\}_{i=1,...,N}$. 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

$$J_{\psi,i} = \lambda^2 \frac{\psi_{i+1} - \psi_i}{x_{i+1/2} - x_{1-1/2}}, \tag{3.1}$$

and

$$\begin{aligned} J_{n,i} &= \mu_n \exp(\alpha\psi - \alpha\phi_n)\nabla(\alpha\phi_n), \quad \text{on } D_i, \\ J_{n,i}\nabla\exp(-\alpha\psi) &= (\nabla\alpha\psi)\mu_n\nabla\exp(-\alpha\phi_n), \\ J_{n,i} &= \mu_n \frac{\exp(-\alpha\phi_{n,i+1}) - \exp(-\alpha\phi_{n,i})}{\exp(-\alpha\psi_{i+1}) - \exp(-\alpha\psi_i)} \cdot \frac{\alpha\psi_{i+1} - \alpha\psi_i}{x_{i+1/2} - x_{i-1/2}}. \end{aligned} \tag{3.2}$$

Similarly an expression is found for $J_{p,i}$,

$$J_{p,i} = \mu_p \frac{\exp(+\alpha\phi_{p,i+1}) - \exp(+\alpha\phi_{p,i})}{\exp(+\alpha\psi_{i+1}) - \exp(+\alpha\psi_i)} \cdot \frac{\alpha\psi_{i+1} - \alpha\psi_i}{x_{i+1/2} - x_{i-1/2}}. \tag{3.3}$$

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

$$\iint_{\Omega_i} (e^{\alpha(\phi_p - \psi)} - e^{\alpha(\psi - \phi_n)} + \bar{D}) \, d\Omega \approx \tag{3.4}$$

$$\approx (e^{\alpha(\phi_{p,i} - \psi_i)} - e^{\alpha(\psi_i - \phi_{n,i})} + \bar{D}(x_{i-1/2})) \cdot (x_i - x_{i-1}),$$

and the left hand side is computed as

$$-\int_{\Omega_i} J_\psi \vec{\nu} \, d\Gamma = -J_\psi|_{x_{i-1}}^{x_i} = \frac{\lambda^2(\psi_i - \psi_{i-1})}{x_{i-1/2} - x_{i-3/2}} - \frac{\lambda^2(\psi_{i+1} - \psi_i)}{x_{i+1/2} - x_{i-1/2}}, \tag{3.5a}$$

$$-\int_{\Omega_i} J_n \vec{\nu} \, d\Gamma = -J_n|_{x_{i-1}}^{x_i} = J_{n,i-1} - J_{n,i}, \tag{3.5b}$$

$$-\int_{\Omega_i} J_p \vec{\nu} \, d\Gamma = -J_p|_{x_{i-1}}^{x_i} = J_{p,i-1} - J_{p,i}. \tag{3.5c}$$

A careful implementation of the scheme, avoiding rounding errors for small differences, is obtained by computing

$$J_{n,i} = \mu_n \, \mathrm{dxp}(-\alpha\psi_{i+1}, -\alpha\psi_i, -\alpha\phi_{n,i+1}, -\alpha\phi_{n,i}) \cdot \frac{\phi_{n,i+1} - \phi_{n,i}}{x_{i+1/2} - x_{i-1/2}}, \tag{3.6a}$$

$$J_{p,i} = \mu_p \, \mathrm{dxp}(+\alpha\psi_{i+1}, +\alpha\psi_i, +\alpha\phi_{p,i+1}, +\alpha\phi_{p,i}) \cdot \frac{\phi_{p,i+1} - \phi_{p,i}}{x_{i+1/2} - x_{i-1/2}}, \tag{3.6b}$$

where the function dxp is defined by

$$\mathrm{dxp}(a,b,c,d) = \frac{a - b}{\exp(a) - \exp(b)} \cdot \frac{\exp(c) - \exp(d)}{c - d}. \tag{3.7}$$

Analogous to (2.6), the discrete equation obtained from (3.4), (3.5), (3.6) is written in symbolic form as

$$N_h(q_h) = r_h(q_h), \tag{3.8}$$

where $N_h : S_h \to V_h = \mathbb{R}^{3N}$ denotes the nonlinear difference operator and $r_h : S_h \to V_h$ is the discrete right-hand-side.

