A Singular Perturbation Approach for the Analysis of the Fundamental Semiconductor Equations

PETER A. MARKOWICH, CHRISTIAN A. RINGHOFER, SIEGFRIED SELBERHERR, MEMBER, IEEE, AND MARIANELA LENTINI

Abstract—This paper is concerned with a singular perturbation analysis of the two-dimensional steady-state semiconductor equations and of the usual finite difference scheme consisting of the five point discretization of Poisson's equation and of the Scharfetter–Gummel discretization of the continuity equations. By appropriate scaling we transform the semiconductor equations into a singularly perturbed elliptic system with nonsmooth data. The singular perturbation parameter is defined as the minimal Debeye-length of the device under consideration. Singular perturbation theory allows to distinguish between regions of strong and of weak variation of solutions, so called layers and smooth regions, and to describe solutions qualitatively in these regions. This information is used to analyze the stability and convergence of the discretization scheme. Particular emphasis is put on the construction of efficient grids. It is shown that the Scharfetter–Gummel method is uniformly convergent, i.e., the global error contribution coming from the continuity equations is small when the maximal mesh size is small, independent of the gradient of the solution. Layer jumps are automatically resolved. The five point scheme however is not uniformly convergent. Large gradients of solutions require a graded mesh if solutions inside the layers are to be resolved accurately. This can lead to an intolerably large number of gridpoints. Therefore, we present a modification of the five point scheme which is uniformly convergent.

Key Words—Semiconductor equations, singularly perturbed elliptic systems, internal layer, uniformly convergent difference schemes.

I. INTRODUCTION

IN THIS PAPER we present a singular perturbation analysis of the two-dimensional steady-state semiconductor equations and of the finite difference method used by Franz *et al.* (1982). The singular perturbation approach works as follows. The carrier densities, the doping profile and the independent variables are scaled to (maximally) 0(1) such that Poisson's equation assumes the form

$$\lambda^2 \Delta \psi_s = \rho_s, (x, y) \epsilon \Omega_s \qquad (1.1)$$

where ψ_s, ρ_s is the scaled potential and space-charge, respectively and Ω_s is a domain in R^2 of diameter 0(1) representing the device geometry after scaling. λ is the dimensionless minimal Debeye length which is small if the maximum of the absolute value of the doping profile is large. This is the usual situation for modern devices. Therefore, (1.1) subjected to mixed Dirichlet–Neumann boundary conditions and supplemented by the scaled continuity equations, represents a singularly perturbed elliptic boundary value problem [6] which can be analysed by adapting well-known asymptotic methods (like matched asymptotic expansions). This was done for the one-dimensional static semiconductor equations by D. Smith [17], Markowich *et al.* [9], [10], Vasilieva and Stelmakh [18]. It turns out that in every closed subset of Ω_s, where the doping profile varies "moderately", the solutions of the semiconductor equations are approximated uniformly up to $0(\lambda)$ by smooth, slowly varying functions which are independent of λ and which fulfill the "reduced" equations obtained by setting $\lambda = 0$ in (1.1). Between these subsets there is a curve across which the reduced solutions have a jump-discontinuity. Physically, this curve is the junction between differently doped regions of the device. We derive equations for the limits of the reduced solutions as the independent variables tend to the junction from both sides and show that the jumps of the reduced carrier densities depend exponentially on the potential drop across the junction.

Close to this junction, that is in sets where the doping profile varies strongly, there are thin regions (of width $0(\lambda|\ln\lambda|)$) of rapid variation of the potential and the carrier densities, so called internal layers. Within these layers, the solutions are qualitatively and quantitatively described by the solution of the layer equation, which is a second order ordinary differential equation. The ith derivatives in perpendicular direction to the junction are of the order of magnitude λ^{-i}.

The analysis shows that even large changes of the doping profile within layer regions only cause $0(\lambda)$ changes of the solution outside the layer regions. This property of the semiconductor equations carries over to the discretization scheme. It causes discretization errors occurring in the layers to decay rapidly.

We also show that the electron and hole current density components, which are perpendicular to the junction, do not exhibit layer behavior, while layers may very well occur in the tangential components.

Manuscript received November 2, 1982; revised April 10, 1983. This work was sponsored by the United States Army under Contract DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant MCS-7927062, Mod. 2.

P. A. Markowich and S. Selberherr are with the Institut fuer Angewandte und Numerische Mathematik, Technische Universitaet Wien, A-1040, Wien, Austria.

C. A. Ringhofer is with the Mathematics Research Center, University of Wisconsin-Madison, Madison, WI 53706.

M. Lentini is with the Universidad Simon Bolivar, Caracas, Venezuela.

0018-9383/83/0900-1165$01.00 ©1983 IEEE

Boundary layers can occur where the reduced solutions do not fulfill the boundary conditions asymptotically (as $\lambda \to 0+$). This happens for example at oxide–semiconductor interfaces (for MOS-transistors) and at Schottky contacts but not at ohmic contacts and isolating boundaries.

Using the qualitative and quantitative information on the solutions of the semiconductor equations we analyse the widely used difference scheme which is obtained by discretizing the Laplace operator by the usual five point formula and by applying the Scharfetter–Gummel [14] discretization to the continuity equations. Due to the strongly different behavior of the potential and the carrier densities inside and outside layer regions it is apparent that the construction of grids has to be done with particular care.

We demonstrate, that the chosen discretization of the continuity equations is uniformly convergent, which means that for *every* grid G the global discretization error $e(G)$ fulfills the estimate

$$e(G) \leqslant \text{const.}\,(\bar{h} + \bar{k} + \lambda|\ln\lambda|) \tag{1.2}$$

where $\bar{h}, \bar{k}$ are the maximal mesh sizes in x and y directions, respectively and the constant in (1.2) is independent of the grid and of λ. The Scharfetter–Gummel scheme resolves layer-jumps accurately, even without using a fine grid inside layer regions.

Contrary to this, the five-point discretization of Poisson's equation is not uniformly convergent. The linearized scheme is uniformly stable (i.e., has an inverse which is bounded independently of the grid and of λ), but large discretization errors within layer regions can destroy uniform convergence, particularly when $0(\lambda)$-mesh sizes are chosen.

Therefore, in order to achieve a certain given (global) error tolerance, it is necessary and sufficient to control the grid (only) for Poisson's equation, since the error-contribution from the continuity equations only depends on the maximal grid sizes and on λ.

We show that there are two possibilities of grid-control for the five-point formula. The first is a layer-ignoring grid. That means, all grid sizes are chosen to be much larger than λ which implies that only very few mesh points are located within layer regions. We show, that for such a mesh, the solutions of the discretization scheme of the semiconductor equations converge to the reduced solution and we give an error estimate for this case.

Of course the choice of such a mesh only makes sense if one is not interested in the solutions within layer regions.

Therefore we also derive a layer-resolving mesh, obtained by equidistributing the local discretization error [11], Ascher and Weiss [2], [3]. The construction of this mesh is based upon the fact, that the global error of the scheme is less or equal than the (linear) stability constant times the maximal local discretization error. This requires the information on the exact solution acquired by the singular perturbation analysis.

The so obtained grid is coarse in regions where the solution is approximated by the reduced solution and it is fine within layers in order to balance large derivatives of solutions.

Storage restrictions normally allow one to use the equidistributing mesh if only vertical or horizontal junctions occur, however junctions, which are not parallel to the x- or y-axis usually require too many grid points. In the latter case a rigorous equidistribution is virtually impossible. The reason for this is, that in the case of a horizontal or vertical junction only the y or, respectively x derivative is large (the perpendicular derivative is large, not the tangential), and therefore only the y or respectively x-grid sizes have to be chosen small compared to λ while the mesh sizes in tangential direction may be independent of λ. If the junction is not aligned to the coordinate system, all partial derivatives can get large, then the mesh sizes in both directions have to be small compared to λ within the layer. Nevertheless it is not necessary to abandon this approach. We show, that even if too few grid-points are placed inside such a layer the discrete solutions of the five-point formula are qualitatively correct inside layer regions and they are qualitatively correct in "smooth" regions.

Since this situation is not completely satisfactory we present a modification of the five-point formula, which is uniformly convergent and therefore does not require any grid-restrictions (except sufficiently small $\bar{h}, \bar{k}$). Grid points can be placed wherever the solution is needed. Again, the asymptotic results on the solutions are heavily used.

We also give an analysis of the "finite boxes" approach [7], which allows gridlines to terminate outside layer regions. "Missing" difference quotients are approximated by interpolation. We show that convergence is not influenced.

