International Series of Numerical Mathematics, Vol. 98, © 1991 Birkhäuser Verlag Basel

An adaptive multigrid approach for the solution of the 2D semiconductor equations

P.W. Hemker and J. Molenaar

Abstract

An adaptive multigrid method is presented for the solution of the two-dimensional steady state Van Roosbroeck equations for semiconductor device modeling. The discretisation is based on the (hybrid) mixed finite element method on rectangles. The integrals involved are approximated by the trapezoidal rule. In this way, in the interior of the domain, the classical Scharfetter Gummel discretisation is retained. A 5-point collective Vanka-type relaxation procedure is used as a smoother.

The mixed finite elements give rise to a cell-centered multigrid method and the multigrid grid-transfer operators are chosen in agreement with the discretisation. The main difficulties are the proper use of very coarse grids and the construction of suitable initial estimates. In order to admit very coarse grids, it appears necessary to take special measures and to introduce local damping of the residual in the coarse grid correction.

It is shown that, under these conditions, a fast convergence can be obtained that seems to be independent of the grid size. Hence, in combination with nested iteration, an efficient procedure is obtained. Results are shown for a realistic two-dimensional transistor model.

1 Introduction

The usual mathematical model to describe the electric behaviour of semiconductor devices is the drift diffusion model, that was essentially proposed by Van Roosbroeck [18] in 1950. It is given by a nonlinear system of three second order partial differential equations. Let $\Omega \subset I\!R^n$, n=1,2,3, be an open bounded region with a piecewise smooth boundary $\partial\Omega$, then, scaled to dimensionless form, the equations are [11] [15]

$$\begin{aligned} &\text{div } (\lambda^2 \text{grad } \psi) = n - p - D, \\ &\frac{\partial n}{\partial t} = \text{div } (\mu_n(\text{grad } n - n \,\text{grad } \psi - n \,\text{grad } \log n_i)) - R, \\ &\frac{\partial p}{\partial t} = \text{div } (\mu_p(\text{grad } p + p \,\text{grad } \psi - p \,\text{grad } \log n_i)) - R. \end{aligned} \tag{1}$$

Here the dependent variables n and p, denote the local density of free electrons and holes in the device respectively, and ψ is the electrostatic potential. These variables are functions of $x \in \Omega$ and $t \geq 0$; λ^2 is associated with the dielectric constant; D(x), the net doping function, describes the location of the impurities that characterises the device. Also n_i, the effective intrinsic carrier density, is a function of x; μ_n and μ_p the carrier mobilities are functions of x, n, p, grad ψ, and the net recombination rate R, that models the recombination and generation of electrons and holes, is a function of n and p. The first equation in (1) is the

Poisson equation for the electric field, whereas the second and the third are the continuity equations for the electrons and holes.

The usual approach for the numerical solution of (1) is the application of a box method (finite volume method) for the discretisation, where the Scharfetter-Gummel exponential fitting scheme is used for the approximation of the fluxes between the control volumes. Damped Newton methods are generally used for the solution of the discrete nonlinear systems. For details about the equations and the techniques for their numerical solution, we refer to [11] and [15]. In this paper we investigate a nonlinear multigrid technique for the solution of eq.(1) in order to see whether it could be advantageously applied.

In order to reduce the large number of technical difficulties involved, we restrict ourselves to the computation of the steady state, and we assume the intrinsic carrier density n_i and the mobilities μ_n and μ_p to be constant. With these simplifications, still many of the essential difficulties remain. In the first place the equations are singularly perturbed because the parameter λ, that can be related with the Debye length of the device, is generally small compared with the size of the device. Moreover, a strong convective behaviour of the equations is caused by the possibly large coefficient grad ψ in the continuity equations.

In the past, several attempts have already been made to explore the possibilities of multigrid techniques for the solution of the equations (1). However, up to now the question of whether the multigrid technique is feasible for practical application for these equations is still open. It appears that the use of multigrid is not straightforward at all and that a number of difficulties are encountered with its application. In this paper we want to show some progress made towards an applicable MG method for semiconductor device modeling.

Adaptive grids and nested meshes

In this paper we apply the nonlinear multigrid approach to the solution of the discretised equation (1). The difficulties lie in the bad scaling, the strong nonlinearity and the singular perturbation character. Because of the singular perturbation behaviour, sharp shifts will appear in the solution, whereas in other regions the variables vary only gradually. This makes it unfeasible to represent the solution on a regular mesh. A priori some physical insight may be available about the location of the various regions, but the true solution is only known after numerical approximation. Therefore automatic adaptive local mesh refinements are introduced. This is done in a more or less straightforward sense, as basicly treated e.g in Brandt [2] or McCormick [12]. One difference is that the method is now applied in its cell-centered form. The solution is represented by its values at cell centers and cells are divided into smaller cells to obtain finer meshes. This approach is advantageous in the case conservation laws play a role and it has direct consequences for the transfer operators used.

The adaptive algorithm is flexible in the sense that it allows a completely arbitrary refinement of already existing cells. This is done by allowing any rectangular cell to be divided into four smaller rectangles of the same shape. The smaller cells are part of the next level of refinement in the sequence of discretisations used for the multigrid method. In this way all cells belong to a quad-tree structure and each cell has at most four neighbours on the same refinement level. The same quad-tree structure is used in the program to store the data. The domain of definition for the equations needs to be covered only by the very coarsest grid. Finer grids may cover the domain only partially. This freedom allows another (independent) algorithm to take full responsibility of the grid refinement procedure.

Multigrid

In the first place multigrid can be used for nested iterations, i.e. to obtain initial estimates for the solution on a coarser grid than the one that is required to give an accurate representation of the solution. The usual way of solving the equation for a set of boundary conditions, is by starting at a zero bias and incrementing the boundary values in small steps until the desired conditions are reached. Such a continuation process requires a number of intermediate calculations that are best made on a grid that is as coarse as possible. As soon as the problem has been approximately solved on the coarse mesh, the mesh can be refined to yield higher accuracy.

The iteration process to solve the problem on each level might be an approximate Newton method where multigrid is used to solve the linear systems. For this approach see e.g. [1]. A drawback of global linearisation is the time-consuming evaluation of Jacobian entries and the large memory requirements to store them. This can be avoided by the use of nonlinear multigrid where linearisations are made only locally. We are interested in such nonlinear multigrid techniques for the solution of the large nonlinear systems.

Because of the strong local variations in the solution, one difficulty to deal with is to know what information, available from a very coarse grid solution, may still be useful for the acceleration of the convergence on the fine grid. Such problems were also known from CFD problems with shocks, where -for the Euler equations- these difficulties could be solved by strict adherence to the discrete conservation laws. For the semiconductor equations this problem appears to be much harder because source terms, that are related to approximate truncation errors, may be so large that -without special measures- no longer a (positive) solution for the coarse grid problem can be guaranteed.

2 The equations

After the simplifications, mentioned in Section 1, the equations to be solved are

$$\begin{array}{lll} \text{div } J_\psi - n + p + D = 0, & J_\psi = \lambda^2 \text{grad } \psi, \\ \text{div } J_n - R = 0, & J_n = +\mu_n(\text{grad } n - n\text{grad } \psi), \\ \text{div } J_p + R = 0, & J_p = -\mu_p(\text{grad } p + p\text{grad } \psi). \end{array} \tag{2}$$

In shorthand we write these equations also as $N(\psi, n, p) = 0$. For the recombination rate we assume the Shockley-Read-Hall model

$$R = \frac{np - 1}{\tau_p(n+1) + \tau_n(p+1)}. \tag{3}$$

In order to bring the variables to quantities of the same dimension, it is useful to introduce the quasi-Fermi potentials ϕ_n and ϕ_p by

$$\begin{array}{l} n = \exp(\psi - \phi_n), \\ p = \exp(\phi_p - \psi). \end{array} \tag{4}$$

As a starting point for the discretisation, we use the Slotboom variables

$$\begin{array}{l} \Phi_n = \exp(-\phi_n), \\ \Phi_p = \exp(+\phi_p), \end{array} \tag{5}$$

for which the equations appear in symmetric positive definite form:

$$\begin{aligned} &-\text{div}\,(\mu_\psi \text{grad}\ \psi) + Q = D, \\ &-\text{div}\,(\mu_n \exp(+\psi)\text{grad}\ \Phi_n) + R = 0, \\ &-\text{div}\,(\mu_p \exp(-\psi)\text{grad}\ \Phi_p) + R = 0, \end{aligned} \tag{6}$$

where we use the notation $\mu_\psi = \lambda^2$, $Q = e^{\psi}\Phi_n - e^{-\psi}\Phi_p$. The boundary conditions are of Dirichlet type (ψ, ϕ_n, ϕ_p given) at the Ohmic contacts, and homogeneous Neumann conditions ($J_\psi = J_n = J_p = 0$) at the remaining parts of the boundary.