For the solution of the system of nonlinear discrete equations thus obtained, we need also derivatives of the equations with respect to the discrete variables. These derivatives are easily computed if the following relations are used

$$\frac{\partial}{\partial a} \mathrm{dxp}(a,b,c,d) = -\mathrm{cxp}(b-a) \, \mathrm{dxp}(a,b,c,d),$$

$$\frac{\partial}{\partial b} \mathrm{dxp}(a,b,c,d) = -\mathrm{cxp}(a-b) \, \mathrm{dxp}(a,b,c,d), \tag{3.9}$$

$$\frac{\partial}{\partial c}\ \mathrm{dxp}(a,b,c,d) = +\mathrm{cxp}(d-c)\ \mathrm{dxp}(a,b,c,d),$$

$$\frac{\partial}{\partial d}\ \mathrm{dxp}(a,b,c,d) = +\mathrm{cxp}(c-d)\ \mathrm{dxp}(a,b,c,d),$$

where cxp is defined by

$$\mathrm{cxp}(z) = z^{-1} + (1 - e^{z})^{-1}. \tag{3.10}$$

The treatment of the boundary conditions is completely analogous to the discretisation in the interior of the domain. For the Dirichlet boundary conditions a state $q_B = (\psi_B, \phi_{n,B}, \phi_{p,B})$ is prescribed at the boundary. In the dual half cells D_0 and D_N, the fluxes J_ψ, J_n, and J_p are assumed to be constant and computed by the expressions (3.1), (3.2) and (3.3), where the size of the dual cell, $x_{i-1/2} - x_{i+1/2}$, is replaced by $x_{1/2} - x_0$ or $x_N - x_{N-1/2}$ respectively. For Dirichlet boundary conditions this describes the numerical boundaries completely. For Neumann boundary conditions the fluxes J_ψ, J_n, and J_p are directly available.

4. Nested discretisations

In fact, by the above construction we have derived a cell-centered version of the well-known Scharfetter-Gummel scheme. What is important is the derivation of this scheme as a Galerkin or weighted residual method. We can define a residual weighting, or *restriction operator*, $\bar{R}_h : V \to V_h$ by

$$\bar{R}_h u = u_h \tag{4.1}$$

where

$$(u_h)_i = \int\!\!\int_{\Omega_i} u(x)\, d\Omega, \qquad i = 1,2,\ldots,N.$$

An interpolation, or *prolongation operator*, $P_h : S_h \to S$ is defined by the assumptions (i) that $(P_h q_h)(x_{i-1/2}) = q_i$, (ii) that $P_h q_h$ has piecewise constant fluxes J_ψ, J_n and J_p on the dual mesh $\{D_i\}$, and (iii) that $P_h q_h$ satisfies the boundary conditions for (2.2) at x_0 and x_N. For (ψ,ϕ_n,ϕ_p) such that $P_h q_h = q = (\psi,\phi_n,\phi_p)$, this implies that q satisfies the boundary conditions, and that ψ is a piecewise linear function which interpolates the values $\{\psi_i\}_{i=1,\ldots,N}$. For ϕ_n and ϕ_p it leads to piecewise exponential interpolation, as is derived from (3.2.a). We find for $x \in D_i$

$$\exp(-\alpha\phi_n(x)) - \exp(-\alpha\phi_{n,i}) = \tag{4.2}$$

$$= (\exp(-\alpha\psi(x)) - \exp(-\alpha\psi_i)) \cdot \frac{\exp(-\alpha\phi_{n,i+1}) - \exp(-\alpha\phi_{n,i})}{\exp(-\alpha\psi_{i+1}) - \exp(-\alpha\psi_i)}.$$