The presented analysis demonstrates the power of the singular perturbation approach in obtaining information on the analytical solutions of the semiconductor equations as well as in the construction and analysis of numerical methods, particularly as far as grid construction and error estimates are concerned.

II. Singular Perturbation Analysis

As shown by Van Roosbroeck [13] the equations describing potential distribution, carrier and current-distributions in a semiconductor in the two-dimensional static case are

$$\epsilon\Delta\psi = q(n - p - C) \text{ (Poisson's equation)} \tag{2.1}$$

$$J_n = -q(\mu_n n \operatorname{grad}\psi - D_n \operatorname{grad} n) \text{ (electron current relation)} \tag{2.2}$$

$$J_p = -q(\mu_p p \operatorname{grad}\psi + D_p \operatorname{grad} p) \text{ (hole current relation)} \tag{2.3}$$

$$\operatorname{div} J_n = qR(n, p) \text{ (electron-continuity equation)} \tag{2.4}$$

$$\operatorname{div} J_p = -qR(n, p) \text{ (hole-continuity equation)} \tag{2.5}$$

for $(x, y) \in \Omega \subset R^2$ (where Ω is a bounded, convex domain representing the device geometry) subject to Dirichlet boundary conditions on Γ_c (ohmic contacts) and homogeneous Neumann boundary conditions for ψ, n, p on Γ_i

(isolating boundaries) with $\partial\Omega = \Gamma_c \cup \Gamma_i$. Also Dirichlet boundary conditions for $J \equiv J_n + J_p$ and oxide–semiconductor interface conditions can be desired for the simulation of certain devices.

However, the exact formulation of the boundary conditions is not necessary for our purposes since we only investigate internal layer phenomena. The numerical and analytical treatment of boundary layers is completely analgous.

We take the Shockley–Read–Hall (SRH) thermal recombination term

$$R(n,p) = \frac{np - n_i^2}{\tau_p(n+n_i) + \tau_n(p+n_i)}. \tag{2.6}$$

In order to model high-injection conditions, the SRH term has to be supplemented by more complicated generation–recombination terms (see Schütz *et al.* [15]). We also assume the validity of Einstein's relation

$$D_n = \mu_n U_T,\ D_p = \mu_p U_T. \tag{2.7}$$

For simplicity we assume that D_n, D_p and μ_n, μ_p are constants. In "reality" they are weakly varying functions of $n, p, \operatorname{grad}\psi$ and of the doping profile. This does not influence the following singular perturbation analysis.

An existence theorem for (2.1)–(2.5) under simplifying assumptions on the boundary data and on the device geometry is given by Mock [12]. However, no qualitative information on the solutions can be obtained from this theorem.

Let l be the characteristic length of the device under investigation (for example the length of a diode).

The following scaling is basic for the singular perturbation approach

$$\psi_s = \frac{\psi}{U_T},\ n_s = \frac{n}{\bar{C}},\ p_s = \frac{p}{\bar{C}},\ D = \frac{C}{\bar{C}} \tag{2.8}$$

$$J_{n_s} = \frac{l J_n}{D_n q \bar{C}},\ J_{p_s} = \frac{l J_p}{D_p q \bar{C}},\ \phi_{n_s} = \frac{\phi_n}{U_T},\ \phi_{p_s} = \frac{\phi_p}{U_T} \tag{2.9}$$

where $\bar{C} := \max_{(x,y)\in\Omega\cup\partial\Omega}|C(x,y)|$ and

$$x_s = \frac{x}{l},\ y_s = \frac{y}{l}. \tag{2.10}$$

Then (2.1)–(2.5) reads, after dropping the subscript s

$$\lambda^2 \Delta\psi = n - p - D(x,y,\lambda) \tag{2.11}$$

$$J_n = -n \operatorname{grad}\psi + \operatorname{grad} n \tag{2.12}$$

$$J_p = -p \operatorname{grad}\psi - \operatorname{grad} p \tag{2.13}$$

$$\operatorname{div} J_n = \beta_n S(n,p,\gamma\lambda) \tag{2.14}$$

$$\operatorname{div} J_p = -\beta_p S(n,p,\gamma\lambda) \tag{2.15}$$

with

$$\lambda^2 = \left(\frac{\lambda_D}{l}\right)^2 = \frac{\epsilon U_T}{l^2 q \bar{C}},\ \gamma^2 = \frac{l q n_i}{\epsilon U_T},\ \beta_n = \frac{l^2}{D_n \tau_n},\ \beta_p = \frac{l^2}{D_p \tau_p} \tag{2.16}$$

and

$$S(n,p,\gamma\lambda) = \frac{np - \gamma^4\lambda^4}{n + p + 2\gamma^2\lambda^2}. \tag{2.17}$$

Equations (2.11)–(2.15) holds in the domain $\Lambda = \{(x,y) | (lx, ly) \in \Omega\}$ and is subjected to (the scaled) boundary conditions. λ_D is the minimal Debeye length. When $\mu_{n,p}, D_{n,p}$ are not constant D_n, D_p in (2.9) have to be substituted by "characteristic" values $\bar{D}_n, \bar{D}_p$ with $D_{n,p} = 0(\bar{D}_{n,p})$. The scaled continuity equations (2.12), (2.13) have to be changed accordingly.

In the sequel we also need the scaled Boltzmann statistics

$$n = \gamma^2\lambda^2 e^{\psi - \phi_n},\ p = \gamma^2\lambda^2 e^{\phi_p - \psi}.$$

Now assume we deal with a silicon device with characteristic length $l = 2.5\times10^{-3}$ cm and the doping is such that $\bar{C} = 10^{17}$ cm^{-3}. Then, at approximately room temperature $T \approx 300$ K we compute the following numerical values

$$\lambda^2 \approx 0.4\times10^{-6},\ \gamma^2 \approx 0.25,\ \beta_n \approx \beta_p \approx 0.25.$$

Obviously $\lambda \ll 1$ holds while the other constants are 0(1). Also, the larger the maximal doping gets (in absolute value), the smaller λ gets. For example, $\bar{C} = 10^{20}$ cm^{-3} gives $\lambda^2 \approx 0.4\times10^{-9}$. Therefore, for sufficiently large doping, we can regard (2.11)–(2.15) as singularly perturbed elliptic system with singular perturbation parameter $\lambda = \lambda_D/l$.

Markowich *et al.* [10] pointed out that the singular perturbation analysis requires that the intrinsic number n_i is much less than the maximal doping $\bar{C}$ and that $C(x,y)/\bar{C}$ is much larger than λ^2 except in domains of small area (i.e., layers) and, of course that λ is sufficiently small.

In order to analyze internal layer phenomena we assume that the scaled doping profile D has an internal layer along the y-axis

$$D(x,y,\lambda) = \bar{D}(x,y) + \hat{D}\left(\frac{x}{\lambda}, y\right) + 0(\lambda).$$

$\bar{D}, \hat{D}$ have a jump discontinuity across the y-axis, i.e.,

$$\lim_{x\to0+} \bar{D}(x,y) \neq \lim_{x\to0-} \bar{D}(x,y);$$

$$\lim_{x\to0+} \hat{D}\left(\frac{x}{\lambda}, y\right) \neq \lim_{x\to0-} \hat{D}\left(\frac{x}{\lambda}, y\right) \tag{2.18}$$

for all y for which $(0,y)\in\Lambda$ and

$$\left|\hat{D}\left(\frac{x}{\lambda}, y\right)\right| \leq c_1 e^{-c_2|x/\lambda|} \tag{2.19}$$

holds for some constants $c_1 > 0, c_2 > 0$ independent of x, y, λ.

Physically this represents a vertical p-n junction along the y-axis. If $\hat{D} \equiv 0$ then the junction is abrupt, if (2.19) holds with $c_1 \neq 0$ then D is assumed to be continuous and the junction is exponentially graded.

We use a vertical junction because this heavily simplifies the following analysis. Later on we state the generalization to curved junctions.

We expect the internal layer in the doping to cause layers in the dependent variables and employ the "ansatz"

$$\begin{aligned}
&\text{(a)} \quad \psi(x,y,\lambda)=\bar{\psi}(x,y)+\hat{\psi}\Big(\frac{x}{\lambda},y\Big)+\cdots\\
&\text{(b)} \quad n(x,y,\lambda)=\bar{n}(x,y)+\hat{n}\Big(\frac{x}{\lambda},y\Big)+\cdots\\
&\text{(c)} \quad p(x,y,\lambda)=\bar{p}(x,y)+\hat{p}\Big(\frac{x}{\lambda},y\Big)+\cdots\\
&\text{(d)} \quad J_n(x,y,\lambda)=\bar{J}_n(x,y)+\hat{J}_n\Big(\frac{x}{\lambda},y\Big)+\cdots\\
&\text{(e)} \quad J_p(x,y,\lambda)=\bar{J}_p(x,y)+\hat{J}_p\Big(\frac{x}{\lambda},y\Big)+\cdots \qquad (2.20)
\end{aligned}$$

where the dots denote a power series in λ starting with the $0(\lambda)$ terms.

