Numerical Methods and Approximation Theory III (Niš, August 18-21, 1987)

NUMERICAL METHODS IN SEMICONDUCTOR DEVICE SIMULATION

P.A. MARKOWICH, C. SCHMEISER* and S. SELBERHERR

Abstract: The simulation of the electrical behavior of semiconductor devices involves the solution of initial-boundary value problems for a nonlinear elliptic-parabolic system. Two major difficulties in the numerical solution of these problems are discussed:

a) The construction of discretisations is not obvious as the equations are singularly perturbed.
b) The discretised problems are very large systems of nonlinear algebraic equations which have to be solved iteratively.

1. INTRODUCTION

The electrical behavior of a semiconductor device is determined by the flow of two types of free charge carriers, the electrons in the conduction band (density $n(x,t)$) and the defect electrons or holes in the valence band (density $p(x,t)$). Well accepted models for the flow of electrons and holes are the Boltzmann transport equations, but their complexity is prohibitive for the numerical simulation of complicated devices. Perturbation arguments lead to the simplified drift-diffusion approximation of the current densities:

$$(1.1a)\qquad \begin{aligned} J_n &= \mu_n(\nabla n+nE) \ , \\ J_p &= -\mu_p(\nabla p-pE) \ . \end{aligned}$$

(All the appearing variables and parameters are already in scaled dimensionless form.)

*The work of the second author was supported by "Österreichischer Fonds zur Förderung der wissenschaftlichen Forschung".

In (1.1a) the parameters μ_n, μ_p denote mobilities and E is the electric field which is related to the electrostatic potential ψ by

(1.1b) $$E = -\nabla\psi \ .$$

Common models for the mobilities depend on n,p,E and the position x.

Maxwell's equations imply the continuity equations

(1.1c) $$\begin{aligned} \operatorname{div} J_n - n_t &= R \ , \\ \operatorname{div} J_p + p_t &= -R \end{aligned}$$

and Poisson's equation

(1.1d) $$\lambda^2 \Delta\psi = n - p - C(x) \ ,$$

where the source term R, called the recombination-generation rate, is the number of electron-hole pairs which are generated $(R<0)$ or disappear $(R>0)$ per unit time. It is usually modelled as a given function of n,p,E and position. The function $C(x)$, the so called doping profile, denotes the concentration of impurity ions. The dimensionless parameter λ is the scaled minimal Debye length and takes small values for realistic semiconductor devices.

The unscaled equations (1.1) are due to Van Roosbroeck [21]. For a derivation from Maxwell's equations and the Boltzmann transport equation see Selberherr [18]. The scaling which leads to (1.1) can be found in Markowich [8].
Mathematically a semiconductor device is given by the doping profile $C(x)$ defined in a bounded domain $\Omega \subseteq \mathbb{R}^3$ which represents the semiconductor part of the device. For the purpose of simulation it often makes sense to reduce the dimension of Ω. Thus, we take $\Omega \subset \mathbb{R}^k$, $k = 1,2$ or 3. The boundary $\partial\Omega$ splits into the union of contact segments $\partial\Omega_D$ where Dirichlet boundary conditions for n,p and ψ are given

(1.2a) $$n\big|_{\partial\Omega_D} = n_D \ , \quad p\big|_{\partial\Omega_D} = p_D \ , \quad \psi\big|_{\partial\Omega_D} = \psi_D \ ,$$

and the insulating part $\partial\Omega_N$ where the homogeneous Neumann conditions

(1.2b) $$(J_n,\nu)\big|_{\partial\Omega_N} = (J_p,\nu)\big|_{\partial\Omega_N} = (E,\nu)\big|_{\partial\Omega_N} = 0$$

hold. In (1.2b) ν denotes the outward normal vector of $\partial\Omega$.

Substituting (1.1a) into (1.1c) shows that (1.1) is a system of two parabolic equations for n and p coupled to an elliptic equation for ψ. In order to complete the formulation of an initial-boundary value problem initial conditions for the densities

$$(1.3)\qquad n(x,0) = n_I(x)\ ,\quad p(x,0) = p_I(x)\ ,\quad x \in \Omega$$

have to be prescribed. The potential at $t = 0$ can be determined by solving Poisson's equation. Several existence and uniqueness results for (1.1)-(1.3) can be found in the literature (see e.g. Mock [12]). Existence results for the corresponding stationary problem are contained in [8] and [12]. Uniqueness cannot be expected in general (see Steinrück [19]).