3 The discretisation by Mixed Finite Elements

Each of the equations (6) can be cast in the form

$$\begin{array}{lcll} \left.\begin{array}{l}\boldsymbol{\sigma} \\ \text{div}\,\boldsymbol{\sigma}\end{array}\right\} & \begin{array}{l}= a\,\text{grad}\ u \\ = f(u)\end{array} & \text{on} & \Omega, \\ u & = g & \text{on} & \Gamma_D, \\ \mathbf{n}\cdot\boldsymbol{\sigma} & = 0 & \text{on} & \Gamma_N, \end{array} \tag{7}$$

where Γ_D and Γ_N denote the parts of the boundary with Dirichlet or homogeneous Neumann boundary conditions, respectively. The sign is chosen such that $a(x) > 0$. As a starting point for the discretisation we use its variational form: find $\boldsymbol{\sigma} \in \mathbf{H}^{BC}(\text{div}\ ,\Omega)$ and $u \in L^2(\Omega)$ such that

$$\left\{ \begin{array}{lcl} \int_\Omega a^{-1}\boldsymbol{\sigma}\cdot\mathbf{v}\ d\Omega + \int_\Omega u\ \text{div}\ \mathbf{v}\ d\Omega & = & \oint_{\Gamma_D} g\,\mathbf{v}\cdot\mathbf{n}\ ds\ , \quad \forall \mathbf{v}\in\mathbf{H}^{BC}(\text{div}\ ,\Omega), \\ \int_\Omega \phi\,\text{div}\ \boldsymbol{\sigma}\ d\Omega & = & \int_\Omega \phi\, f(u)\ d\Omega, \quad \forall \phi \in L^2(\Omega), \end{array} \right. \tag{8}$$

where $\mathbf{H}^{BC}(\text{div}\ ,\Omega) = \{\mathbf{v}\in H(\text{div}\ ,\Omega) \mid \mathbf{v}\cdot\mathbf{n} = 0 \text{ on } \Gamma_N\}$.

For the discretisation we assume that Ω can be divided by a regular partitioning in open disjoint rectangular cells Ω_i, $\overline{\Omega} = \cup\overline{\Omega}_i$. We denote by E_j the edges of the rectangles, by ϵ_i the characteristic function on Ω_i, and we use the notation

$$d_{ij} = \left\{ \begin{array}{ll} +1 & \text{if } E_j \text{ is a N- or E-edge of } \Omega_i, \\ -1 & \text{if } E_j \text{ is a S- or W-edge of } \Omega_i, \\ 0 & \text{if } E_j \text{ is not an edge of } \Omega_i. \end{array} \right. \tag{9}$$

By $\mathbf{e}_j \in \mathbf{H}(\text{div}\ ,\Omega)$ we denote the *tent function* for E_j, i.e. a vector function $\mathbf{e}_j$ of which each component is linear on each Ω_i and which satisfies $\mathbf{e}_j \cdot \mathbf{n}_k = \delta_{jk}$, where $\mathbf{n}_k$ is the unit normal on edge E_k (in the positive x- or y- direction); δ_{jk} is the Kronecker delta. We introduce $\bar{\epsilon}_\ell$, a function defined on all edges E_j, by

$$\bar{\epsilon}_\ell(x) = \left\{ \begin{array}{ll} 1 & \text{if } x \in E_\ell, \\ 0 & \text{if } x \notin E_\ell, \end{array} \right. \tag{10}$$

and the *half tent function* $\mathbf{e}_{ij}$ defined by $\mathbf{e}_{ij} = \mathbf{e}_j\epsilon_i$.

We define the discrete spaces

$$\begin{array}{ll} L_h(\Omega) & = \text{Span}(\epsilon_i), \\ M_h & = \text{Span}(\bar{\epsilon}_\ell), \\ \mathbf{H}_h^{BC}(\text{div}\ ,\Omega) & = \text{Span}(\mathbf{e}_i) \bigcap \mathbf{H}^{BC}(\text{div}\ ,\Omega), \\ W_h(\Omega) & = \text{Span}(\mathbf{e}_{ij}). \end{array} \tag{11}$$

The mixed finite element (MFE) discretisation of equation (6) reads:
find $\boldsymbol{\sigma}_h \in \mathbf{H}_h^{BC}(\mathrm{div}\,,\Omega)$ and $u_h \in L_h^2(\Omega)$ such that

$$\begin{cases} \int_\Omega a^{-1}\boldsymbol{\sigma}_h \cdot \mathbf{v}_h \, d\Omega + \int_\Omega u_h \mathrm{div}\, \mathbf{v}_h \, d\Omega &= \oint_{\Gamma_D} g\, \mathbf{v}_h \cdot \mathbf{n}, \quad \forall \mathbf{v}_h \in \mathbf{H}_h^{BC}(\mathrm{div}\,,\Omega), \\ \int_\Omega \phi_h \,\mathrm{div}\, \boldsymbol{\sigma}_h \, d\Omega &= \int_\Omega \phi_h f(u)\, d\Omega, \quad \forall \phi \in L_h^2(\Omega). \end{cases} \tag{12}$$

We notice that f may be a nonlinear function of u, so linearisation yields a discrete linear system of the form

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} \sigma_j \\ u_i \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}, \tag{13}$$

where

$$\begin{array}{lll} a_{k,j} &= \int_\Omega a^{-1}\mathbf{e}_k \cdot \mathbf{e}_j \, d\Omega, & \\ b_{k,i} &= \int_{\Omega_i} \mathrm{div}\, \mathbf{e}_k \, d\Omega & = \sum_j d_{ij} h_j, \\ c_{m,i} &= -\delta_{m,i} \int_{\Omega_m} \frac{\partial f}{\partial u} \, d\Omega, & \\ b_k &= \int_{E_k} g\, \mathbf{e}_k \cdot \mathbf{n} \, d\Gamma & = \sum_i d_{ik} \int_{E_k} g \, d\Gamma, \\ f_m &= \int_{\Omega_m} f(u) \, d\Omega. & \end{array} \tag{14}$$

Here h_j denotes the length of E_j and σ_j and u_i are the coefficients in

$$\boldsymbol{\sigma}_h = \sum_j \sigma_j \, \mathbf{e}_j, \qquad u_h = \sum_i u_i \, \epsilon_i. \tag{15}$$

(Notice that $C = 0$ in classical MFE theory.) In the usual stable case (no avalanche) the sign of f is such that $c_{i,i} \leq 0$. One of the advantages of the mixed finite element method is that the second equation in its discrete form guarantees the property of discrete current conservation.