The functions marked with "$\bar{}$" are independent of λ and denote the reduced solutions, the functions marked with "$\hat{}$" are the layer terms which are defined for $\hat{\tau}=x/\lambda\epsilon(-\infty,\infty)$ and all y in the device.

The layer terms are supposed to decay away from both sides of the junction

$$\hat{\psi}(\pm\infty,y)=\hat{n}(\pm\infty,y)=\hat{p}(\pm\infty,y)=0 \qquad (2.21a)$$

$$\hat{J}_n(\pm\infty,y)=\hat{J}_p(\pm\infty,y)=0. \qquad (2.21b)$$

Now we insert the expansions (2.20) into the semiconductor equations (2.11)–(2.14).

Neglecting $0(\lambda)$ terms and evaluating away from the junction gives the reduced problem

$$0=\bar{n}-\bar{p}-\bar{D} \qquad (2.22)$$

$$\bar{J}_n=-\bar{n}\,\mathrm{grad}\,\bar{\psi}+\mathrm{grad}\,\bar{n} \qquad (2.23)$$

$$\bar{J}_p=-\bar{p}\,\mathrm{grad}\,\bar{\psi}-\mathrm{grad}\,\bar{p}\ (x,y)\epsilon\Lambda,\quad x\neq 0 \qquad (2.24)$$

$$\mathrm{div}\bar{J}_n=\beta_n S(\bar{n},\bar{p},0) \qquad (2.25)$$

$$\mathrm{div}\bar{J}_p=-\beta_p S(\bar{n},\bar{p},0). \qquad (2.26)$$

The reduced solutions $\bar{n}$, $\bar{p}$, $\bar{\psi}$ are discontinuous along $x=0$ because $\bar{D}$ has a jump discontinuity there. In the sequel we denote a vector $a=\binom{a^x}{a^y}\epsilon R^2$. Evaluating close to the junction, but to the left of $x=0$ and comparing $0(1/\lambda)$ terms gives the left layer problem

$$\begin{aligned}
&\text{(a)} \quad \hat{\psi}_{\tau\tau}=\hat{n}-\hat{p}-\hat{D}\\
&\text{(b)} \quad \hat{n}_\tau=(\hat{n}+\bar{n}(0-,y))\hat{\psi}_\tau\\
&\text{(c)} \quad \hat{p}_\tau=-(\hat{p}+\bar{p}(0-,y))\hat{\psi}_\tau \qquad -\infty<\tau<0\\
&\text{(d)} \quad \hat{J}^x_{n\tau}=0\\
&\text{(e)} \quad \hat{J}^x_{p\tau}=0 \qquad (2.27)
\end{aligned}$$

where we set $f(0\pm)=\lim_{x\to 0\pm}f(x)$. Subscripts denote partial derivatives. Analogously we obtain the right internal layer problem

$$\begin{aligned}
&\text{(a)} \quad \psi_{\tau\tau}=\hat{n}-\hat{p}-\hat{D}\\
&\text{(b)} \quad \hat{n}_\tau=(\hat{n}+\bar{n}(0+,y))\hat{\psi}_\tau\\
&\text{(c)} \quad \hat{p}_\tau=-(\hat{p}+\bar{p}(0+,y))\hat{\psi}_\tau \qquad 0<\tau<\infty\\
&\text{(d)} \quad \hat{J}^x_{n\tau}=0\\
&\text{(e)} \quad \hat{J}^x_{p\tau}=0. \qquad (2.28)
\end{aligned}$$

The decay of the layer terms and (2.27), (2.28d), (2.28e) immediately imply that

$$\hat{J}^x_n\equiv\hat{J}^x_p\equiv 0 \qquad (2.29)$$

holds. The current components normal to the junction have no layers.

Because of the discontinuity of $\bar{D}$ we need interface conditions at $x=0$. These conditions are obtained from the (natural) assumption that the solutions ψ, n, p, J_n, J_p of (2.11)–(2.15) and grad ψ are continuous along the y-axis. This implies that the sum of the reduced and layer components of the expansions (2.20) have to be continuous across the y-axis

$$\begin{aligned}
&\text{(a)} \quad \bar{\psi}(0-,y)+\hat{\psi}(0-,y)=\bar{\psi}(0+,y)+\hat{\psi}(0+,y)\\
&\text{(b)} \quad \bar{n}(0-,y)+\hat{n}(0-,y)=\bar{n}(0+,y)+\hat{n}(0+,y)\\
&\text{(c)} \quad \bar{p}(0-,y)+\hat{p}(0-,y)=\bar{p}(0+,y)+\hat{p}(0+,y)\\
&\text{(d)} \quad \bar{J}^x_n(0-,y)=\bar{J}^x_n(0+,y)\\
&\text{(e)} \quad \bar{J}^y_n(0-,y)+\hat{J}^y_n(0-,y)=\bar{J}^y_n(0+,y)+\hat{J}^y_n(0+,y)\\
&\text{(f)} \quad \bar{J}^x_p(0-,y)=\bar{J}^x_p(0+,y)\\
&\text{(g)} \quad \bar{J}^y_p(0-,y)+\hat{J}^y_p(0-,y)=\bar{J}^y_p(0+,y)+\hat{J}^y_p(0+,y)
\end{aligned} \qquad (2.30)$$

and by comparing $0(1/\lambda)$ terms

$$\hat{\psi}_\tau(0-,y)=\hat{\psi}_\tau(0+,y).$$

Integrating (2.27), (2.28b), (2.28c) gives

$$\hat{n}(\tau,y)=\begin{cases}\bar{n}(0+,y)(e^{\hat{\psi}(\tau,y)}-1), & \tau>0\\ \bar{n}(0-,y)(e^{\psi(\tau,y)}-1), & \tau<0\end{cases} \qquad (2.31)$$

$$\hat{p}(\tau,y)=\begin{cases}\bar{p}(0+,y)(e^{-\hat{\psi}(\tau,y)}-1), & \tau>0\\ \bar{p}(0-,y)(e^{-\hat{\psi}(\tau,y)}-1), & \tau<0\end{cases} \qquad (2.32)$$

and (2.30a), (2.30b), (2.30c) give the interface conditions for the reduced carrier densities

$$\begin{aligned}
&\text{(a)} \quad \bar{n}(0+,y)=e^{\bar{\psi}(0+,y)-\bar{\psi}(0-,y)}\bar{n}(0-,y)\\
&\text{(b)} \quad \bar{p}(0+,y)=e^{\bar{\psi}(0-,y)-\bar{\psi}(0+,y)}\bar{p}(0-,y).
\end{aligned} \qquad (2.33)$$

The layer jumps of the carrier densities depend exponentially on the voltage drop across the junction.

The reduced problem (2.22)–(2.26) can be written as nonlinear elliptic system (after a lengthy, but simple calculation)

$$\text{(a)}\quad \Delta\bar{\psi} + \frac{1}{2\bar{n}-\bar{D}}\operatorname{grad}(2\bar{n}-\bar{D})\cdot\operatorname{grad}\bar{\psi} - \frac{\Delta\bar{D}}{2\bar{n}-\bar{D}} - \frac{\beta_n-\beta_p}{2\bar{n}-\bar{D}}S(\bar{n},\bar{n}-\bar{D},0) = 0$$

$$\text{(b)}\quad \Delta\bar{n} - (\operatorname{grad}\bar{n} - \frac{n}{2\bar{n}-\bar{D}}\operatorname{grad}(2\bar{n}-\bar{D}))\cdot\operatorname{grad}\bar{\psi} - \frac{\bar{n}}{2\bar{n}-\bar{D}}(\Delta\bar{D} - (\beta_n-\beta_p)S(\bar{n},\bar{n}-\bar{D},0)) = \beta_n S(\bar{n},\bar{n}-\bar{D},0)$$

for $(x, y)\in\Lambda$, $x\neq 0$ subject to the Neumann and Dirichlet interface condition

$$\text{(c)}\quad \bar{n}(0+,y) = e^{\bar{\psi}(0+,y)-\bar{\psi}(0-,y)}\bar{n}(0-,y)$$

$$\text{(d)}\quad \bar{n}(0+,y) - \bar{D}(0+,y) = e^{\bar{\psi}(0-,y)-\bar{\psi}(0+,y)}(\bar{n}(0-,y) - \bar{D}(0-,y))$$

$$\text{(e)}\quad \bar{n}(0+,y)\bar{\psi}_x(0+,y) - \bar{n}_x(0+,y) = \bar{n}(0-,y)\bar{\psi}_x(0-,y) - \bar{n}_x(0-,y)$$

$$\text{(f)}\quad (\bar{n}(0+,y) - \bar{D}(0+,y))\bar{\psi}_x(0+,y) + \bar{n}_x(0+,y) - \bar{D}_x(0+,y) = (\bar{n}(0-,y) - \bar{D}(0-,y))\bar{\psi}_x(0-,y) + \bar{n}_x(0-,y) - \bar{D}_x(0-,y). \quad (2.34)$$

Equations (2.34e), (2.34f) correspond to (2.30d), (2.30f).