This formula gives a kind of exponential interpolation formula for $\phi_n(x)$, interpolating the values $\{\phi_{n,i}\}_{i=1,\ldots,N-1}$. A similar formula is found for ϕ_p:

$$\exp(\alpha\phi_p(x)) - \exp(\alpha\phi_{p,i}) = (\exp(\alpha\psi(x)) - \exp(\alpha\psi_i)) \cdot \frac{\exp(\alpha\phi_{p,i+1}) - \exp(\alpha\phi_{p,i})}{\exp(\alpha\psi_{i+1}) - \exp(\alpha\psi_i)}.$$

The discrete operator $N_h q_h$ was constructed as the Galerkin operator $N_h = \bar{R}_h N(P_h q_h)$. Notice that the complete discretisation of $(N_h - r_h)$ is not a true Galerkin approximation because of the quadrature approximation (3.4).

Given Ω_H, a discretisation of Ω, we can construct a sequence of finer and finer discretisations by successively doubling the number of boxes. Thus, we obtain $\Omega_{H/2}$, $\Omega_{H/4}$ etc.. In these discretisations all *boxes are nested*, i.e. a single box on a coarser discretisation contains a number of complete boxes in a finer discretisation. Notice that the corresponding dual boxes are *not* nested.

For each discretisation Ω_h in the sequence, we have spaces S_h, V_h and operators P_h, $\bar{R}_h$, N_h and r_h. Based on (4.1), a restriction operator $\bar{R}_{2h,h}: V_h \to V_{2h}$ can be introduced by

$$(\bar{R}_{2h,h} v_h)_i = \sum_{\substack{j \\ \Omega_{h,j} \subset \Omega_{2h,i}}} v_{h,j}, \tag{4.3}$$

which satisfies the relation $\bar{R}_{2h,h} \bar{R}_h = \bar{R}_{2h}$.
Because the dual boxes are not nested, we can not find a prolongation $P_{h,2h}: S_{2h} \to S_h$ that satisfies the similar relation $P_{2h} = P_h P_{h,2h}$. Nevertheless we can construct a $P_{h,2h}$ such that

$$\bar{R}_{2h,h} N_h (P_{h,2h} q_{2h}) = N_{2h}(q_{2h}) \tag{4.4}$$

for all $q_{2h} \in S_{2h}$.

A sufficient condition for (4.4) to hold is that for $J = J_\psi, J_n, J_p$ the following equality holds

$$\sum_{\substack{j \\ \Omega_{h/2,j} \subset \Omega_{h,i}}} \int_{\Omega_{h/2,j}} J \vec{\nu} \, d\Gamma = \int_{\Omega_{h,i}} J \vec{\nu} \, d\Gamma. \tag{4.5}$$

This is achieved by the construction of a $P_{h/2,h}$ such that J at the boundary of all coarse boxes $\Omega_{h,i}$ is the same when derived from either $q_{h/2}$ or $P_{h/2,h} q_h$; i.e.

$$J(q_h)|_{x=x_i} = J(P_{h/2,h} q_h)|_{x=x_i} \tag{4.6}$$

at the boundary x_i of $\Omega_{h,i}$, for all i.

The construction of such a $P_{h/2,h}$ can be based on the interpolation $P_h: S_h \to S$ described in (4.2). Let x_i be a boundary point in Ω_h and let $x_{i\pm 1/2}$ be the boundaries of the corresponding dual box D_i. Then, on the refined mesh $\Omega_{h/2}$ the intervals $(x_{i-1/2}, x_i)$ and $(x_i, x_{i+1/2})$ are boxes, and $(x_{i-1/4}, x_{i+1/4})$ is the dual box corresponding with x_i. The discrete variables on Ω_h and $\Omega_{h/2}$ are respectively denoted by $q_h = \{(\psi(x_{i-1/2}), \phi_n(x_{i-1/2}), \phi_p(x_{i-1/2})), x \in \Omega, i = 0, 1, ..., N\}$ and $q_{h/2} = \{(\psi(x_{i\pm 1/4}), \phi_n(x_{i\pm 1/4}), \phi_p(x_{i\pm 1/4}), x \in \Omega, i = 0, 1, ..., N\}$.