Lumping, Scharfetter-Gummel

Taking piecewise constant approximations for f and g, all entries in the system (13) are simple to evaluate, except $a_{k,j}$. This coefficient may give rise to problems because in the continuity equations $a(x)$ can be a rapidly varying exponential function. The quadrature used to approximate $a_{k,j}$ is the weighted trapezoidal rule for rectangles

$$\int_{\Omega_i} w(x)\, z(x)\, d\Omega \;\cong \sum_{\nu=1,2,3,4} z(x_\nu) \int_{\Omega_i^\nu} w(x)\, d\Omega,$$

where x_ν are the four vertices and Ω_i^ν are the four quarter rectangles, parts of Ω_i, associated with these vertices respectively. We use $w(x) = a^{-1}$, and $z(x) = \sum_i \mathbf{e}_{ij} \cdot \mathbf{e}_{ik}$. The use of this quadrature rule is called *lumping* because it makes the matrix A diagonal. This is seen e.g. in the case of constant a (Poisson equation), where exact quadrature would yield

$$\int_{\Omega_i} a^{-1}\mathbf{e}_{ij} \cdot \mathbf{e}_{jk} \, d\Omega = \begin{cases} \frac{1}{3}a^{-1}a_i & \text{if } k = j, \\ \frac{1}{6}a^{-1}a_i & \text{if } E_k \text{ and } E_j \text{ are opposite neighbours,} \\ 0 & \text{otherwise,} \end{cases}$$

where $a_i =$ area(Ω_i). In the case of the trapezoidal rule we obtain

$$\int_{\Omega_i} a^{-1}\mathbf{e}_{ij} \cdot \mathbf{e}_{jk} \, d\Omega = \begin{cases} \frac{1}{2}a^{-1}a_i & \text{if } k = j, \\ 0 & \text{otherwise.} \end{cases}$$

It was shown by Schilders [16] that the trapezoidal rule is advantageous, because the lumped form of the discretisation still yields an M-matrix. In the non-lumped case it is easily shown that the matrix obtained after elimination of σ is not necessarily an M-matrix in the case of a non-zero matrix C. Hence stability problems may rise. (In the non-lumped linear one-dimensional case with constant coefficients, $f(u) = f'u$ with $f' > 0$, the matrix is an M-matrix only if $h^2 f'/a \leq 6$!) Therefore, in the remainder of this paper we restrict ourselves to the lumped case only.

By the trapezoidal rule we get the approximation

$$a_{j,k} = \sum_i \int_{\Omega_i} a^{-1} \mathbf{e}_{ij} \cdot \mathbf{e}_{ik} \, d\Omega \cong \sum_i \sum_{\nu=1,2,3,4} (\mathbf{e}_{ij} \cdot \mathbf{e}_{ik})(x_\nu) \int_{\Omega_i^\nu} a^{-1}(x) \, d\Omega =$$

$$= \sum_i \sum_{\nu=1,2,3,4} \delta_{jk} \bar{\epsilon}_k(x_\nu) \int_{\Omega_i^\nu} a^{-1}(x) \, d\Omega = \delta_{jk} \int_{\Omega^k} a^{-1} \, d\Omega,$$

where $\Omega^k = \bigcup_{\{i,\nu | \bar{\Omega}_i^\nu \cap E_k \neq \emptyset\}} \Omega_i^\nu$; i.e. Ω^k is the dual box related with the edge E_k. If we approximate ψ in Ω^k by a linear function, interpolating the values ψ_i from the neighbouring cell centers, then $a = \exp(\pm\psi)$, and we obtain

$$\int_{\Omega^k} a^{-1} \, d\Omega = \text{area}(\Omega_k) \, \text{Bexp}^{-1}(\mp\psi_{i_1}, \mp\psi_{i_2}),$$

where we introduced the function

$$\text{Bexp}(x, y) = \frac{x - y}{e^x - e^y}. \tag{16}$$

Thus we retain the well-known Scharfetter-Gummel scheme (cf. [3]). In [16] it was shown that currents may be computed more accurately by the present MFE method than by the classical box scheme.

We see that, after lumping, the variables σ_j may be eliminated to obtain a five-point difference scheme between the variables u_i. For the discretisation of (6), we apply the above scheme for $u = (\psi, \Phi_n \Phi_p)$, so that $\sigma = (J_\psi, J_n, J_p)$, and $a = (\mu_\psi, \mu_n \exp(\psi), \mu_p \exp(-\psi))$, to obtain the system

$$\begin{aligned}
&\textstyle\sum_j h_j d_{ij} J_{\psi j} = e^{\phi_{p_i} - \psi_i} - e^{\psi_i - \phi_{n_i}} + D(x_i), && J_{\psi j} = -\tfrac{h_j d_{ij}}{a_j} \mu_\psi (\psi_j - \psi_i), \\
&\textstyle\sum_j h_j d_{ij} J_{nj} = +R(\psi_i, \phi_{n_i}, \phi_{p_i}), && J_{nj} = -\tfrac{h_j d_{ij}}{a_j} \mu_n \frac{\text{Bexp}\,(-\psi_j, -\psi_i)}{\text{Bexp}\,(-\phi_{n_j}, -\phi_{n_i})} (\phi_{n_j} - \phi_{n_i}), \\
&\textstyle\sum_j h_j d_{ij} J_{pj} = -R(\psi_i, \phi_{n_i}, \phi_{p_i}), && J_{pj} = -\tfrac{h_j d_{ij}}{a_j} \mu_p \frac{\text{Bexp}\,(\psi_j, \psi_i)}{\text{Bexp}\,(\phi_{p_j}, \phi_{p_i})} (\phi_{p_j} - \phi_{p_i}),
\end{aligned} \tag{17}$$

where $a_j = \text{area}(\Omega_j)$.

Green boundaries

In the case of partially refined grids, green boundaries appear. Green boundaries are those boundaries of a fine grid that are not part of the boundary of the domain Ω. Such green

boundaries (green edges) separate areas where finest cells have different mesh sizes. Here the finer mesh needs an additional boundary condition. Hence, values of the potentials u are needed at edges E_j, to serve as boundary conditions for discretisation on the fine mesh. These values are obtained from the coarse grid by the use of half tent functions as weighting functions in equation (12a), and by forcing sufficient continuity by the introduction of a Lagrange multiplier, as is usual for the hybrid mixed finite element method [3]. We denote these test functions by τ_h^*. The value of the potential at wall E_k, denoted by λ_k, is derived from the variational equation: find $(u_h^*, \sigma_h^*, \lambda_h) \in L_h(\Omega) \times W_h(\Omega) \times M_h(\Omega)$ such that

$$\begin{cases} \sum_i \int_{\Omega_i} a^{-1} \boldsymbol{\sigma}_h^* \cdot \tau_h^* + \sum_i \int_{\Omega_i} u_h^* \mathrm{div}\, \tau_h^* & = \sum_i \oint_{\partial\Omega_i} \lambda_h \tau_h^* \cdot \mathbf{n}_i, \\ \sum_i \int_{\Omega_i} \phi_h \mathrm{div}\, \boldsymbol{\sigma}_h^* & = \sum_i \int_{\Omega_i} \phi_h \ f(u)\ d\Omega, \\ \sum_i \oint_{\partial\Omega_i} \mu_h \boldsymbol{\sigma}_h^* \cdot \mathbf{n}_i & = 0, \end{cases} \tag{18}$$

for all $(\phi_h, \tau_h^*, \mu_h) \in L_h(\Omega) \times W_h(\Omega) \times M_h(\Omega)$. Now the third equation guarantees that the fluxes in the solution satisfy $\boldsymbol{\sigma}_h^* \in \mathbf{H}^{BC}(\mathrm{div}\,, \Omega)$. Hence, in the interior the solution of system (18) is the same as the solution of (12), and λ_h can be interpreted as the value of the potentials at the edges. The values λ_k are the coefficients in $\lambda_h = \sum_k \lambda_k \bar{\epsilon}_k$, with $\bar{\epsilon}_k$ the characteristic function on E_k, and the λ_k can be expressed as

$$\lambda_k = u_{i_1} \frac{\int a^{-1}\ d\Omega_{k,i_1}}{\int a^{-1}\ d\Omega_k} + u_{i_2} \frac{\int a^{-1}\ d\Omega_{k,i_2}}{\int a^{-1}\ d\Omega_k}, \tag{19}$$

where i_1 and i_2 denote adjacent cells. This actually comes down to linear interpolation for the Poisson equation, or exponential interpolation for the continuity equations as was used for the one-dimensional case in [5].

4 Vanka type relaxation

For the efficiency of the multigrid method the choice of a proper relaxation procedure is of prime importance. Several procedures are available to solve the system of equations that arises from the mixed finite element method. Blockwise relaxation with current conservation has been used by Schmidt and Jacobs [17] for the solution of a Poisson problem with Neumann boundary conditions, Maitre c.s. [10] give an analysis of Uzawa relaxation. Vanka [19] describes a block-implicit method applied to the incompressible Navier-Stokes equations. In that study the equations associated with the pressure in a cell and the velocities over the cell faces are solved in a coupled manner.