The reduced problem has to be supplemented by boundary conditions (on $\partial\Lambda$) derived by setting $\lambda = 0$ in the scaled boundary conditions for the full problem (2.11)–(2.15) on ohmic contacts and isolating boundaries and by making boundary layer expansions on oxide–semiconductor interfaces and Schottky contacts similarly to (2.20).

The reduced hole density $\bar{p}$ can be computed from the vanishing space-charge conditions (2.22).

The internal layer problem is obtained from (2.27), (2.28a), (2.31), (2.32)

$$\text{(a)}\quad \hat{\psi}_{\tau\tau} = \bar{n}(0+,y)e^{\hat{\psi}(\tau,y)} - \bar{p}(0+,y)e^{-\hat{\psi}(\tau,y)} - (\bar{D}(0+,y) + \hat{D}(\tau,y)), \qquad \tau > 0$$

$$\text{(b)}\quad \hat{\psi}_{\tau\tau} = \bar{n}(0-,y)e^{\hat{\psi}(\tau,y)} - \bar{p}(0-,y)e^{-\hat{\psi}(\tau,y)} - (\bar{D}(0-,y) + \hat{D}(\tau,y)), \qquad \tau < 0$$

$$\text{(c)}\quad \hat{\psi}(\infty,y) = \hat{\psi}(-\infty,y) = 0$$

$$\text{(d)}\quad \hat{\psi}(0+,y) - \hat{\psi}(0-,y) = \bar{\psi}(0-,y) - \bar{\psi}(0+,y)$$

$$\text{(e)}\quad \hat{\psi}_\tau(0+,y) = \hat{\psi}_\tau(0-,y). \quad (2.35)$$

Markowich et al. [10] showed that the ordinary second-order differential equation (2.35) possesses a unique solution $\hat{\psi}$ which decays exponentially as $\tau\to\pm\infty$ if (2.18), (2.19) holds. The derivatives fulfill

$$\left|\frac{\partial^i\hat{\psi}}{\partial\tau^i}(\tau,y)\right| \leqslant K_1 e^{-K_2|\tau|} \quad (2.36)$$

where $K_1, K_2 > 0$ only depend on i. The n- and p-layer terms have to be computed from (2.31), (2.32), they also decay exponentially as $\tau\to\pm\infty$.

After having shown the validity of the expansion (2.20) (see Markowich *et al.* [10] for the one-dimensional problem) it follows that the solutions of the reduced problem (which are smooth, slowly varying functions) approximate the solutions of the semiconductor equations up to $0(\lambda)$ outside the layer region, which is a strip of width $0(\lambda|\ln\lambda|)$ about the y-axis. Within this internal layer the solutions vary rapidly, for example

$$\psi(x,y,\lambda) \sim \begin{cases} \bar{\psi}(0+,y) + \hat{\psi}\left(\frac{x}{\lambda},y\right), & x>0 \\ \bar{\psi}(0-,y) + \hat{\psi}\left(\frac{x}{\lambda},y\right), & x<0. \end{cases}$$

There the solutions are, up to the reduced solution evaluated at the junction, given by the exponentially decaying (in $\tau = x/\lambda$) layer terms. The ith derivative in perpendicular direction to the junction fulfills

$$\left|\frac{\partial^i}{\partial x^i}\psi(x,y,\lambda)\right| \leqslant D_1\lambda^{-i}e^{-D_2|x/\lambda|} \quad (2.37)$$

for $D_1, D_2 > 0$ independent of λ while the tangential derivatives (y-direction) are 0(1). The same statement holds for n and p, but not for the "slow" variables J_n^x, J_p^x, which do not exhibit an internal layer. They are; even close to the y-axis, approximated up to 0(1) by the corresponding reduced solutions $\bar{J}_n^x$, $\bar{J}_p^x$.

So far our analysis did not give any information on the tangential current components J_n^y, J_p^y except that they might exhibit layer behavior.

A complicated asymptotic analysis (given in [8]), which uses the substitution

$$n = e^{\psi}u,\quad p = e^{-\psi}v \quad (2.38)$$

shows that

$$\hat{J}_n^y(\tau,y) = \begin{cases} e^{\bar{\psi}(0+,y)}\left(e^{\hat{\psi}(\tau,y)}-1\right)\bar{u}_y(0,y), & \tau>0 \\ e^{\bar{\psi}(0-,y)}\left(e^{\hat{\psi}(\tau,y)}-1\right)\bar{u}_y(0,y), & \tau<0 \end{cases} \quad (2.39)$$

$$\hat{J}_p^y(\tau,y) = \begin{cases} e^{-\bar{\psi}(0+,y)}\left(e^{-\hat{\psi}(\tau,y)}-1\right)\bar{v}_y(0,y), & \tau>0 \\ e^{-\bar{\psi}(0-,y)}\left(e^{-\hat{\psi}(\tau,y)}-1\right)\bar{v}_y(0,y), & \tau<0 \end{cases} \quad (2.40)$$

where $\bar{u}, \bar{v}$ denote the reduced solutions u, v (which are continuous across $x = 0$ because of (2.33)). Therefore, if the

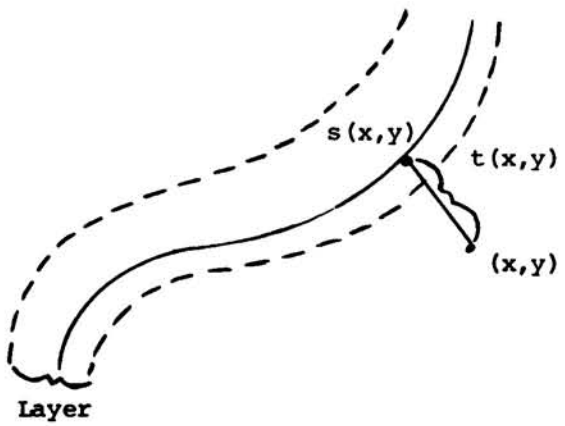

Fig. 1. Local coordinates.

tangential derivatives of the scaled n and p quasi-fermi levels ϕ_n, ϕ_p at the junction do not vanish, then the tangential current density components exhibit layer behavior.

All results immediately carry over to a horizontal junction by exchanging the x and y coordinate, however a curved junction extremely complicates the analysis. Therefore we only state the results, the proofs can be found in [8].

Assume that there is a curve $\Gamma \subset \Lambda$ along which the (scaled) profile D has a jump discontinuity (abrupt junction) or decays exponentially (as in (2.19)). Then, for points (x, y) sufficiently close to the junction we denote by $t(x, y)$ the closest oriented distance to Γ (that means t is positive on one side of Γ and negative on the other) and by $s(x, y) = (s_1(x, y), s_2(x, y))$ the point on Γ closest to (x, y) (see Fig. 1).