Using the interpolation formulas (4.2) we see that J is constant on $(x_{i-1/2}, x_{i+1/2})$ and, hence, (4.6) holds if

$$\frac{\psi(x_{i\pm 1/4}) - \psi(x_{i-1/2})}{x_{i\pm 1/4} - x_{i-1/2}} = \frac{\psi(x_{i+1/2}) - \psi(x_{i-1/2})}{x_{i+1/2} - x_{i-1/2}},$$

$$\frac{\exp(-\alpha\phi_n(x_{i\pm 1/4})) - \exp(-\alpha\phi_n(x_{i-1/2}))}{\exp(-\alpha\psi(x_{i\pm 1/4})) - \exp(-\alpha\psi(x_{i-1/2}))} = \frac{\exp(-\alpha\phi_n(x_{i+1/2})) - \exp(-\alpha\phi_n(x_{i-1/2}))}{\exp(-\alpha\psi(x_{i+1/2})) - \exp(-\alpha\psi(x_{i-1/2}))},$$

$$\frac{\exp(+\alpha\phi_p(x_{i\pm 1/4})) - \exp(+\alpha\phi_p(x_{i-1/2}))}{\exp(+\alpha\psi(x_{i\pm 1/4})) - \exp(+\alpha\psi(x_{i-1/2}))} = \frac{\exp(+\alpha\phi_p(x_{i+1/2})) - \exp(+\alpha\phi_p(x_{i-1/2}))}{\exp(+\alpha\psi(x_{i+1/2})) - \exp(+\alpha\psi(x_{i-1/2}))}.$$

Hence, identifying $\psi(x_{i-1/2}) = \psi_i^h$, $\phi_n(x_{i-1/2}) = \phi_{n,i}^h$, $\phi_p(x_{i-1/2}) = \phi_{p,i}^h$, $\psi(x_{i-1/4}) = \psi_{2i}^{h/2}$, $\phi_n(x_{i-1/4}) = \phi_{n,2i}^{h/2}$, $\phi_p(x_{i-1/4}) = \phi_{p,2i}^{h/2}$, $\psi(x_{i+1/4}) = \psi_{2i+1}^{h/2}$, $\phi_n(x_{i+1/4}) = \phi_{n,2i+1}^{h/2}$, $\phi_p(x_{i+1/4}) = \phi_{p,2i+1}^{h/2}$, we obtain the prolongation operator $P_{h/2,h}$: $S_{h/2} \to S_h$ by $q_{h/2} = P_{h/2,h}\, q_h$.

Assuming an equidistant spacing between $(x_{i-1/2}, x_{i-1/4}, \tilde{x}_i, x_{i-1/4}, x_{i-1/2})$, we find the interpolation rules

$$\psi_{2i}^{fine} = \frac{3}{4}\psi_i^{coarse} + \frac{1}{4}\psi_{i+1}^{coarse} ,$$

$$\psi_{2i+1}^{fine} = \frac{1}{4}\psi_i^{coarse} + \frac{3}{4}\psi_{i+1}^{coarse} ,$$

$$\phi_{n,2i}^{fine} = \frac{1}{\alpha}\log\left[\frac{(1 + e^m + e^{2m})\Phi_{n,i}^{coarse} + e^{3m}\Phi_{n,i+1}^{coarse}}{(1 + e^{2m})(1 + e^m)}\right] ,$$

$$\phi_{n,2i+1}^{fine} = \frac{1}{\alpha}\log\left[\frac{e^{-3m}\Phi_{n,i}^{coarse} + (e^{-2m} + e^{-m} + 1)\Phi_{n,i+1}^{coarse}}{(1 + e^{-2m})(1 + e^{-m})}\right] ,$$