For the construction and analysis of numerical methods some a priori knowledge of the solution structure is extremely important. This can be gained from a singular perturbation analysis by exploiting the smallness of the parameter λ^2 in (1.1d). In the stationary case such an analysis shows that the solution can be approximated by setting $\lambda = 0$ except in thin layer regions where it varies rapidly (see [8]). For the time dependent problem additionally an initial layer appears (see Ringhofer [14], Szmolyan [20], Markowich [9]). In this paper we will be concerned with the stationary problem. Its analysis is facilitated by the transformation

$$(1.4)\qquad n = e^{\psi}u\ ,\quad p = e^{-\psi}v$$

which takes the stationary differential equations to the form

$$(1.5)\qquad \begin{aligned} &\lambda^2\Delta\psi = e^{\psi}u - e^{-\psi}v - C(x) \\ &\operatorname{div}(\mu_n e^{\psi}\nabla u) = R \\ &\operatorname{div}(\mu_p e^{-\psi}\nabla v) = R \end{aligned}$$

The continuity equations are in self-adjoint form now. Besides u and v are so called slow variables which means that they do not exhibit layer behavior. As opposed to (1.1d) the potential can be determined from the reduced $(\lambda=0)$ Poisson's equation. Subject to the appropriate boundary conditions each of the equations in (1.5) represents a well posed problem for the variable which appears with the highest differential order, when the other two variables are considered as known.

These properties make it much easier to design numerical methods which are well suited for (1.5) than for the original system. Unfortunately the potential becomes rather large in many applications such that u and v are so out of range that they are impossible to compute with (for different choices of variables and related conditioning questions see Bank et al. [3], Schmeiser et al. [17], Ascher et al. [1]). These facts led to the following approach: Methods are designed and analysed for (1.5). In computations the transformation (1.4) is applied on the discrete level to be able to compute with the original variables ψ, n and p.

2. DISCRETISATIONS

In this section we shall present discretisations for the steady state semiconductor equations which take into account the singular perturbation nature of the problem. The properties of system (1.5) allow for a decoupled approach, where each equation is treated separately.

2.1. Poisson's equation is a semilinear elliptic equation for the potential when u and v are considered to be known. The solution is approximated by a solution of the reduced equation except close to regions of rapid variation of the doping profile and possibly close to the boundaries where the solution varies rapidly. When trying to solve the problem numerically one would expect to be forced to use grids which are fine enough in the regions of rapid variation to resolve the solution structure. For the simulation of complex devices the cost of using such a grid is prohibitive. In order to get around this difficulty, discretisations are used which mimic the above described properties of the continuous problem by the use of lumping for the evaluation of the right hand side. A finite element of finite difference discretisation at node x_i then takes the form

$$(2.1) \qquad \lambda^2(\Delta_h\psi_h)_i = e^{\psi_i}u_i - e^{-\psi_i}v_i - C(x_i)$$

where Δ_h is a discretised version of the Laplace-operator (see Markowich [8], Selberherr [18]). The effect of lumping is that the reduced equations in the continuous and the discrete

case are the same. For any discretisation which inherits the stability properties of the continuous operator (maximum principle) the solution structure is similar for the discrete and continuous problems even if a coarse mesh is used. The main difference is that layers in the discrete case may be wider $(O(h))$ than in the continuous case $(O(\lambda))$. This fact will be demonstrated in the following section. It has two effects of major importance. First, even when starting on a very coarse grid adaptive grid refinement will be able to detect the correct solution structure. Second, as the solution is approximated well away from the thin layer regions the approximation error will be small if measured in integral norms although large pointwise errors may occur. The importance of this effect will also be demonstrated in section 3.