In the present study we use a method similar to the procedure used by Vanka. In our relaxation all cells on a given level are scanned in a predetermined order, either lexicographically or in a red-black ordering. When a cell is visited the variables related with that cell and the fluxes over its four edges are relaxed simultaneously. In this way 5 variables are relaxed for each equation in (6), and in the relaxation of a single cell 15 equations are solved simultaneously.

This system of equations for $(\psi_i, \phi_{n_i}, \phi_{p_i})$ and $(J_{\psi k}, J_{nk}, J_{pk})$, $k = N, E, S, W$, has the

following form

$$
\begin{array}{ll}
(a1) & \sum_k d_{i,k} h_k J_{\psi k} = (e^{\phi_{p_i} - \psi_i} - e^{\psi_i - \phi_{n_i}}) + D(x_i), \\
(a2) & J_{\psi k} = -\frac{d_{i,k} h_k}{a_k} \mu_\psi (\psi_k - \psi_i), \\
(b1) & \sum_k d_{i,k} h_k J_{nk} = +R(\psi_i, \phi_{n_i}, \phi_{p_i}), \\
(b2) & J_{nk} = -\frac{d_{i,k} h_k}{a_k} \mu_n (\phi_{n_k} - \phi_{n_i}) \frac{\text{Bexp}\,(-\psi_k, -\psi_i)}{\text{Bexp}\,(-\phi_{n_k}, -\phi_{n_i})}, \\
(c1) & \sum_k d_{i,k} h_k J_{pk} = -R(\psi_i, \phi_{n_i}, \phi_{p_i}), \\
(c2) & J_{pk} = -\frac{d_{i,k} h_k}{a_k} \mu_p (\phi_{p_k} - \phi_{p_i}) \frac{\text{Bexp}\,(\psi_k, \psi_i)}{\text{Bexp}\,(\phi_{p_k}, \phi_{p_i})}.
\end{array}
\tag{20}
$$

Due to the structure of the equations, the computational work in each cell is limited. We can exploit the linear appearance of J in the equations, and, as was the case in [19], the linearised form of the equations can be arranged in a block structure

$$
\begin{bmatrix}
b_N & 0 & 0 & 0 & +h_N \\
0 & b_E & 0 & 0 & +h_E \\
0 & 0 & b_S & 0 & -h_S \\
0 & 0 & 0 & b_W & -h_W \\
h_N & h_E & -h_S & -h_W & -h_N h_W f\prime(x_i)
\end{bmatrix}
\begin{bmatrix} \sigma_N \\ \sigma_E \\ \sigma_S \\ \sigma_W \\ u_i \end{bmatrix}
=
\begin{bmatrix} S_N + u_N h_N \\ S_E + u_E h_E \\ S_S - u_S h_S \\ S_W - u_W h_W \\ S_i + h_N h_W f(x_i) \end{bmatrix},
\tag{21}
$$

where $b_k = \text{area}(\Omega_k)/\mu$, $k = N, E, S, W$, for the Poisson equation,
or $b_k = \text{area}(\Omega_k)\mu^{-1}\text{Bexp}(\mp\psi_i, \mp\psi_k)$ for the continuity equations. S_k denotes a possible source term and u_k the potential in the neighbouring cells. The upper 4×4 block in this system is inverted analytically, which comes down to the local elimination of the fluxes.

Because the equations associated with the edges of a cell are satisfied as soon as that cell has been relaxed, it is a property of our 5-point Vanka relaxation that all equations related to the fluxes (i.e. eq. (17,a2,b2,c2)) are satisfied as soon as a complete relaxation sweep has been performed. (Notice that an over or under relaxation would spoil this property.) The residuals left are associated with the cells and describe the extent to which the conservation property is not satisfied.

Newton vs Gummel

What remains in the relaxation of a cell is the solution of the nonlinear part of the equations. For this we resort to two approaches (1) Newton's iteration, and (2) Gummel's iteration. (Notice that we apply these methods locally, in contrast with the usual approaches where these methods are used for all points in the grid simultaneously.) The advantage of Newton's method is its quadratic convergence in the neighbourhood of the solution. This well known phenomenon makes Newton's method efficient when good initial approximations are available. For practical problems it appeared that relaxation based on Newton's iteration took about 60% of the computing time needed by Gummel's iteration. (This figure depends on a number of factors, but it gives some qualitative impression.) For our equations, the problem with Newton's method is the strong nonlinearity in the potential variables and a possible lack of good initial approximations. Much of the nonlinearity is characterised by the fact that the variables ψ, ϕ_n and ϕ_p, appear as exponents in exponential functions.

Correction transformations and initial estimates

Because the equations are better linearised with respect to the variables n and p than to ϕ_n and ϕ_p (see Section 1) Schilders' correction transformation [15] is used, both in the case of Newton's and Gummel's method. This means that first the linearisation is made with respect to the potentials and the corresponding correction is computed. Then this correction is transformed to the correction that would have been obtained if a linearisation with respect to n or p were made. From equation (4) it follows that the corrections, expressed in the quasi-Fermi potentials, are related by

$$\Delta\phi_n^{\text{new}} = \Delta\psi^{\text{old}} - \log(1 - (\Delta\phi_n^{\text{old}} - \Delta\psi^{\text{old}})), \tag{22}$$

$$\Delta\phi_p^{\text{new}} = \Delta\psi^{\text{old}} + \log(1 + (\Delta\phi_p^{\text{old}} - \Delta\psi^{\text{old}})). \tag{23}$$

Such a transformation, introduced in [15] for the continuity equations, can also be used for the nonlinear Poisson equation. There we have to determine what part of the equation is dominating, the linear part or the nonlinear (exponential) part. We took the following strategy. Without loss of generality the Poisson equation (20 a) can be written as

$$a \sinh\psi + b\psi = 1. \tag{24}$$

If $|b| > |a\cosh\psi|$ we decide that the linear part is dominating and we apply the correction

$$\psi^{\text{new}} = \psi^{\text{old}} + \Delta\psi,$$

if $|b| < |a\cosh\psi|$ the nonlinear part is dominating and we take

$$\psi^{\text{new}} = \operatorname{arsinh}(\sinh\psi^{\text{old}} + \Delta\psi\cosh\psi^{\text{old}}).$$

If we need an initial estimate for Gummel's method, we can also start from equation (24). Depending on the size of ψ, we can find two approximations for the solution: $\psi = 1/(a+b)$, or $\psi = \operatorname{arsinh}(1/a)$. In order to decide which one is the more appropriate, we select the one for which the functional

$$G(\psi) = a\cosh\psi + \frac{1}{2}b\psi^2 - \psi$$

is minimal.

The convergence of pointwise Gummel iteration

Newton's method is used in the later stages of the local solution process, when good initial estimates are already available; in the absence of good initial estimates we use (locally) Gummel's iteration because it is more robust.

Little is known about the convergence of the pointwise Gummel iteration. Therefore we present here an analysis of the convergence of Gummel's decoupling method for the solution of the system (20). The objective is to obtain a more precise understanding of the convergence properties of this iterative scheme. The analysis presented predicts that the convergence of Gummel's method depends only on the difference in the values of ψ in the neighbouring control volumes, and not on the initial estimate or on the properties of the doping profile $D(x)$.

In the spirit of [8] we study the Gummel iteration as a fixed-point mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$, that maps a pair (ϕ_n, ϕ_p) onto a pair $(\tilde{\phi}_n, \tilde{\phi}_p) = T(\phi_n, \phi_p)$. To compute $T(\phi_n, \phi_p)$, first the electric potential $\psi(\phi_n, \phi_p)$ is computed as an intermediate result by the solution of (20 a). The values $\tilde{\phi}_n$ and $\tilde{\phi}_p$ are obtained from this $\psi(\phi_n, \phi_p)$ by the solution of (20 b,c). Existence of a solution in $A \subset \mathbb{R}^2$ follows when T is a contraction mapping on A. Then Gummel's iteration converges and the contraction factor may give an indication of the convergence speed of the iteration. To measure the distance in $\mathbb{R}^2$ we use the max-norm:

$$\|(\phi_n^1, \phi_p^1) - (\phi_n^2, \phi_p^2)\| = \max(|\phi_n^1 - \phi_n^2|, |\phi_p^1 - \phi_p^2|). \tag{25}$$

In order to be able to be more specific, we restrict the analysis to the zero recombination case. This enables us to find explicit expressions for the iterates.