$$\phi_{p,2i}^{fine} = \frac{1}{\alpha}\log\left[\frac{(1 + e^{-m} + e^{-2m})\Phi_{p,i}^{coarse} + e^{-3m}\Phi_{p,i+1}^{coarse}}{(1 + e^{-2m})(1 + e^{-m})}\right] ,$$

$$\phi_{p,2i+1}^{fine} = \frac{1}{\alpha}\log\left[\frac{e^{3m}\Phi_{p,i}^{coarse} + (e^{2m} + e^m + 1)\Phi_{p,i+1}^{coarse}}{(1 + e^{2m})(1 + e^m)}\right] .$$

where $\Phi_n = \exp(-\alpha\phi_n)$, $\Phi_p = \exp(\alpha\phi_p)$, and where $m = \alpha(\psi_{i+1}^{coarse} - \psi_i^{coarse})/4$. Expressed in ϕ_n and ϕ_p, this is

$$\phi_{n,2i}^{fine} = \phi_{n,i}^{coarse} - \frac{1}{\alpha}\log\left[\frac{(1+e^m+e^{2m}) + e^{3m}\exp(-\alpha(\phi_{n,i+1}^{coarse} - \phi_{n,i}^{coarse}))}{(1+e^m)(1+e^{2m})}\right] ,$$

$$\phi_{n,2i+1}^{fine} = \phi_{n,i}^{coarse} - \frac{1}{\alpha}\log\left[\frac{1+(e^m+e^{2m}+e^{3m})\exp(-\alpha(\phi_{n,i+1}^{coarse} - \phi_{n,i}^{coarse}))}{(1+e^m)(1+e^{2m})}\right] ,$$

$$\phi_{p,2i}^{fine} = \phi_{p,i}^{coarse} + \frac{1}{\alpha}\log\left[\frac{(e^m+e^{2m}+e^{3m}) + \exp(\alpha(\phi_{p,i+1}^{coarse} - \phi_{p,i}^{coarse}))}{(1+e^m)(1+e^{2m})}\right] ,$$

$$\phi_{p,2i+1}^{fine} = \phi_{p,i}^{coarse} + \frac{1}{\alpha}\log\left[\frac{e^{3m}+(1+e^m+e^{2m})\exp(\alpha(\phi_{p,i+1}^{coarse} - \phi_{p,i}^{coarse}))}{(1+e^m)(1+e^{2m})}\right] .$$

For large values of $|m|$, this simply reduces to a special upwind interpolation:

$$\phi_{n,2i}^{fine} = \phi_{n,2i+1}^{fine} = \phi_{n,i+1}^{coarse} ,$$
$$\phi_{p,2i}^{fine} = \phi_{p,2i+1}^{fine} = \phi_{p,i}^{coarse} \quad \text{if } m \gg 1 ,$$

and

$$\phi_{n,2i}^{fine} = \phi_{n,2i+1}^{fine} = \phi_{n,i}^{coarse} ,$$
$$\phi_{p,2i}^{fine} = \phi_{p,2i+1}^{fine} = \phi_{p,i+1}^{coarse} \quad \text{if } m \ll 1 .$$

5. The Newton Method (Schilders' Correction Transformation)

As part of the non-linear solution process, a Newton method is used and some linearised systems have to be solved. An important feature in our method is the *correction transformation* a device introduced for the solution of the semiconductor equations in [9] . The dependent variables in our computation are (ψ,ϕ_n,ϕ_p). However, the equations have a less strongly nonlinear behaviour when expressed in the variables (ψ,n,p) or (ψ,Φ_n,Φ_p)), where $\Phi_n = \exp(-\alpha\phi_n)$, $\Phi_p = \exp(\alpha\phi_p)$. Therefore, the Newton process can better be based on linearisation with respect to (ψ,n,p) or (ψ,Φ_n,Φ_p). The correction transformation is the technique to transform the correction $(d\psi,d\phi_n,d\phi_p)$ computed by linearisation with respect to (ψ,ϕ_n,ϕ_p), to the correction for the same variables that would have been obtained when linearisation were applied to the better behaving variables.