For $(x, y) \in \Gamma$ we call $\vec{n}(x, y) = \begin{pmatrix} t_x(x,y) \\ t_y(x,y) \end{pmatrix}$ the normal vector to Γ and $\vec{t}(x, y) = \begin{pmatrix} -t_y(x,y) \\ t_x(x,y) \end{pmatrix}$ the tangential vector and set $\tau = t(x, y)/\lambda$. It turns out that the asymptotic expansions (2.20) remain valid when x is substituted by $t(x, y)$ and y by $s(x, y)$ in the layer terms. That means, the layer is a strip of width $0(\lambda|\ln\lambda|)$ about Γ (see Fig. 1). The exponential decay of the layer terms occurs in perpendicular direction to the junction. Both partial derivatives are large if $t_x(x, y), t_y(x, y)$ are nonzero, since

$$\text{(a)}\quad \frac{\partial\psi(x,y,\lambda)}{\partial x} \sim \frac{\partial\bar{\psi}(x,y)}{\partial x} + \frac{1}{\lambda}\frac{\partial\hat{\psi}\left(\frac{t}{\lambda}, s\right)}{\partial\tau} t_x + \frac{\partial\hat{\psi}\left(\frac{t}{\lambda}, s\right)}{\partial\sigma} s_x + 0(1)$$

$$\text{(b)}\quad \frac{\partial\psi(x,y,\lambda)}{\partial y} \sim \frac{\partial\bar{\psi}(x,y)}{\partial y} + \frac{1}{\lambda}\frac{\partial\hat{\psi}\left(\frac{t}{\lambda}, s\right)}{\partial\tau} t_y + \frac{s\hat{\psi}\left(\frac{t}{\lambda}, s\right)}{\partial s} s_x + 0(1) \qquad (2.41)$$

holds where $\hat{\psi}$ and $\partial\hat{\psi}/\partial\tau$ fulfill (2.36) with y substituted by s. The first and the last terms on the right-hand side of (2.41) are 0(1). Moreover $\bar{J}_n \cdot \vec{n}$, $\bar{J}_p \cdot \vec{n}$ are continuous across the junction, while $J_n \cdot \vec{t}$, $J_p \cdot \vec{t}$ generally have layers.

The current density components in normal direction have no layer, while the tangential components may very well have one (depending on whether the tangential derivative of the quasifermi levels vanish or not). Details are given in [3].

III. The Basic Discretization Scheme

For the discretization of the (scaled) semiconductor equations (2.11)–(2.15) we use a rectangular grid G with grid sizes h_i in x-direction and k_j in y-direction. The grid points have coordinates (x_i, y_j) where

$$\text{(a)}\quad x_{i+1} = x_i + h_i,\ y_{j+1} = y_j + k_j \qquad (3.1)$$

holds. The subscripts are chosen such that no gridline left of $x = x_0$ and below $y = y_0$ intersects $\Lambda \cup \partial\Lambda$. i runs from 0 to maximally N and j from 0 to maximally M.

The point (x_i, y_j) is called interior gridpoint if its four neighbors (x_{i-1}, y_j), (x_{i+1}, y_j), (x_i, y_{j+1}), (x_i, y_{j-1}) are in $\Lambda \cup \partial\Lambda$, otherwise it is called exterior gridpoint.

In the sequel we denote the evaluation of a gridfunction f at (x_i, y_j) by f_{ij} and we define

$$\text{(b)}\quad \bar{h} = \max_i h_i,\ \bar{k} = \max_j k_j. \qquad (3.1)$$

We use the standard five-point formula for Poisson's equation (2.11)

$$\text{(a)}\quad \lambda^2\left[\frac{2}{h_i + h_{i-1}}\left(\frac{\psi_{i+1,j} - \psi_{ij}}{h_i} - \frac{\psi_{ij} - \psi_{i-1,j}}{h_{i-1}}\right) + \frac{2}{k_j + k_{j-1}}\left(\frac{\psi_{i,j+1} - \psi_{ij}}{k_j} - \frac{\psi_{ij} - \psi_{i,j-1}}{k_{j-1}}\right)\right] = n_{ij} - p_{ij} - D(x_i, y_j, \lambda). \qquad (3.2)$$

For the discretization of the current relations and the continuity equations we need the function

$$\text{(b)}\quad \gamma(z) = \left(\frac{z}{2}\right)\coth z. \qquad (3.2)$$

We discretize the current relations (2.12)–(2.15) by the following difference scheme, obtained by "exponential interpolation"

$$\text{(c)}\quad J^x_{nij} = \gamma\left(\frac{\psi_{i+1,j} - \psi_{ij}}{2}\right)\frac{n_{i+1,j} - n_{ij}}{h_i} - \frac{\psi_{i+1,j} - \psi_{ij}}{h_i}\frac{n_{i+1,j} + n_{ij}}{2}$$

$$\text{(d)}\quad J^y_{nij} = \gamma\left(\frac{\psi_{i,j+1} - \psi_{ij}}{2}\right)\frac{n_{i,j+1} - n_{ij}}{k_j} - \frac{\psi_{i,j+1} - \psi_{ij}}{k_j}\frac{n_{i,j+1} + n_{ij}}{2}$$

$$\text{(e)} \quad J^x_{pij} = -\gamma\left(\frac{\psi_{i,j+1}-\psi_{ij}}{2}\right)\frac{p_{i+1,j}-p_{ij}}{h_i} - \frac{\psi_{i+1,j}-\psi_{ij}}{h_i}\,\frac{p_{i+1,j}+p_{ij}}{2}$$

$$\text{(f)} \quad J^y_{pij} = -\gamma\left(\frac{\psi_{i,j+1}-\psi_{ij}}{2}\right)\frac{p_{i,j+1}-p_{ij}}{k_j} - \frac{\psi_{i,j+1}-\psi_{ij}}{k_j}\,\frac{p_{i,j+1}+p_{ij}}{2}$$

$$\text{(g)} \quad \frac{2}{h_i+h_{i-1}}(J^x_{n_{i,j}} - J^x_{n_{i-1,j}}) + \frac{2}{k_j+k_{j-1}}(J^y_{n_{i,j}} - J^y_{n_{i,j-1}}) = \beta_n S(n_{ij}, p_{ij}, \lambda)$$

$$\text{(h)} \quad \frac{2}{h_i+h_{i-1}}(J^x_{p_{i,j}} - J^x_{p_{i-1,j}}) + \frac{2}{k_j+k_{j-1}}(J^y_{p_{i,j}} - J^y_{p_{i,j-1}}) = \beta_p S(n_{ij}, p_{ij}, \lambda) \tag{3.2}$$

Equation (3.2) only holds at interior gridpoints. For the actual computation of solutions (3.2c), (3.2d) and (3.2e), (3.2f) are inserted into (3.2g) and (3.2h), respectively.

It is an easy exercise to show that the difference scheme (3.2c)–(3.2h) is equivalent to the Scharfetter–Gummel discretization [14], however the form given in (3.2) is more convenient for the analysis. Equation (3.2) is used by many authors (see, for example Selberherr [16] and Franz *et al.* [7]).

The boundary conditions are discretized in the normal way (see Franz *et al.* [7]).

In the following sections we apply the results from the singular perturbation analysis to construct grids which allow the global error of the discretization scheme (3.2) to be less than a prescribed tolerance without employing too many unnecessary grid points.

IV. Analysis of the Scharfetter–Gummel (SG) Method

The use of the function $\gamma(z)$ as a factor multiplying the discrete derivatives of n and p in (3.2c)–(3.2f) appears striking at the first glance. Originally the discretizations of the continuity equations were obtained by using the exponential dependence of the carrier densities on the potential. However, by applying the exponential fitting methods described in Section V for Poisson's equation it turns out that (up to small terms) the SG method is the only discretization of the continuity equation whose (discrete) solutions have the same asymptotic properties (as $\lambda \to 0+$) as the "continuous" solutions for *every grid* with $\bar{h}, \bar{k}$ sufficiently small but independent of λ.

Taylor series expansion shows that

$$\gamma(z) = 1 + 0(z^2) \text{ as } z \to 0. \tag{4.1}$$

Therefore, in regions where ψ_{ij} differs only by $0(\bar{h}+\bar{k})$ from its four neighboring values $\psi_{i+1,j}, \psi_{i-1,j}, \psi_{i,j+1}, \psi_{i,j-1}$ (for example outside layers), the discretization (3.2c)–(3.2f) is up to $0((\bar{h}+\bar{k})^2)$ equal to the standard trapezoidal rule type-discretization of the current relations (with $\gamma \equiv 1$). However, when ψ_{ij} differs significantly from one of its four neighbors, the SG method behaves differently. To demonstrate this we investigate how the SG scheme would solve the continuity equation and current relation for n and J_n if ψ is a given function with the qualitative properties derived in Section II. For this investigation we restrict ourselves to the one-dimensional case. It should be noted, that for the two-dimensional case the analysis becomes much more complicated. However numerical experiments show, that the following results carry over to two dimensions even in the case of general curved *p-n* junctions. The analysis for the two-dimensional case will be reported in a separate paper. The continuity equations—current relations in one dimension are of the form (assuming $R \equiv 0$)

$$\text{(a)} \quad n' - \psi' n = J_n, \qquad -1 \leqslant x \leqslant 1$$

$$\text{(b)} \quad J_n' = 0, \qquad -1 \leqslant x \leqslant 1$$

$$\text{(c)} \quad n(-1) = n_-, n(1) = n_+ \tag{4.2}$$

We assume that the potential $\psi = \psi(x, \lambda)$ is a given function with an exponential internal layer at $x = 0$ of width $0(\lambda|\ln\lambda|)$ (see Fig. 2) as given by the singular perturbation analysis of Section II.