2.2. The continuity equations. We shall only deal with the electron continuity equation as the necessary modifications for the hole continuity equation are obvious. Let us first consider the one-dimensional situation. As the variables u and J_n are slow variables - in the language of singular perturbation theory - in this case, the discretisation of

$$J_n' = R \;, \quad J_n = \mu_n e^{\psi} u' \tag{2.2}$$

is not very critical. For simplicity we assume an equidistant grid and replace the first equation at the gridpoint x_i by

$$J_{n,i+1/2} - J_{n,i-1/2} = h\, R_i \tag{2.3a}$$

where R_i denotes an approximation of the recombination-generation rate at x_i. The second equation is approximated between gridpoints by

$$J_{n,i+1/2} = \mu_{n,i+1/2} (e^{\psi})_{i+1/2} \frac{u_{i+1}-u_i}{h} \;, \tag{2.3b}$$

where the approximation $\mu_{n,i+1/2}$ for μ_n at $\frac{x_i+x_{i+1}}{2}$ depends on the model which is used. For the approximation $(e^{\psi})_{i+1/2}$ two obvious choices are

$$\frac{1}{2}(e^{\psi_i}+e^{\psi_{i+1}}), \quad \exp(\frac{\psi_i+\psi_{i+1}}{2}) \;.$$

A third possibility is obtained by replacing μ_n and J_n by constants and ψ by a linear function in $[x_i, x_{i+1}]$ and solving the second equation in (2.2) explicitely. This results in the approximation

$$(2.3c) \qquad (e^{\psi})_{i+1/2} = \frac{\psi_{i+1} - \psi_i}{e^{-\psi_i} - e^{-\psi_{i+1}}} .$$

This procedure could have also been applied to the equation (1.1a) in the original variable n. The so obtained discretisation which is equivalent to (2.3) is an example of an exponentially fitted method (see Doolan et al. [5]) and bears the names of the engineers Scharfetter and Gummel [16] in the semiconductor device simulation literature.

The difference between the above mentioned discretisations is an unsettled issue from the theoretical point of view, but in practically all of the existing device simulation software the Scharfetter-Gummel scheme is used.

Extensions to finite difference methods in the two- and three-dimensional cases are straightforward (see Selberherrr [18]). Finite element methods which are generalisations of the Scharfetter-Gummel method to the two-dimensional situation can be found in Buturla and Cottrell [4] and Markowich and Zlamal [10]. It can be shown that the errors only depend on the variation of the current density J_n (see [10], Mock [13]). The drawback in the multidimensional situation is that J_n is not a slow variable in general (see Markowich [8]) which makes it necessary to use fine grids in regions of rapid variation of J_n. However, in most practical situations J_n varies much less than ψ,n and p and the computational effort remains reasonable.

The above error considerations dealt with each equation separately. In order to prove convergence results for the full system one has to assume wellposedness of the problem. Then the error estimates for the single equations can be combined (see [8]).

3.A UNIFORM CONVERGENCE RESULT

When talking about numerical methods for singular perturbation problems uniform convergence means roughly that errors can be estimated independently of the singular perturbation parameter. In particular, errors are even small if the grid ignores layers. Results of this kind can be proven for pointwise errors when using exponentially fitted methods (see Doolan et al. [5]). Such a result cannot be expected for the discretisations of the semiconductor device equations discussed in the preceding section, but this is of minor importance when the goals of device modeling are considered. These goals are basically twofold. One aim is to reveal the solution structure inside the device, the second is to obtain the relation between applied voltages - which enter the Dirichlet boundary conditions - and outflow currents - which are computed by integrals of the current densities along contact segments. Only for the latter part the accuracy of the method is of decisive importance. In this section we prove for a model problem that both aims can be met with reasonable computational effort.

We consider a one-dimensional problem with constant mobilities and vanishing recombination-generation rate. System (1.5) reads

$$
\begin{aligned}
&\lambda^2\psi'' = e^{\psi}u - e^{-\psi}v - C(x)\ ,\\
(3.1)\qquad &(e^{\psi}u')' = 0\ ,\\
&(e^{-\psi}v')' = 0
\end{aligned}
$$

in this case. The simulation domain is $\Omega = (0,1)$. System (3.1) is subject to Dirichlet boundary conditions at $x = 0$ and $x = 1$. We consider an equidistant grid on $[0,1]$. Poisson's equation is discretised by using the common three point formula for the approximation of ψ''. The approximate solution ψ_h is obtained by linear interpolation between the gridpoints.