We show this first for the variables (ψ,n,p). From eq. (2.8) we see

$$\begin{aligned} d\psi &= d\psi, \\ d\bar{n} &= \alpha\bar{n}(d\psi - d\phi_n), \\ d\bar{p} &= \alpha\bar{p}(d\phi_p - d\psi), \end{aligned} \tag{5.1}$$

From this we derive

$$\begin{aligned} \psi^{(n+1)} &= \psi^{(n)} + d\psi \\ \bar{n}^{(n+1)} &= \bar{n}^{(n)}(1 + \alpha(d\psi - d\phi_n)), \\ \bar{p}^{(n+1)} &= \bar{p}^{(n)}(1 + \alpha(d\phi_p - d\psi)), \\ \phi_n^{(n+1)} &= \phi_n^{(n)} + d\psi - \log(1 + \alpha(d\psi - d\phi_n))/\alpha, \\ \phi_p^{(n+1)} &= \phi_p^{(n)} + d\psi + \log(1 + \alpha(d\phi_p - d\psi))/\alpha, \end{aligned} \tag{5.2}$$

Similarly, the correction can be transformed for linearisation with respect to $\psi,\Phi_n,\Phi_p)$. Then we obtain

$$\begin{aligned} \Phi_n^{(n+1)} &= \Phi_n^n(1 - \alpha d\phi_n), \\ \Phi_p^{(n+1)} &= \Phi_p^n(1 + \alpha d\phi_p), \\ \phi_n^{(n+1)} &= \phi_n^n - \log(1 - \alpha d\phi_n)/\alpha, \\ \phi_p^{(n+1)} &= \phi_p^n + \log(1 + \alpha d\phi_p)/\alpha. \end{aligned} \tag{5.3}$$

It is clear that for small corrections the correction transformation has a negligible effect. For larger corrections the effect will be stronger. If the argument of the logarithm is larger than one, the original correction will be damped. For arguments smaller than one, the correction is blown up. However, large corrections may yield negative arguments for the logarithmic function. This will happen in cases where linearisation doesn't make much sense. In this case we want to damp the correction. In practice this is done by replacing the function $\log(s)$ in (5.2) or (5.3) by a $C^1(-\infty,\infty)$-function, identical with $\log(s)$ for $s > s_0$, viz.

$$\log(s_0) + \text{sign}(s - s_0)\,|\log(|s - s_0| + s_0) - \log(s_0)| \tag{5.4}$$

where s_0 is some small positive number. The approximation of this modified $\log(s)$

function for $s<s_0$ (i.e. $s_0 = 0.0000005$) is called *damping* in [8].

6. The Multigrid Method

To solve the nonlinear system

$$M_h(q_h) := N_h(q_h) - r_h(q_h) = f_h \tag{6.1}$$

we use a nonlinear multigrid (FAS) method [3, 6]. For a vanishing right-hand-side f_h this system is the system of equations (3.8). The FAS method is an iterative process, in which each cycle consists of:

1. a number of p nonlinear relaxation sweeps;
2. a coarse grid correction;
3. another q nonlinear relaxation sweeps.

As a *relaxation* procedure we use a nonlinear Collective Symmetric Gauss Seidel (CSGS) relaxation. In this procedure all boxes are successively scanned in forward and backward direction, and for each box in its turn the 3 nonlinear equations are (approximately) solved. The *coarse grid correction* consists of the following steps

$$d_{2h} = \bar{R}_{2h,h}(f_h - M_h(q_h^{(n)})), \tag{6.2a}$$

$$M_{2h}(\tilde{q}_{2h}) = M_{2h}(q_{2h}) + d_{2h}/\mu, \tag{6.2b}$$

$$q_h^{(n+1)} = q_h^{(n)} + \mu(P_{h,2h}\tilde{q}_{2h} - P_{h,2h}q_{2h}). \tag{6.2c}$$