The problem (4.2) can be thought of as modeling a *p-n* junction device where the junction between n and p regions is at $x = 0$. For the sake of simplicity the recombination rate has been set to 0 which makes sense close to thermal equilibrium.

A simple calculation gives the exact solution of (4.2)

$$n(x,\lambda) = e^{\psi(x,\lambda)-\psi(-1,\lambda)}n_- + J_n e^{\psi(x,\lambda)}\int_{-1}^{x} e^{-\psi(s,\lambda)}\,ds \tag{4.3}$$

$$J_n \equiv \frac{n_+ - e^{\psi(1,\lambda)-\psi(-1,\lambda)}n_-}{e^{\psi(1,\lambda)}\displaystyle\int_{-1}^{x} e^{-\psi(s,\lambda)}\,ds} \tag{4.4}$$

At first we investigate the SG method applied to the homogeneous initial value problem

$$n' = \psi' n, \qquad x \geqslant -1, \qquad n(-1) = n_- \tag{4.5}$$

which has the solution

$$n(x,\lambda) = n_- e^{\psi(x,\lambda)-\psi(-1,\lambda)}. \tag{4.6}$$

The SG scheme for (4.5) reads

$$\gamma\left(\frac{\psi_{i+1}-\psi_i}{2}\right)\frac{n_{i+1}-n_i}{h_i} = \frac{\psi_{i+1}-\psi_i}{h_i}\,\frac{n_{i+1}+n_i}{2}, \qquad i \geqslant 0;\ n_0 = n_-. \tag{4.7}$$

We obtain the recursion

$$n_{i+1} = \sigma\left(\frac{\psi_{i+1}-\psi_i}{2}\right)n_i \tag{4.8}$$

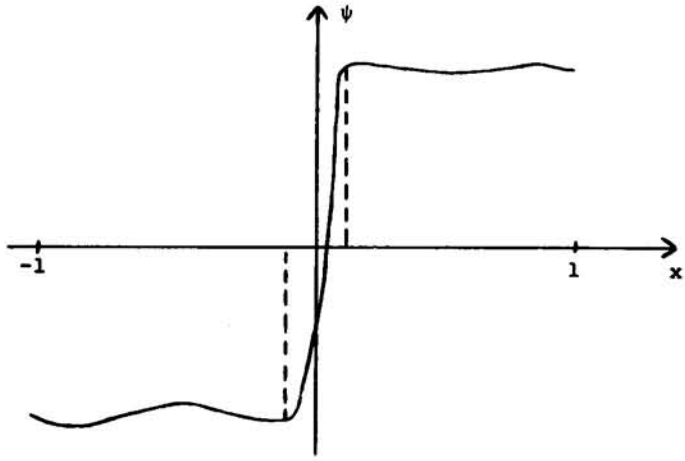

Fig. 2. Potential.

with the growth function

$$\sigma(z) = e^{2z}. \tag{4.9}$$

Therefore the solution n_i is given by

$$n_i = \prod_{j=0}^{i-1} \sigma\left(\frac{\psi_{j+1} - \psi_i}{2}\right) n_- = e^{\psi_i - \psi_0} n_- = e^{\psi(x_i,\lambda) - \psi(-1,\lambda)} n_-. \tag{4.10}$$

The SG method intergrates the initial value problem (4.5) exactly on *every* grid. Internal layer jumps are resolved accurately.

The trapezoidal rule, that is (4.6) with $\gamma \equiv 1$, gives

$$n_{i+1} = \phi\left(\frac{\psi_{i+1} - \psi_i}{2}\right) n_i \tag{4.11}$$

with

$$\phi(z) = \frac{1+z}{1-z}. \tag{4.12}$$

The solution of the trapezoidal rule is

$$n_i = \prod_{j=0}^{i-1} \phi\left(\frac{\psi_{j+1} - \psi_j}{2}\right) n_-. \tag{4.13}$$

Now take a grid which is such that no gridpoint is placed inside the layer. Let x_I be the largest grid point "left of" the layer, then x_{I+1} is already "right of" the layer.

Standard theory implies that

$$n_I \approx n(x_I) = e^{\psi(x_I,\lambda) - \psi(-1,\lambda)} n_- \tag{4.14}$$

holds if $\bar{h}$ is sufficiently small. From (4.13) and (4.14) we derive

$$n_{I+1} = n_I \phi\left(\frac{\psi_{I+1} - \psi_I}{2}\right) \approx e^{\psi(x_I,\lambda) - \psi(-1,\lambda)} \phi\left(\frac{\psi_{I+1} - \psi_I}{2}\right) n_-. \tag{4.15}$$

Since $n(x_{I+1},\lambda) = e^{\psi(x_{I+1},\lambda) - \psi(-1,\lambda)} n_-$ holds, n_{I+1} only approximates $n(x_{I+1},\lambda)$ if $\phi(\psi_{I+1} - \psi_I/2)$ approximates $e^{\psi(x_{I+1},\lambda) - \psi(x_I,\lambda)}$. However $\phi(z)$ only approximates e^{2z} if z is small. Since the internal layer jump of ψ, i.e., $\psi(x_{I+1},\lambda) - \psi(x_I,\lambda)$, can be arbitrarily large, the trapezoidal rule generally does not resolve layer jumps, at least not for layer-ignoring grids. It can be shown that the trapezoidal rule converges if and only if the grid sizes inside the layer are small compared to λ. No better performance can be expected for the boundary value problem and for the two-dimensional problem. This result was anticipated by J. Barnes [4] who used arguments based on physical grounds. The superiority of the SG method for the initial value problem is due to the exact resolution of internal layers for arbitrary grids.

We now turn to the boundary value problem (4.4). To solve the recursion we observe that $J_{n_i} \equiv A$ holds. Then (4.4a) has to be solved for general A by using (4.10) and variation of constants. Then A and the second integration constant have to be determined from the boundary conditions (4.4c). We obtain

(a)
$$n_i = e^{\psi(x_i,\lambda) - \psi(-1,\lambda)} n_- + \sum_{j=0}^{i-1} J_{n_j} h_j e^{\psi(x_i,\lambda) - \psi(x_{j+1},\lambda)} w\left(\frac{\psi(x_{j+1},\lambda) - \psi(x_j,\lambda)}{2}\right)$$

(b)
$$J_{n_i} \equiv \frac{n_- e^{\psi(-1,\lambda) - \psi(+1,\lambda)} n_-}{\sum_{j=0}^{N-1} e^{\psi(1,\lambda) - \psi(x_{j+1},\lambda)} h_j w\left(\frac{\psi(x_{j+1},\lambda) - \psi(x_j,\lambda)}{2}\right)} \tag{4.16}$$

where

$$w(z) = e^z \frac{\sinh z}{z} \tag{4.17}$$

holds. The structure of the discrete solution is analogous to the structure of the continuous solution as given by (4.3), only the integrals are approximated by sums.

Standard analysis shows that

$$\left| \int_{x_I}^{x_J} e^{-\psi(s,\lambda)}\, ds - \sum_{j=I}^{J-1} e^{-\psi_{j+1}} h_j w\left(\frac{\psi_{j+1} - \psi_j}{2}\right) \right| \leq K\bar{h} \tag{4.18}$$

holds if the layer of ψ is not in $[x_I, x_J]$. Since the width of the layer is $0(\lambda|\ln\lambda|)$ we obtain the uniform convergence estimate

$$\max_{0 \leq i \leq N} \left(|n_i - n(x_i,\lambda)| + |J_n - J_{ni}|\right) \leq c(\bar{h} + \lambda|\ln\lambda|) \tag{4.19}$$

for an arbitrary mesh. If the mesh sizes outside the layer are constant, then $\bar{h}$ can be substituted by $\bar{h}^2$. The constant c only depends on upper bounds for n, ψ and on bounds of derivatives of n and ψ up to order three outside layers.

The estimate (4.19) carries over to the hole continuity equation when n is substituted by p.

Similar results hold for the problem with nonzero recombination rate and for the two-dimensional problem. We do

not present the proofs here since they require a large technical apparatus.

The estimate (4.19) implies that the error contribution of the discretization of the continuity equations only depends on the maximal grid sizes and on the singular perturbation parameter, but not on the mesh inside the layer.