The Scharfetter-Gummel method amounts to replacing ψ by ψ_h in the continuity equations and solving them explicitely because of the assumptions on μ_n, μ_p and R. Problem (3.1) can be written as a fixed point problem by denoting the solutions of the continuity equations for given ψ

$$u(x) = u(0) + (u(1)-u(0)) \int_0^x e^{-\psi} \Big/ \int_0^1 e^{-\psi} ,$$

(3.2)

$$v(x) = v(0) + (v(1)-v(0)) \int_0^x e^{\psi} \Big/ \int_0^1 e^{\psi}$$

by $u(\psi), v(\psi)$ and the solution of

$$\lambda^2 \phi'' = e^{\phi} u(\psi) - e^{-\phi} v(\psi) - C(x)$$

plus boundary conditions by $\phi = T(\psi)$. A fixed point of the operator T corresponds to a solution.

The discretised problem can be written as

$$\frac{\lambda^2}{h^2} (\psi_{i+1} - 2\psi_i + \psi_{i-1}) = e^{\psi_i} u_i - e^{-\psi_i} v_i - C(x_i) ,$$

(3.3)

$$u_h = u(\psi_h) , \quad v_h = v(\psi_h) .$$

Our convergence analysis will be based on the

Lemma 3.1: Let the Frechet derivative of the operator $(I-T)$ at ψ_h be invertible and the inverse be bounded as operator from $L^1(\Omega)$ to $L^1(\Omega)$ independently of λ and h.
Let $\| \psi_h - T(\psi_h) \|_1$ be sufficiently small, where $\| . \|_p$ denotes the L^p-Norm on $(0,1)$.
Then (3.1) has a locally unique solution ψ^* and

$$\| \psi^* - \psi_h \|_1 \leq K_1 \| \psi_h - T(\psi_h) \|_1$$

with K_1 independent of λ and h holds.

The proof is a straightforward application of the implicit function theorem (For similar results see [8],[12]).

Because of Lemma 3.1 we only have to estimate the L^1-Norm of the error in solving Poisson's equation. This is contained in

Lemma 3.2: Let $C(x)$ have a finite number of jump discontinuities in $[0,1]$ and Lipschitz-continuous first derivatives between those points. Let $u(0), u(1), v(0), v(1) > 0$ hold. Then

$$\| \psi_h - T(\psi_h) \|_1 \leq K_2 (\lambda + h)$$

holds with K_2 independent of λ and h.

Outline of a <u>proof</u>: A priori estimates (see [8],[12]) show that

$$e^{\psi_i} u_i + e^{-\psi_i} v_i \geq K > 0$$

holds for the derivative with respect to ψ_i of the right hand side of (3.3). Thus the discrete operator in (3.3) is of inverse monotone type (see Meis-Markowitz [11]). This allows the use of comparison functions for estimates of the solution. Comparison functions can be constructed which are roughly the sum of a solution of the reduced equation and of terms which decay exponentially away from the boundaries and the discontinuities of the doping profile. The L^1-Norm of the decaying terms can be computed and shown to be of the order $O(\lambda+h)$. The argument that the layer terms in the continuous solution are $O(\lambda)$ with respect to the L^1-Norm completes the proof.

A combination of the above lemmata yields the main result of this section

<u>Theorem 3.3</u>: Let the assumptions of the Lemmata 3.1 and 3.2 hold. If the total current density is denoted by $J = J_n + J_p$, the estimate

$$\|\psi^*-\psi_h\|_1 + \|u^*-u_h\|_\infty + \|v^*-v_h\|_\infty + |J-J_h| \leq K_3(\lambda+h)$$

holds with K_3 independent of λ and h.

<u>Proof</u>: The estimate for the error in the potential follows directly from the preceding lemmata. Considering the representation (3.2) for u and v and

$$J_n = (u(1)-u(0)) \Big/ \int_0^1 e^{-\psi}, \quad J_p = (v(0)-v(1)) \Big/ \int_0^1 e^{\psi}$$

for the current densities the proof of the remaining estimates is also immediate.

Supposedly the above result can be extended to one-dimensional problems with less stringent assumptions one the mobilities and the recombination-generation rate. In the multidimensional situation a similar result cannot be expected to hold because layers in the current densities have to be resolved which requires grid-spacings of the order $O(\lambda)$.