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 (6.2.b), or it may be an approximated by a number of σ multigrid cycles for the solution of (6.2.b) applied to the initial approximation q_{2h}. In this way a recursive procedure is obtained in which a sequence of coarser and coarser grids is used. Only on the coarsest grid a (smaller) nonlinear system is to be solved by other means. The parameter $\mu \in \mathbb{R}$ is a number to control the right-hand-side in the equation (6.2.b). In our applications we use $\mu = 1$ throughout. The numbers $p, q, \sigma \in \mathbb{N}$ determine the *strategy* of the multigrid method; $\sigma = 1$ defines a V-cycle, $\sigma = 2$ a W-cycle. In most experiments reported in [8] we take a fixed strategy with $p = q = \sigma = 1$. The operators $\bar{R}_{2h,h}$ and $P_{h,2h}$ are described in section 4.

There is a difference between the usual FAS algorithm and the present one, due to the nonlinearity of the prolongation. Generally, the last step in the coarse grid correction is written

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

which is equivalent to (6.2.c) only for a linear prolongation.

With the property $N_{2h} = \bar{R}_{2h,h} N_h P_{h,2h}$ it can be shown [7] that the restriction of the residual will be small after a coarse grid correction. In fact, we see that the restriction of the residual

$$\bar{R}_{2h,h}(f_h - M_h(q_h^{(n+1)})) \approx (\bar{R}_{2h,h} r'_h P_{h,2h} - r'_{2h})(\tilde{q}_{2h} - q_{2h})$$

depends on the integration error in the right-hand-side, which will be at most $\mathcal{O}(h)$. In general $\tilde{q}_{2h} - q_{2h}$ will be at most $\mathcal{O}(h)$, and the restriction of the residual will be $\mathcal{O}(h^2)$.

Because $\bar{R}_{2h,h}$ adds residual components of small boxes to form a residual component of a coarse grid box, a small value of $\bar{R}_{2h,h}(f_h - M_h(q_h^{(n+1)}))$ implies that large components in $f_h - M_h(q_h^{(n+1)})$ must be high-frequency components. The success of the MG method is based on the fact that relaxation methods as CSGS are effective means to efficiently reduce these high frequency components in the error/residual.

7. Relaxation.

For the smoothing step we may consider various relaxations. Based on previous experience with other equations, a good choice seems Collective Symmetric Gauss Seidel relaxation. This relaxation can be performed in different ways. In all cases the boxes are successively scanned, first in the forward later in the backward direction, and in each box in its turn the 3 nonlinear equations are approximately solved. How the solution of these small systems is approximated makes the difference. The first possibility is to use Newtons method. Another possibility is pointwise Gummel iteration. In Gummel iteration, first the variable ψ is solved for fixed values of ϕ_n and ϕ_p, and then ϕ_n and ϕ_p are solved for the new value of ψ. In pointwise Gummel relaxation, this process is iterated for each particular cell until the solution of the nonlinear 3×3 system is obtained with a specified accuracy.

At convergence, the result of this pointwise iteration process and the result of pointwise Newton iteration are the same. Differences are the faster convergence of the Newton process near the solution, and the better global convergence properties of the pointwise Gummel iteration. An additional advantage of Gummel iteration is that no (possibly ill conditioned) linear 3×3 systems have to be solved. The better global convergence can be understood by the stable discretisation of the elliptic scalar equations (2.4b) and (2.4c) as equations in $\exp(-\alpha\phi_n)$ and $\exp(+\alpha\phi_p)$ respectively. This guarantees that for each separate equation the correction in the solution will be bounded in terms of the right-hand-side. For a vanishing right-hand-side the correction equation satisfies a maximum principle. This stability property guarantees that intermediate approximations remain limited to a reasonable neighbourhood of the solution.