Actually, this estimate can be refined, such that the right hand side of (4.19) is substituted by

$$c\Big(\max_{i,j}|l_{ij}|+\lambda|\ln\lambda|\Big) \tag{4.20}$$

where l_{ij} is the local discretization error of the (two-dimensional) SG scheme and the maximum is only taken outside the strip of width $0(\lambda|\ln\lambda|)$ around the junction. l_{ij} only depends on the "local" mesh sizes $h_i, h_{i-1}, k_j, k_{j-1}$ and on derivatives up to order four of n and ψ (the local discretization error is computed by inserting the exact solution into the SG scheme and by Taylor expansion (see next section)). Therefore the local discretization error outside the layer can be equidistributed (see next section) and the grid inside the layer may be chosen arbitrarily at least as far as the continuity equations are concerned. In the next section we will show that the grid inside the layer has to be constructed with respect to the discretization of Poisson's equation.

V. The Five-Point Formula

The scaled Boltzmann statistics

$$n=\gamma^2\lambda^2e^{\psi-\phi_n},\ p=\gamma^2\lambda^2e^{\phi_p-\psi} \tag{5.1}$$

transform Poisson's equation to

$$\lambda^2\Delta\psi=\gamma^2\lambda^2(e^{\psi-\phi_n}-e^{\phi_p-\psi})-D(x,y,\lambda),(x,y)\epsilon\Lambda. \tag{5.2}$$

The linearization of (5.2) with respect to ψ along some function w is given by

$$\lambda^2\Delta w=\gamma^2\lambda^2(e^{\psi-\phi_n}+e^{\phi_p-\psi})w-E(x,y,\lambda),(x,y)\epsilon\Lambda \tag{5.3}$$

where E has a layer at the same position as D. The function

$$\alpha=\gamma^2\lambda^2(e^{\psi-\phi_n}+e^{\phi_p-\psi}) \tag{5.4}$$

is positive. Since discretizations of (5.3) (including boundary data) have to be solved in every step of the (Newton) iteration of the five point formula for (5.2) we investigate the five-point formula applied to the linear singularly perturbed elliptic problem

$$\text{(a)}\quad \lambda^2\Delta w=\alpha(x,y,\lambda)w-E(x,y,\lambda),(x,y)\epsilon\Lambda \tag{5.5}$$

subject to mixed Neumann–Dirichlet boundary conditions on $\partial\Lambda$, where ξ denotes the outward unit normal to $\partial\Lambda$

$$\text{(b)}\quad w\Big|_{(\partial\Lambda)_c}=\hat{w},\ \frac{\partial w}{\partial\xi}\Big|_{(\partial\Lambda)_i}=0,(\partial\Lambda)_c\cup(\partial\Lambda)_i=\partial\Lambda. \tag{5.5}$$

For simplicity we assume that the junction is abrupt, i.e., the function E has a jump discontinuity along some curve $\Gamma\subseteq\Lambda$ and that E is independent of λ. Due to the built-in potential which behaves like $\pm\ln(1/\gamma^2\lambda^2)$ outside a p-n layer, the function α is uniformly in λ bounded away from zero outside the p-n layer. For simplicity we assume that α is uniformly bounded away from zero *everywhere*, that means

$$0<\underline{\alpha}\leqslant\alpha(x,y,\lambda)\leqslant\bar{\alpha},(x,y)\epsilon\Lambda\cup\partial\Lambda \tag{5.6}$$

with $\underline{\alpha},\bar{\alpha}$ independent of λ. This assumption does not influence the results, it only simplifies the technical approach needed for the analysis.

Because of (5.4) $\alpha(\cdot,\cdot,0)$ has a discontinuity along Γ. The five-point formula applied to (5.5a) reads

$$\lambda^2L_{ij}w_{ij}=\alpha_{ij}w_{ij}-E(x_i,y_j) \tag{5.7}$$

at interior grid points where $L_{ij}w_{ij}$ is the discrete Laplace operator

$$L_{ij}w_{ij}=\frac{2}{h_i+h_{i-1}}\left(\frac{w_{i+1,j}-w_{ij}}{h_i}-\frac{w_{ij}-w_{i-1,j}}{h_{i-1}}\right)+\frac{2}{k_j+k_{j-1}}\left(\frac{w_{i,j+1}-w_{ij}}{k_j}-\frac{w_{i,j}-w_{i,j-1}}{k_{j-1}}\right) \tag{5.8}$$

and $\alpha_{ij}=\alpha(x_i,y_j,\lambda)$.

For the following analysis we take a rectangular device, that means Λ is the interior of a rectangle in the x, y plane. Neumann boundary conditions can be discretized in the obvious way, i.e., $(w_{1j}-w_{0j})/h_0\approx w_x(x_0,y_j),(w_{Nj}-w_{N-1,j})/h_{N-1}\approx w_x(x_N,y_j)$ and analogously for w_y. This leads to a first-order approximation of the boundary conditions. A second-order approximation can be obtained by "mirror imaging," that is by introducing the exterior grid points $(x_{-1},y_j),(x_{N+1},y_j)$ for the approximation of $w_x(x_0,y_j),w_x(x_N,y_j)$ and analogously for w_y. After arranging the gridpoints row-wise into the grid vector $w_{h,k}$, the linear system representing (5.7) and the boundary conditions can be written as

$$L_{h,k}(\lambda)w_{h,k}=f_{h,k} \tag{5.9}$$

where $(f_{h,k})_{ij}=-E(x_i,y_j)$ at interior gridpoints and $(f_{h,k})_{ij}$ is equal to the boundary data at boundary gridpoints.

In the Appendix we prove that the system of difference equations (5.9) is uniformly stable in λ in the maximum norm for every grid, that means there is a constant L independent of the grid and of λ such that

$$\|L_{h,k}^{-1}(\lambda)\|\leqslant L \tag{5.10}$$

holds, where we denote with $\|\cdot\|$ the row sum norm of matrices as well as the maximum norm of vectors.

Stability is one ingredient for convergence, the second is consistency. Therefore, we insert the exact solution w, that is the solution of (5.5), into the difference scheme. Taylor expansion up to the third term gives the "local discretiza-

tion" error

$$l_{ij} = \lambda^2 L_{ij} w(x_i, y_j) - \alpha(x_i, y_j, \lambda) w(x_i, y_j) + E(x_i, y_j)$$
$$= \lambda^2 \left[\frac{\partial}{\partial x^3} w(\xi_{1i}, y_j) \frac{h_i^2}{3(h_i + h_{i-1})} - \frac{\partial}{\partial x^3} w(\xi_{2i}, y_j) \frac{h_{i-1}^2}{3(h_i + h_{i-1})} + \frac{\partial}{\partial y^3} w(x_i, \eta_{1j}) \frac{k_j^2}{3(k_j + k_{j-1})} - \frac{\partial}{\partial y^3} w(x_i, \eta_{2j}) \frac{k_{j-1}^2}{3(k_j + k_{j-1})} \right] \quad (5.11)$$

where $\xi_{1i}, \xi_{2i} \epsilon (x_{i-1}, x_i); \eta_{1j}, \eta_{2j} \epsilon (y_{j-1}, y_j)$. For (5.11) we assumed that w is three times continuously differentiable in Λ. The local error component from the first order discretization of the Neumann conditions at vertical boundaries is

$$l_{0j} = \frac{h_0}{2} \frac{\partial}{\partial x^2} w(\xi_0, y_j), \; l_{Nj} = \frac{h_{N-1}}{2} \frac{\partial}{\partial x^2} w(\xi_{N-1}, y_j) \quad (5.12)$$

where $\xi_0 \epsilon (x_0, x_1), \xi_{N-1} \epsilon (x_{N-1}, x_N)$. Similar expressions hold at horizontal boundaries. For general grids the five-point formula is of first order, that means

$$|l_{ij}| = 0(\bar{h} + \bar{k}) \quad (5.13)$$

holds depending on bounds of the second and third derivatives of w. It should be noted that by using one more term in the Taylor expansion the truncation error l_{ij} in (5.11) can be written as

$$l_{ij} = \lambda^2 \left[\frac{1}{3}(h_i - h_{i-1}) \frac{\partial^3 w}{\partial x^2}(\xi_{1i}, y_j) + \frac{h_i^3 + h_{i-1}^3}{12(h_i + h_{i-1}} \frac{\partial^4 w}{\partial x^4}(\xi_{2i}, y_j) + \frac{1}{3}(k_j - k_{j-1}) \frac{\partial^3 w}{\partial y^3}(x_i, \eta_{1j}) + \frac{k_j^3 + k_{j-1}^3}{12(k_j + k_{j-1})} \frac{\partial^4 w}{\partial y^4}(x_i, \eta_{2j}) \right] \quad (5.11a)$$

Equation (5.11a) implies, that the local error is of second order for equidistant grids ($h_1 = h_{i-1}$ and $k_j = k_{j-1}$). If the grid is not equidistant the truncation error is outside the layer still of order $(\bar{h}^2 + \bar{k}^2)$ as long as $|h_i - h_{i-1}| < \text{const}\, \bar{h}^2$ and $|k_j - k_{j-1}| < \text{const}\, \bar{k}^2$ holds. However this second-order estimate involves the fourth derivative of the exact solution. Thus as we will see, it represents no improvement over (5.11) inside the layer regions.