Theorem 1 *If the variation in the ψ-values in the four neighbouring points is sufficiently small* ($\max_k \psi_k - \min_k \psi_k < 12$), *then the operator T for the pointwise Gummel iteration is a contraction, i.e.*

$$\|T(\phi_n^1, \phi_p^1) - T(\phi_n^2, \phi_p^2)\| \leq C \|(\phi_n^1, \phi_p^1) - (\phi_n^2, \phi_p^2)\|, \tag{26}$$

with $C = \frac{1}{12}(\max_k \psi_k - \min_k \psi_k)$ *and for all* $(\phi_n^i, \phi_p^i) \in \mathbb{R}^2$, $i = 1, 2$.

Proof: The proof is given in two parts. We consider the iteration sequence

$$(\phi_n^i, \phi_p^i) \to \psi^i \to (\tilde{\phi}_n^i, \tilde{\phi}_p^i), \quad i = 1, 2, \tag{27}$$

so that $\psi^i = \psi(\phi_n^i, \phi_p^i)$ and $(\tilde{\phi}_n^i, \tilde{\phi}_p^i) = T(\phi_n^i, \phi_p^i)$. In the first part we prove

$$|\psi^1 - \psi^2| \leq \|(\phi_n^1, \phi_p^1) - (\phi_n^2, \phi_p^2)\|, \tag{28}$$

and in the second part we show

$$\|(\phi_n^1, \phi_p^1) - (\phi_n^2, \phi_p^2)\| \leq C |\psi^1 - \psi^2|. \tag{29}$$

In fact we show (29) only for ϕ_p,

$$|\phi_p^1 - \phi_p^2| \leq C |\psi^1 - \psi^2|, \tag{30}$$

because a similar result for ϕ_n follows by analogy, and both results together yield (29).

In order to prove equation (28) we consider (20a), which yields for $i = 1, 2$,

$$\sum_k w_k \mu_\psi (\psi_k - \psi^i) + (e^{\phi_p^i - \psi^i} - e^{\psi^i - \phi_n^i}) + D(x) = 0,$$

with $w_k = h_k^2 / \text{area}(\Omega_k)$. By subtraction we obtain

$$\sum_k w_k \mu_\psi (\psi_2 - \psi_1) + (e^{\phi_p^1 - \psi^1} - e^{\psi^1 - \phi_n^1} - e^{\phi_p^2 - \psi^2} + e^{\psi^2 - \phi_n^2}) = 0$$

or

$$(\psi_1 - \psi_2) \mu_\psi \sum_k w_k = (e^{\psi^1 - \phi_n^1} (e^{(\phi_n^1 - \phi_n^2) - (\psi^1 - \psi^2)} - 1) + e^{\phi_p^2 - \psi^2} (e^{(\phi_p^1 - \phi_p^2) - (\psi^1 - \psi^2)} - 1)). \tag{31}$$

From this equality, the inequality (28) follows for the following reason.

Assume that (28) is *not* true, then we consider two cases: either $\psi^1-\psi^2 > 0$ or $\psi^1-\psi^2 < 0$. In the former case from the negation of (28) follows that $\psi^1 - \psi^2 \geq \phi_n^1 - \phi_n^2$ and $\psi^1 - \psi^2 \geq \phi_p^1 - \phi_p^2$. It follows that the left-hand side of the equality (31) is positive and the right-hand side is negative. This is a contradiction. Similarly, if $\psi^1 - \psi^2 < 0$ it follows that $\psi^1 - \psi^2 \leq \phi_n^1 - \phi_n^2$ and $\psi^1 - \psi^2 \leq \phi_p^1 - \phi_p^2$. Now it follows that the left-hand side of the equality (31) is negative and the right-hand side is positive. This also yields a contradiction. Because (28) is trivially satisfied for $\psi^1 = \psi^2$, we may conclude that (28) holds.

In order to prove the second part (30), we consider (20c). With zero recombination this yields for $i = 1, 2$, (dropping the subscript p)

$$\sum_k w_k(\phi_k - \phi^i)\frac{\text{Bexp}\,(\psi_k, \psi^i)}{\text{Bexp}\,(\phi_k, \phi^i)} = 0,$$

using the definition of Bexp for the denominators, we obtain

$$e^{\phi^i}\sum_k w_k \text{Bexp}\,(\psi_k, \psi^i) = \sum_k w_k\, e^{\phi_k}\, \text{Bexp}\,(\psi_k, \psi^i). \tag{32}$$

First we notice that all factors and terms in this expression are positive, and hence $\min_k e^{\phi_k} \leq e^{\phi^i} \leq \max_k e^{\phi_k}$, for $i = 1, 2$, which yields (without any restriction on ψ_k)

$$\min_k \phi_k \leq \phi^i \leq \max_k \phi_k, \quad \text{for } i = 1, 2,$$

and

$$\phi^1 - \phi^2 \leq |\max_k \phi_k - \min_k \phi_k|.$$

Further, from (32) we derive

$$e^{\phi^1-\phi^2} = \frac{\sum_k w_k \text{Bexp}\,(\psi_k, \psi^2)}{\sum_k w_k \text{Bexp}\,(\psi_k, \psi^1)} \cdot \frac{\sum_k w_k \text{Bexp}\,(\psi_k, \psi^1)e^{\phi_k}}{\sum_k w_k \text{Bexp}\,(\psi_k, \psi^2)e^{\phi_k}}.$$

Now we define ψ_A to be the value of ψ_k for which

$$\frac{\text{Bexp}\,(\psi_A, \psi^2)}{\text{Bexp}\,(\psi_A, \psi^1)} \geq \frac{\text{Bexp}\,(\psi_k, \psi^2)}{\text{Bexp}\,(\psi_k, \psi^1)} \tag{33}$$

for all k, and similarly ψ_B such that

$$\frac{\text{Bexp}\,(\psi_B, \psi^1)}{\text{Bexp}\,(\psi_B, \psi^2)} \geq \frac{\text{Bexp}\,(\psi_k, \psi^1)}{\text{Bexp}\,(\psi_k, \psi^2)}$$

for all k, then

$$e^{\phi^1-\phi^2} \leq \frac{\text{Bexp}\,(\psi_A, \psi^2)}{\text{Bexp}\,(\psi_A, \psi^1)} \cdot \frac{\text{Bexp}\,(\psi_B, \psi^1)}{\text{Bexp}\,(\psi_B, \psi^2)}. \tag{34}$$

Taking the logarithm and introducing the function $g(x) = \log\left(\frac{x}{e^x-1}\right)$, we may write (34) as

$$\phi^1 - \phi^2 \leq g(\psi^2 - \psi_A) - g(\psi^1 - \psi_A) - g(\psi^2 - \psi_B) + g(\psi^1 - \psi_B)$$

or

$$\phi^1 - \phi^2 \leq \int_{-\psi_B}^{-\psi_A} \int_{\psi_2}^{\psi_1} (-g''(x+y))\, dx\, dy\ .$$

Since

$$g''(x) = \frac{1}{2\cosh(x) - 2} - \frac{1}{x^2}$$

we know that $0 < -g''(x) \leq 1/12$ and

$$\phi^1 - \phi^2 \leq \frac{1}{12}(\psi_B - \psi_A)(\psi^1 - \psi^2).$$

To determine ψ_A and ψ_B we consider

$$\log\left(\frac{\text{Bexp}\ (\psi, \psi_2)}{\text{Bexp}\ (\psi, \psi_1)}\right) = \int_{\psi_1}^{\psi_2} g'(x - \psi)dx = (\psi_2 - \psi_1)g'(\psi_m - \psi)$$

for some $\psi_m \in (\psi^1, \psi^2)$. Because $g'(\psi_m - \psi)$ is a monotonically increasing function of ψ we find $\psi_A = \max_k \psi_k$ and $\psi_B = \min_k \psi_k$ if $\psi_2 > \psi_1$, and if $\psi_2 < \psi_1$ we have $\psi_A = \min_k \psi_k$ and $\psi_B = \max_k \psi_k$. It follows that

$$\phi^1 - \phi^2 \leq \frac{1}{12}(\max_k \psi_k - \min_k \psi_k)|\psi^1 - \psi^2|.$$

Because the superscripts 1 and 2 may be interchanged without changing the meaning of the right-hand side, this proves (30) and hence the theorem. □

The proof of the theorem, valid for zero recombination and zero source term, clearly shows that convergence may be slower if a source term for the continuity equations takes values that make the right-hand side of (32) smaller. No solution exists for the local nonlinear problem, if the source term makes the right-hand side of (32) negative. This means that large source terms can cause the non-existence of a solution. Hence, we have to face the possibility that the correction equations in the multigrid process have no solution if the right-hand side of the equation gets too large.