The pointwise Gummel iteration allows an additional technique to be used to enhance the convergence. To solve the discrete equivalent of (2.4.a) at a single point, a variation of ψ contributes to the residual by a linear term $\nabla(\lambda^2\nabla\psi)$ end by two exponential terms. If $|\phi_n + \phi_p - 2\psi|$ is large, the influence of one of the exponentials can be neglected and the residual grows exponential with the other. This is an argument to linearise the equation w.r.t. $\exp(+\psi)$ (or $\exp(-\psi)$) instead of w.r.t. ψ, if $\phi_n + \phi_p - 2\psi > (<) 0$. This is a motivation to apply a correction transformation also for the ψ -correction, similar to (5.3)

$$d\psi := |\log(1 + \alpha|d\psi| * \alpha)| / \alpha$$

The discrete analogues of (2.4.b,c) are linear in Φ_n and Φ_p and, hence, the corresponding corrections are transformed by (5.3)

CONCLUSION.

In this paper a multigrid method is described for the solution of the nonlinear system of equations that arises from the discretisation of the equations for numerical semiconductor device simulation. Special attention has been paid to the construction of a suitable prolongation, such that a sequence of nested Galerkin discretisations is obtained for the differential operator on the different levels. Further, the use of Schilders' correction transformation makes collective Gauss Seidel relaxation an efficient residual smoother.

Results obtained with the method will be published elsewhere [8] . They show rapid convergence indeed. A few iterations are sufficient to reduce the iteration error to less than the truncation error. Convergence appears to be mesh independent for a diode model problem with different applied voltages, both forward and reversed biased.

REFERENCES

1. R.E. BANK, J.W. JEROME, and D.J. ROSE (1982). Analytical and numerical aspects of semiconductor device modeling, in *Computing Methods in Applied Sciences and Enginering*, 593 - 597, ed. R. Glowinski and J.L. Lions, North-Holland, Amsterdam.
2. R.E. BANK and H.D. MITTELMANN (1986). Continuation and Multi-Grid for nonlinear elliptic systems, in *Multigrid Methods II*, 23 - 37, ed. W. Hackbusch and U. Trottenberg, Springer Verlag.
3. A. BRANDT (1982). Guide to multigrid development, in *Multigrid Methods*, 220-312, ed. W. Hackbusch and U. Trottenberg, Springer Verlag.
4. A. BRANDT (1982.). Multigrid solutions to steady state compressible Navier Stokes equations I, in *Computing Methods in Applied Science and Engineering*, 407-422, ed. R. Glowinski and J.L. Lions, North Holland Publ. Comp..
5. S.P. GAUR and A. BRANDT (1977). Numerical solution of semiconductor transport equations in two dimensions by multi-grid method, in *Advances in Computer Methods for PDEs II*, 327 - 329, ed. R. Vichnevetsky.
6. W. HACKBUSCH (1985). *Multigrid Methods and Applications*, Springer Verlag.
7. P.W. HEMKER (1988). *A multigrid approach for one-dimensional semiconductor device simulation,* Procs. SUPRENUM meeting, Jan. 1988, GMD, Bonn.
8. P.W. HEMKER (1988). A nonlinear multigrid method for the one-dimensional semiconductor device simulation: results for the diode, in *Proceedings of the Conference on Numerical Methods and Applications*, Sofia, Bulgaria.
9. S.J.POLAK, C.DEN HEIJER, W.H.A.SCHILDERS, and P.MARCOWICH (1987). Semiconductor device modelling from the numerical point of view, *Int. J. Num. Meths Engng.*, 24, 763-838.
10. Q.M. SHEIKH. *Systems of nonlinear algebraic equations arising in simulation of semiconductor devices,* UIUCDCS - R83 -1133, Urbana.
11. A.S. SHIEH (1985). On the solution of coupled system of PDE by a multigrid method, *IEEE Trans. ED*, 10, 2083-2086.
12. S. SLAMET (1981). *Numerical simulation of semiconductor devices,* Rept. UIUCDCS-R-81-1072, Dept. Comp.Sc, Univ.Illinois.
13. S. SLAMET (1983). *Quasi-Newton and multigrid methods for semiconductor device simulation,* UIUCDCS - R83 -1154, Urbana.