4. NONLINEAR ITERATION METHODS

By discretising (1.5) we obtain a large system of nonlinear algebraic equations. Their solution requires the use of appropriate iteration methods. Although these methods are applied to the discrete problem we discuss them for the continuous equations for notational convenience. Assuming again constant mobilities and vanishing recombination-generation we have to solve

$$
\begin{aligned}
&\lambda^2\Delta\psi - e^{\psi}u + e^{-\psi}v + C(x) = b_1 = 0\ ,\\
(4.1)\qquad &\operatorname{div}(e^{\psi}\,\nabla u) = b_2 = 0\ ,\\
&\operatorname{div}(e^{-\psi}\nabla v) = b_3 = 0\ .
\end{aligned}
$$

Newton's method for (4.1) reads

$$
\begin{aligned}
&\lambda^2\Delta d\psi - (e^{\psi}u+e^{-\psi}v)d\psi - e^{\psi}du + e^{-\psi}dv = -\,b_1\ ,\\
(4.2)\qquad &\operatorname{div}(J_n d\psi + e^{\psi}\nabla du) = -b_2\ ,\\
&\operatorname{div}(J_p d\psi + e^{-\psi}\nabla dv) = -b_3\ .
\end{aligned}
$$

Its application requires the solution of a large linear system in each iteration step. The computational cost can be reduced by "freezing" the Frechet-derivative for several iteration steps. For efficient strategies of this kind and their analysis see Bank and Rose [2]. Their concept of approximate Newton methods allows for perturbations in the Frechet-derivative. A worthwhile goal is to find perturbations which decouple the linear system (4.2) to a certain extent. One method of this kind relies on the assumption that the current densities are comparatively small. Obviously (4.2) is decoupled if J_n and J_p are replaced by zero. The resulting method amounts to solving the continuity equations for given ψ and then the linearized Poisson's equation with the updated u and v in each step. This method was first proposed by Gummel [6]. An alternative which also carries his name is to solve the nonlinear Poisson's equation in each step which can also be seen as the Picard iteration for the fixed point problem $\psi = T(\psi)$ formulated in the preceding section. Convergence analyses of Gummel's method for small current densities are contained in Markowich [8] and Kerkhoven [7].

When the current densities take values of significant size the convergence of Gummel's method often deteriorates. In view of this situation a different kind of decoupling by approximating the Frechet derivative was introduced in Ringhofer and Schmeiser [15]. Here the singular perturbation character of the linearised problem (4.2) is used. As du and dv are slow variables they are approximated well by the solution of the reduced problem. Thus, we substitute

$$\overline{d\psi} = (-e^{\psi}\overline{du} + e^{-\psi}\overline{dv} + b_1)(e^{\psi}u + e^{-\psi}v)^{-1}$$

into the linearized continuity equations

$$\text{(4.3a)} \quad \begin{aligned} &\operatorname{div}\left(\frac{J_n}{e^{\psi}u+e^{-\psi}v}(-e^{\psi}\overline{du}+e^{-\psi}\overline{dv}+b_1) + e^{\psi}\nabla\overline{du}\right) = -b_2 , \\ &\operatorname{div}\left(\frac{J_p}{e^{\psi}u+e^{-\psi}v}(-e^{\psi}\overline{du}+e^{-\psi}\overline{dv}+b_1) + e^{-\psi}\nabla\overline{dv}\right) = -b_3 . \end{aligned}$$

As $d\psi$ is a fast variable, $\overline{d\psi}$ is a good approximation only away from layers. In order to improve on that the full linearised Poisson's equation has to be solved:

$$\text{(4.3b)} \quad \lambda^2\Delta\widetilde{d\psi} - (e^{\psi}u+e^{-\psi}v)\widetilde{d\psi} - e^{\psi}\overline{du} + e^{-\psi}\overline{dv} = -b_1$$

Instead of the Newton corrections $d\psi, du, dv$ we now use $\widetilde{d\psi}, \overline{du}, \overline{dv}$. In the perturbed problem (4.3) Poisson's equation is decoupled from the continuity equations which are coupled to each other by the terms multiplied by J_n and J_p.