5 The coarse grid correction

If, for the solution of the nonlinear discrete equation,

$$N_h(q_h) = f_h, \tag{35}$$

we consider the usual nonlinear coarse grid correction stage of a two-grid process,

$$N_{2h}(\tilde{q}_{2h}) = N_{2h}(q_{2h}) + \mu\, \overline{R}_{2h,h}(f_h - N_h(q_h^{\text{old}}))\,, \tag{36}$$

$$q_h^{\text{new}} = q_h^{\text{old}} + P_{h,2h}(\tilde{q}_{2h} - q_{2h})/\mu, \tag{37}$$

we recognise five important components that influence the effect of this stage. In the first place, there are the three operators N_{2h}, $\overline{R}_{2h,h}$ and $P_{h,2h}$, and further the starting approximation on the coarser grid, q_{2h}, and the parameter $\mu \in \mathbb{R}$. For a nonlinear problem, the operator N_{2h} is often constructed by the same method as is used for N_h; in our case it is

described in Section 3. In principle, the choice for the operators $\overline{R}_{2h,h}$ and $P_{h,2h}$ is free (as long as they are accurate enough), but in the context of our MFE discretisation there exist a natural prolongation and restriction associated with the discretisation, viz. those induced by the relations $L^2_{2h}(\Omega) \subset L^2_h(\Omega)$, and $\mathbf{H}^{BC}_{2h}(\mathrm{div}\,,\Omega) \subset \mathbf{H}^{BC}_h(\mathrm{div}\,,\Omega)$. These relations imply that the prolongation corresponds for the potentials with piecewise constant interpolation, and for both components of the fluxes with piecewise linear interpolation in one direction and piecewise constant interpolation in the other. The corresponding prolongation stencils are $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, for the potentials (associated with a cell), and $\begin{bmatrix} 1/2 & 1/2 \\ 1 & 1 \\ 1/2 & 1/2 \end{bmatrix}, \begin{bmatrix} 1/2 & 1 & 1/2 \\ 1/2 & 1 & 1/2 \end{bmatrix}$, for the fluxes (associated with a horizontal and a vertical edge respectively). The natural restriction $\overline{R}_{2h,h}$ is the transpose of the natural prolongation $P_{h,2h}$, because the spaces of test and trial functions in (12) are the same.

For strong nonlinear problems, also the choice of the starting approximation q_{2h} and the parameter μ are of importance because they determine to a very large extent the coarse grid problem that is solved. If the distance between q_h^{old} and the solution of (35) is small, it is clear that mainly q_{2h} determines the coarse grid problem, and it is wise to select q_{2h} in such a way that the problem (36) is well conditioned. If (36) is not ill-conditioned, the parameter μ can be used to keep $\tilde{q}_{2h}$ in a sufficiently small neighbourhood of q_{2h}. This may guarantee the existence of a solution of the correction equation. However, the effect of a small μ can be that only a very small neighbourhood of q_{2h} is considered, so that nonlinear effects in N_{2h} are neglected. Moreover, the factor $1/\mu$ in (37) can amplify the errors made in the solution of (36).

For the semiconductor equations (6) without a row scaling, the residual for the continuity equations correspond with the rate-of-change in the carrier concentrations, cf. eq. (1). In this unscaled form, the natural restriction operator has a "physical meaning": the sum of the rate-of-change in four small sub-cells corresponds with the total rate-of-change in the father cell. We believe that this is an advantageous property of the equations in their unscaled form. However, without row-scaling, the size of the residuals (as well as the size of the diagonal elements of the Jacobian matrix) may vary largely in magnitude. This introduces the difficulty that for some parts of the domain Ω the large residual requires a very small μ, whereas a larger μ would be allowed in other parts. An even more awkward situation is encountered if the values for a proper row-scaling differ strongly for the equations related with a coarse grid cell and the corresponding equations on the finer level. In this case a large residual on the fine grid may yield an improper large correction on the coarse grid. This effect is seen in regions where the character of the solution changes rapidly (transition between N- and P-region, depletion layer). The same effect was observed by de Zeeuw in [4] in the 1D-case and it leads to the introduction of a residual damping operator D_{2h}. This D_{2h} is a diagonal operator, depending on the current coarse and fine grid solution, which has entries in $[0,1]$. Hence, for the coarse grid correction we use

$$N_{2h}(\tilde{q}_{2h}) = N_{2h}(q_{2h}) + D_{2h}(q_{2h}, q_h^{\mathrm{old}})\,\overline{R}_{2h,h}(f_h - N_h(q_h^{\mathrm{old}}))\,, \tag{38}$$

$$q_h^{\mathrm{new}} = q_h^{\mathrm{old}} + P_{h,2h}(\tilde{q}_{2h} - q_{2h})\,. \tag{39}$$

This means that the coarse grid correction (38), (39) is not able to reduce *all* components of the residual that can be represented on the coarse grid, but an amount $(I_{2h} - D_{2h})\overline{R}_{2h,h}(f_h -$

$N_h(q_h^{\text{old}})$) remains unaffected (in a single CGC sweep). The elements of D_{2h} are different from 1 only in small regions (the transition regions in the semiconductor), and the effect of the damping is compensated in these regions by additional relaxation on the fine grid. The precise construction of the operator D_{2h} is found in [14].

The selection of a proper coarse grid approximation

For the selection of q_{2h} in (36) or (38) two approaches are in common use. Either $q_{2h} = R_{2h,h}q_h^{\text{old}}$ is used, where $R_{2h,h}$ is a restriction operator for the solution, or for q_{2h} one takes simply the last approximation that is available in the full multigrid process, i.e. one starts with the approximate solution on the coarse grid as obtained in the nested iteration, and later -at each stage of the multigrid process- the last approximate solution on a given level is used as an initial approximation in the next stage.

In practice, it appears that the latter technique performs rather well. However, we consider it unreliable because in all later stages of the process the approximate solutions on the coarser levels depend on the complete history of the computational process, and there is no mechanism that forces such a coarse grid approximation to stay in the neighbourhood of a solution. In fact, such an approximation q_{2h} may loose properties that are required for a proper approximate solution, e.g. symmetry.

The first approach, however, requires the selection of an $R_{2h,h}$ and for our problem there is no reason to assume that e.g. the simple use of L^2-projection of the Slotboom variables -as suggested by the discretisation- will yield a proper problem (38). It seems a better choice to take mean values for ψ and to construct ϕ_n and ϕ_p such that the total amount of electrons and holes in a coarse cell equals the sum of the amounts in the corresponding smaller cells.

A third, more simple technique was adopted because of its good results: compute a reasonably accurate discrete approximation on the coarse grid during the nested iteration, and keep this value as q_{2h} during all the later stages of the computation.

In our case this last technique can be understood as a favourable approach for the following reason. For the homogeneous continuity equations (20b,c) with $R = 0$, the Scharfetter-Gummel discretisation has the property that the row-sum of the discrete matrix $(\partial J_p/\partial \phi_p)$ at cell i is equal to the residual of the discrete equation for that cell: $\sum_l \frac{\partial}{\partial \phi_{p_l}}(\sum_k h_k d_{ik} J_{pk}) = \sum_k h_k d_{ik} J_{pk}$, and, analogously, for the other continuity equation $\sum_l \frac{\partial}{\partial \phi_{n_l}}(\sum_k h_k d_{ik} J_{nk}) = -\sum_k h_k d_{ik} J_{nk}$. This follows from, (cf. equation (20 b2,c2)), $J_{pk} = \frac{\partial J_{pk}}{\partial \phi_{p_k}} + \frac{\partial J_{pk}}{\partial \phi_{p_i}}$, $J_{nk} = \frac{\partial J_{nk}}{\partial \phi_{n_k}} + \frac{\partial J_{nk}}{\partial \phi_{n_i}}$. Hence the row-sums in the Jacobian matrix vanish when q_{2h} is in the neighbourhood of the discrete solution. This implies that in the neighbourhood of the discrete solution the linearised operators in the Gummel process are M-matrices. A positive recombination only improves the situation. This shows that the stability of the linearised operators is better in the neighbourhood of a solution than at some distance from the solution.

Other transfer operators

A priori there is also no reason to assume that the natural grid transfer operators $P_{h,2h}$ and $\overline{R}_{2h,h}$ are the best, or even that they are sufficiently accurate (smooth) in order not to disturb reduction of the the high frequency components in the solution.

Indeed, for the one-dimensional case, in combination with the Vanka relaxation we observe by Fourier analysis that these simple transfer operators are too inaccurate. The non-damped 5-point Vanka relaxation can be considered as eliminating the fluxes and applying a collective Gauss-Seidel procedure to the remaining potentials. After elimination of the fluxes, the differential equations for the potentials are second order, and hence the rule applies that the sum of the HF orders of the prolongation and restriction should at least be two [7]. The orders of the natural prolongation and restriction, however, are one. To obtain a proper convergence of the MG algorithm, one should take more accurate transfer operators. This is analyzed in detail by two-grid Fourier analysis in [13]. The simplest operator that satisfies a sufficient accuracy condition is piecewise linear interpolation and its adjoint as a weighted restriction. The LF and HF order of this restriction is 2.

However, the same effect is not seen in 2D [13]. The Vanka relaxation damps sufficiently the HF modes that are allowed by the transfer operators, and in practical 2D computations piecewise constant interpolation, together with its transpose for a restriction, gave satisfactory results. We did not observe an improvement when more accurate restrictions were used instead of the natural restriction.

Because of the asymmetric character of the convection operator, and in view of the successful use of an asymmetric prolongation in a multigrid method for the one-dimensional semiconductor problem in [4] [5] [6], it is interesting to consider the possibility of an asymmetric prolongation for the two-dimensional problem as well. In 1D such an interpolation was based on the form (cf. eq. (6 b,c))

$$\Phi(x) - \Phi(a) = \int_a^x e^{\pm\psi(\xi)} J d\xi, \tag{40}$$

with the assumption of a piecewise constant J and a piecewise linear ψ over the area of integration (the dual boxes). In our MFE context, the same exponential interpolation formula is found in Section 3 as equation (19). The principle behind the construction of that prolongation in the one-dimensional case is the equal flux over corresponding coarse and fine grid edges. In two dimensions, however, such an explicit prolongation cannot be constructed. This is because in two dimensions the assumption of a piecewise constant J and the existence of a unique function Φ leads to an inconsistency. Independence of $\Phi(x)$ on the integration path means grad $\Phi = \exp(\pm\psi)J$. This relation only holds for ψ and J satisfying

$$0 = \text{rot grad } \Phi = \text{rot } (e^{\pm\psi} J) = e^{\pm\psi}(\text{rot } J \pm J \times \text{grad } \psi). \tag{41}$$

With the assumption of a constant J, this implies that J should be parallel with grad ψ. However, for a two-dimensional case, this is generally too restrictive a condition. From equation (41) follows that J has the general form $J = \text{grad } u \mp u\text{grad } \psi$, for an arbitrary scalar function u.

Assuming that the dependence of the integration path has only a minor influence, we might overlook the non-uniqueness of Φ and select an path, e.g. select the shortest line segment from the coarse cell center (with the known potential) to the fine cell center (where the potential has to be computed). Then the fluxes over corresponding edges in the coarse and the fine mesh are not equal, and in our experiments this interpolation appears no better than the piecewise constant interpolation.

6 Example

As a test problem we used a bipolar NPN transistor from the CURRY example set [9]. The geometry of the transistor is shown in Figure 1. There is an N-type emitter region, a P-type base region and an N-type collector region. The length of the device is 20 micron and the width is 8. For the precise description of the doping profile we refer to [9]. The Shockley-Read-Hall model 3 is used for the recombination, with carrier lifetimes $\tau_p = \tau_n = 10^{-6}$. Dirichlet boundary conditions are given at the contacts. On the remaining boundaries homogeneous Neumann boundary conditions apply.

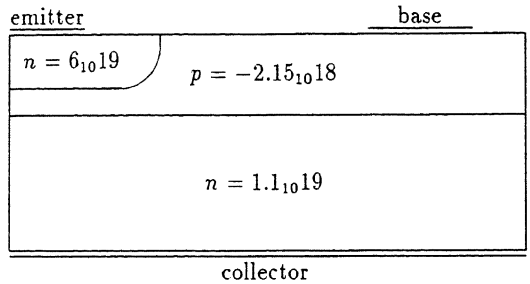


Figure 1: Geometry of the bipolar transistor

The voltages applied to the collector and base are kept constant at $V_{coll} = 1.0$V and $V_{base} = 0.0$V. The simulation is started with zero potential at the emitter. Then no currents are present because all P-N junctions are blocked. The simulation is continued, first by an increase of the emitter voltage to -0.5V and then in steps (-0.05V) to -0.8 Volts. Then currents are clearly present. In Table 1 we show the collector currents computed on a 16 × 40 , 32 × 80 and 64 × 160 mesh, together with a reference solution computed with the CURRY package on a non-uniform 56 × 62 grid. We see that the solution (i.e. the collector current) appears to converge for vanishing mesh-widths.

	MFEM uniform mesh			CURRY non-uniform
V_{emit}	16 × 40	32 × 80	64 × 160	56 × 62
0.0	5.3(-12)	5.1(-11)	5.2(-11)	7.2(-11)
-0.50	9.5(-5)	1.4(-5)	1.0(-5)	9.8(-6)
-0.55	5.8(-4)	9.5(-5)	7.0(-5)	6.7(-5)
-0.60	3.4(-3)	6.4(-4)	4.8(-4)	4.6(-4)
-0.65	1.8(-2)	4.3(-3)	3.3(-3)	3.1(-3)
-0.70	8.4(-2)	2.8(-2)	2.2(-2)	2.1(-2)
-0.75	3.2(-1)	1.7(-1)	1.4(-1)	1.3(-1)
-0.80	1.1(0)	7.9(-1)	7.1(-1)	6.9(-1)

Table 1: Collector currents (A/cm).

For the coarsest mesh, the device is divided into 4 × 10 (!) squares. We notice that this mesh is so coarse that the emitter boundary does not fit the edges of the cells. Therefore, for

the discretisation, an obvious generalisation of the method described in Section 3 was used. In the discretisation, for each cell the current through such a boundary edge is determined by the boundary potential, the potential in that cell and the proportion of the edge that is covered by the contact. This treatment of the boundary prevents the obligation to use fine or irregular cells in the coarsest grids.

The initial estimates for the emitter voltages -0.5(-0.05)-0.8 were obtained from the solutions computed with the previous voltage. First the solution on the coarsest grid was accurately computed, and the solution on the finer grids was computed (approximately, by a few W-cycles) before an interpolation to the next finer grid was made. In the interpolation to the finer grid, the low frequencies in the solution were taken from the coarser grid, whereas the high frequency components were taken from the fine grid solution for the lower voltage. Thus mimicking a well known technique used for time dependent problems.

The MGM used, Convergence results

The multigrid method to solve the transistor problem applies the lumped MFE discretisation as described in Section 3. The natural prolongation and restriction operators were used, together with the residual damping as explained in Section 5. A single additional point-Vanka relaxation sweep was made over all fine grid cells for which the residual was damped in the coarse grid correction. As the initial estimate q_{2h} we kept the solution obtained initially on the coarse grid. Both in the pre- and in the post-relaxation stage a single sweep of the smoothing procedure was used.