Some of the most important semiconductor devices (e.g.MOSFETs) are so called unipolar devices. They are characterised by the property that only one type of charge carriers (i.e. electrons or holes) contributes significantly to the current flow. This means that one current density (for example J_p) is very small compared to the other. This motivates a further decoupling by replacing J_p by zero in (4.3). The resulting method was proven to converge linearly in [15] with a convergence rate of the form

$$\text{(4.4)} \quad \text{const}(c(\lambda)\|J_n\| + \|J_p\|)$$

if the problem is well-posed. In (4.4) $c(\lambda)$ tends to zero as $\lambda \to 0$ and $\|.\|$ denotes a suitable norm. The value of (4.4) is so

small in many applications that the convergence behavior is dominated by the quadratic terms throughout the computations which suggests the use of the term "almost quadratic convergence". The performance of this method was examined in [15] by numerical tests which showed that - compared to Gummel's method - a significant improvement can be achieved.

References:

[1] U.Ascher, P.A.Markowich, C.Schmeiser, H.Steinrück, R.Weiss, Conditioning of the Steady State Semiconductor Device Problem, Techn.Rep.86-18, Comp.Sc., UBC, 1986.

[2] R.E.Banks, D.J.Rose, Global Approximate Newton Methods, Numer.Math.37, 279-295, 1981.

[3] R.E.Bank, D.J.Rose, W.Fichtner, Numerical Methods for Semiconductor Device Simulation, SIAM JSSC 4, No.3, 416-435, 1983.

[4] E.M.Buturla, P.E.Cottrell, Two-Dimensional Finite Element Analysis of Semiconductor Steady Transport Equations, Proc. Int.Conf. "Computer Methods in Nonlinear Mechanics", Austin, Texas, 512-530, 1974.

[5] E.P.Doolan, J.H.H.Miller, W.H.A.Schilders, Uniform Numerical Methods for Problems with Initial and Boundary Layers, Boole Press, Dublin, 1980.

[6] H.K.Gummel, A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations, IEEE Trans.El.Devices, ED-11, 455-465, 1964.

[7] T.Kerkhoven, Convergence of Gummel's Algorithm for Realistic Device Geometries, Proc.NASECODE IV Conf., Boole Press, Dublin, 1985.

[8] P.A.Markowich, The Stationary Semiconductor Device Equations, Springer-Verlag, Wien, 1986.

[9] P.A.Markowich, Spatial-Temporal Structure of Solutions of the Semiconductor Device Problem, to appear in "Lecture Notes on Appl.Math.", 1987.

[10] P.A.Markowich, M.Zlamal, Inverse-Average-Type Finite Element Discretisations of Self-Adjoint Second Order Elliptic Problems, submitted to Math.Comp., 1987.

[11] T.Meis, U.Marcowitz, Numerische Behandlung Partieller Differentialgleichungen, Springer-Verlag, Berlin, 1978.

[12] M.S.Mock, Analysis of Mathematical Models of Semiconductor Devices, Boole Press, Dublin, 1983.

[13] M.S.Mock, Analysis of a Discretisation Algorithm for Stationary Continuity Equations in Semiconductor Device Models I, COMPEL 2, No.3, 117-139, 1983.

[14] C.A.Ringhofer, The Shape of Solutions to the Basic Semiconductor Equations, to appear in "Lecture Notes on Appl.Math.", 1987.

[15] C.A.Ringhofer, C.Schmeiser, An Approximate Newton Method for the Solution of the Basic Semiconductor Device Equations, submitted to SIAM J.Numer.Anal., 1987.

[16] D.L.Scharfetter, H.K.Gummel, Large Signal Analysis of a Silicon Read Diode Oscillator, IEEE Trans.El. Devices, ED-16, 64-77, 1969.

[17] C.Schmeiser, S.Selberherr, R.Weiss, On Scaling and Norms for Semiconductor Device Simulation, Proc. NASECODE IV Conf., Boole Press, Dublin, 1985.

[18] S.Selberherr, Analysis and Simulation of Semiconductor Devices, Springer-Verlag, Wien, 1984.

[19] H.Steinrück, A Bifurcation Analysis of the One Dimensional Steady State Semiconductor Device Equations, submitted to SIAM J.Appl.Math., 1987.

[20] P.Szmolyan, Ein hyperbolisches System aus der Halbleiterphysik, Thesis, Techn.Univ.Wien, 1987.

[21] W.V.Van Roosbroeck, Theory of Flow of Electrons and Holes in Germanium and Other Semiconductors, Bell Syst.Techn.J. 29, 560-607, 1950.