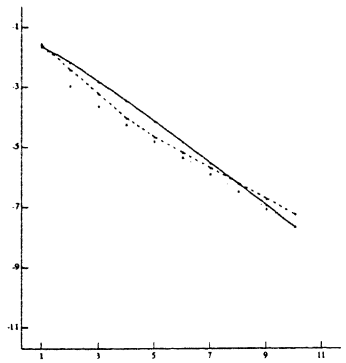
Beside the symmetric lexicographic point-Vanka relaxation, also a (non-symmetric, but horizontal+vertical) line-Vanka relaxation was applied as smoothing procedure. In the results shown, only W-cycle results are given. As was shown earlier [6] [14] V-cycles are less robust for the semiconductor problem.

In Figure 2 convergence histories are shown for the multigrid solution process. On the horizontal axis the number of cycles is given, and on the vertical axis the scaled residual. The residual scaling was made pointwise, by means of the diagonal 3×3 blocks of the Jacobian matrix. Thus the residual corresponds with corrections that would occur if a pointwise collective Jacobi relaxation was used. Hence, the scaled residual can be associated with corrections for (ψ, ϕ_n, ϕ_p). For the resulting scaled residual the maximum was taken over the grid and over the three variables (ψ, ϕ_n, ϕ_p).

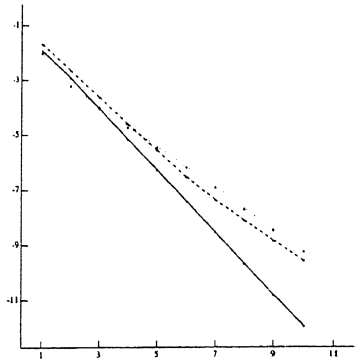
Convergence results are shown for the solution with 3, 4 and 5 grid levels, both for point-Vanka and for line-Vanka relaxation. It is observed that line-Vanka relaxation is the more efficient. It is seen that the convergence is not always stabilised to a constant factor after 10 iterations, but an almost grid independent convergence rate can be expected. In any case, convergence is fast and a limited number of iterations is sufficient to attain truncation error accuracy.

Anknowledgement

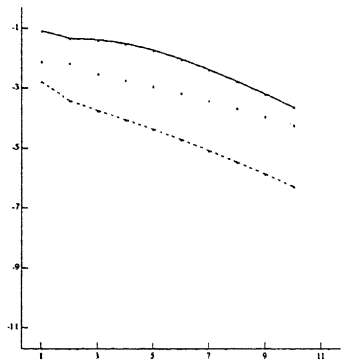
We would like to thank Dr. W.H.A. Schilders for reading the manuscript.



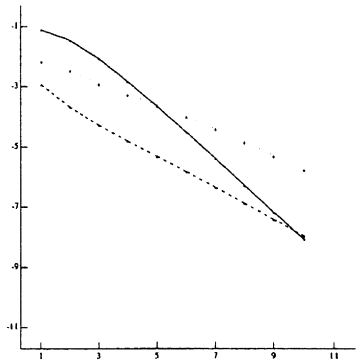
Case a: $V_e = 0.0$
Point-Vanka relaxation



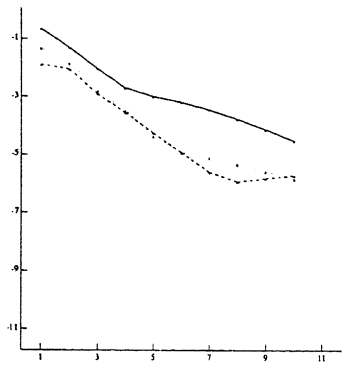
Case b: $V_e = 0.0$
Line-Vanka relaxation



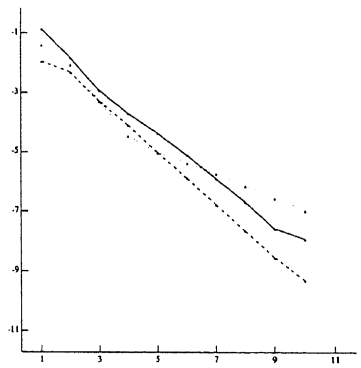
Case c: $V_e = -0.5$
Point-Vanka relaxation



Case d: $V_e = -0.5$
Line-Vanka relaxation



Case e: $V_e = -0.8$
Point-Vanka relaxation



Case f: $V_e = -0.8$
Line-Vanka relaxation

Figure 2: Convergence of the MG method (W-cycle).
The 10-log of the scaled residual against the iteration number.
Solid line: 16×40 mesh; dotted line: 32×80 mesh; dashed line: 64×160 mesh.

References

[1] R.E. Bank, W. Fichtner, D.J. Rose, and R.K. Smith. Algorithms for semiconductor simulation. In P. Deuflhard and B. Engquist, editors, *Numerical Analysis of Singular Perturbation Problems*, pages 3–21. Birkhauser, 1988.

[2] A. Brandt. Multi-level adaptive techniques for singular perturbation problems. In P.W. Hemker and J.J.H. Miller, editors, *Numerical Analysis of Singular Perturbation Problems*, pages 53–142, Londen, New York, 1979. Academic Press.

[3] F. Brezzi, L.D. Marini, and P. Pietra. Two-dimensional exponential fitting and applications to semiconductor device equations. Technical Report 597, Inst. Anal. Numerica CNR, 1987.

[4] P.M. de Zeeuw. Nonlinear multigrid applied to a 1D stationary semiconductor model. Technical Report NM-R8905, CWI, Dept.NM, Amsterdam, 1989.

[5] P.W. Hemker. A nonlinear multigrid method for one-dimensional semiconductor device simulation. In Guo Ben-yu, J.J.H. Miller, and Shi Zhong-ci, editors, *BAIL V, Proceedings of the 5th International Conference on Boundary And Interior Layers - Computational and Asymptotic Methods*, pages 18–29, Dublin, 1988. Boole Press.

[6] P.W. Hemker. A nonlinear multigrid method for one-dimensional semiconductor device simulation: Results for the diode. *J. Comput. Appl. Math*, 30:117–126, 1990.

[7] P.W. Hemker. On the order of prolongations and restrictions in multigrid procedures. *J. Comput. Appl. Math*, 1990. to appear.

[8] T. Kerkhoven. On the effectiveness of Gummel's method. *SIAM J.S.S.C.*, 9:48–60, 1988.

[9] C. Lepoeter. CURRY example set. Technical Report No. 4322.271.6005, Philips, Corp. CAD Centre, Eindhoven, 1987.

[10] J.-F. Maitre, F. Musy, and P. Nigon. A fast solver for the Stokes equations using multigrid with a Uzawa smoother. In D. Braess, W. Hackbusch, and U. Trottenberg, editors, *Advances in Multi-Grid Metods*, volume 11 of *Notes on Numerical Fluid Dynamics*, pages 77–83, Braunschweig, 1985. Vieweg.

[11] P.A. Markowich. *The Stationary Semiconductor Device Equations.* Springer Verlag, Wien, New York, 1986.

[12] S.F. McCormick. *Multilevel Adaptive Methods for Partial Differential Equations*, volume 6 of *Frontiers in Applied Mathematics.* SIAM, 1989.

[13] J. Molenaar. A two-grid analysis of the combination of mixed finite elements and Vanka relaxation. Technical Report to appear, CWI, Dept.NM, Amsterdam, 1990.

[14] J. Molenaar and P.W. Hemker. A multigrid approach for the solution of the 2D semiconductor equations. *IMPACT*, page (to appear), 1990.

[15] S.J. Polak, C. den Heijer, W.H.A. Schilders, and P. Markowich. Semiconductor device modelling from the numerical point of view. *Int. J. Num. Meth. Engng.*, 24:763–838, 1987.

[16] S.J. Polak, W.H.A. Schilders, and H.D. Couperus. A finite element method with current conservation. In G. Baccarani and M. Rudan, editors, *Simulation of Semiconductor Devices an Processes*, volume 3, Bologna, Italy, 1988. Technoprint.

[17] G.H. Schmidt and F.J. Jacobs. Adaptive local grid refinement and multi-grid in numerical reservoir simulation. *J. Comput. Phys.*, 77:140–165, 1988.

[18] W. van Roosbroeck. Theory of flow of electrons and holes in germanium and other semiconductors. *Bell Syst. Techn. J.*, 29:560–607, 1950.

[19] S.P. Vanka. Block-implicit multigrid solution of Navier-Stokes equations in primitive variables. *J. Comp. Phys.*, 65:138–158, 1986.