We denote by w_{ex} the vector with the grid values of the exact solution w (row-wise ordered) and get

$$L_{h,k}(\lambda)(w_{ex} - w_{h,k}) = l_{h,k} \quad (5.14)$$

where $l_{h,k}$ is the vector with components l_{ij}. The stability estimate (5.10) gives the global error estimate

$$\|w_{ex} - w_{hk}\| \leq L \|l_{h,k}\|. \quad (5.15)$$

Therefore, a prescribed global error tolerance δ can be achieved if we require that

$$|l_{ij}| \leq \frac{\delta}{L} = \bar{\delta} \quad (5.16)$$

holds for all i, j (L is given explicitly in the Appendix).

Our goal is now to determine grid sizes h_i, k_j such that the local discretization error as given by (5.11), (5.12) is (approximately) equal to $\bar{\delta}$. This process is called equidistribution of the local error (see Markowich and Ringhofer [11]). The obtained equidistributing grid will be fine inside layer regions (where the derivatives vary rapidly) and it will be coarse outside layers (where the derivatives are 0(1)).

Equidistribution is normally done iteratively. That means solutions of the discretization scheme are computed on an initial mesh. Then the derivatives in the local error are approximated for this initial solution and a new grid is determined by equidistribution. A second approximation to the solution is computed on this new grid and so on. The iteration is stopped by an appropriate criterion, e.g., when two consecutive solutions differ insignificantly or when a solution satisfies the local error condition (5.16) on the mesh it was computed. For the nonlinear problem the old solution is used as initial guess for Newton's method for the computation of the new solution (continuation method, see Ascher, Christiansen, and Russell [1]).

We will now show that the singular perturbation analysis can be used to obtain qualitative and quantitative information on the equidistributing grid inside layers.

Analogously to Section II we derive that the solution of (5.5) has the form

$$w(x, y, \lambda) = \bar{w}(x, y) + \hat{w}\left(\frac{t(x, y)}{\lambda}, s(x, y) \right) + 0(\lambda) \quad (5.17)$$

where t, s are the local coordinates introduced in Section II. The function $\bar{w}$ is the solution of the reduced problem (5.5)

$$\bar{w}(x, y) = \frac{E(x, y)}{\alpha(x, y, 0)} \quad (5.18)$$

which is discontinuous along Γ and $\hat{w}(t/\lambda, s)$ is the internal-layer term which solves the layer equation

$$\begin{aligned} &\text{(a)} \quad \hat{w}_{\tau\tau} = \alpha_+(s)\hat{w}, \qquad \tau > 0 \\ &\text{(b)} \quad \hat{w}_{\tau\tau} = \alpha_-(s)\hat{w}, \qquad \tau < 0 \\ &\text{(c)} \quad \hat{w}(0+, s) - \hat{w}(0-, s) = \bar{w}_-(s) - \bar{w}_+(s) \\ &\text{(d)} \quad \hat{w}_\tau(0+, s) = \hat{w}_\tau(0-, s) \end{aligned} \quad (5.19)$$

where $\alpha_+(s), \alpha_-(s)$ and $\bar{w}_+(s), \bar{w}_-(s)$ are equal to α and $\bar{w}$, respectively, evaluated at the "right" and "left" side of the junction Γ, respectively and $\tau = t/\lambda$. The layer solution $\hat{w}$ then is

$$\text{(e)} \quad \hat{w}(\tau, s) = \begin{cases} \gamma_+(s) e^{-\sqrt{\alpha_+(s)}\tau}, \tau > 0 \\ \gamma_-(s) e^{\sqrt{\alpha_-(s)}\tau}, \tau < 0 \end{cases} \quad (5.19)$$

where

$$\text{(f)}\quad \gamma_+(s)=\frac{\bar{w}_-(s)-\bar{w}_+(s)}{1+\sqrt{\dfrac{\alpha+(s)}{\alpha-(s)}}}$$

$$\text{(g)}\quad \gamma_-(s)=\frac{\bar{w}_+(s)-\bar{w}_-(s)}{1+\sqrt{\dfrac{\alpha+(s)}{\alpha-(s)}}}. \tag{5.19}$$

The asymptotic expansion (5.17) holds in any neighborhood of the junction Γ in which no other layer occurs. We remark that layers can only be due to rapid variation of E and to boundary conditions which are not fulfilled asymptotically by $\bar{w}$ (boundary layers). Differentiation as in (2.41) gives

$$\text{(a)}\quad \frac{\partial}{\partial x^i}w(x,y,\lambda)=\frac{1}{\lambda^i}\frac{\partial^i}{\partial\tau^i}\hat{w}\left(\frac{t(x,y)}{\lambda},s(x,y)\right)t_x+0\left(\frac{1}{\lambda^{i-1}}\right)$$

$$\text{(b)}\quad \frac{\partial}{\partial y^i}w(x,y,\lambda)=\frac{1}{\lambda^i}\frac{\partial^i}{\partial\tau^i}\hat{w}\left(\frac{t(x,y)}{\lambda},s(x,y)\right)t_y+0\left(\frac{1}{\lambda^{i-1}}\right) \tag{5.20}$$

inside the layer. Therefore the local discretization error (5.11) fulfills

$$|l_{ij}|\leqslant \text{const}\left[\left(\frac{h_i+h_{i-1}}{\lambda}|t_x(x_i,y_j)|+\frac{k_{i-1}+k_j}{\lambda}|t_y(x_i,y_j)|\right)\cdot\exp\left(-\frac{\sqrt{\underline{\alpha}}}{\lambda}|t(x_i,y_j)|\right)+h_{i-1}+h_i+k_{j-1}+k_j\right] \tag{5.21}$$

inside the layer, that is for $|t(x,y)|\leqslant\lambda/\sqrt{\underline{\alpha}}\,|\ln\lambda|$.

Let us take the vertical junction $x=0$ at first, such that $t=x$, $s=(0,y)$. Equation (5.21) now reads

$$|l_{ij}|\leqslant \text{const}\left[\frac{h_i+h_{i-1}}{\lambda}\exp\left(-\frac{\sqrt{\underline{\alpha}}}{\lambda}|x_i|\right)+h_{i-1}+h_i+k_{j-1}+k_j\right] \tag{5.22}$$

In order to satisfy the equidistribution condition (5.16) we choose

$$\text{(a)}\quad h_i=c_1\lambda\delta\exp\left(\frac{\sqrt{\underline{\alpha}}}{\lambda}|x_i|\right)$$

$$\text{(b)}\quad k_j=c_2\delta \tag{5.23}$$

for $|x_i|\leqslant(1/\sqrt{\underline{\alpha}})\lambda|\ln\lambda|$ where c_1,c_2 are 0(1) constants. Markowich and Ringhofer [11] showed that the number of grid points in x-direction on every gridline is

$$N^x_{\text{layer}}=0\left(\frac{1}{\delta}\right) \tag{5.24}$$

inside the layer. Due to the exponential grading of h_i the number of gridpoints is independent of λ. The number of y-gridlines is

$$N^y_{\text{layer}}=0\left(\frac{1}{\delta}\right). \tag{5.25}$$

Therefore, the equidistributing grid requires

$$N_{\text{layer}}=N^x_{\text{layer}}N^y_{\text{layer}}=0\left(\frac{1}{\delta^2}\right) \tag{5.26}$$

meshpoints within the layer in order to achieve a $0(\delta)$ global error. The smallest gridsize h_i within the layer is $0(\lambda\delta)$ and the largest is $0(\delta)$. All mesh sizes in y-direction can be chosen as $0(\delta)$.

Equidistribution gives a comparable number of gridpoints for horizontal junctions and for horizontal and vertical boundary layers.

We now assume that the junction Γ is not aligned to the coordinate system. As example we take the line

$$\Gamma: y=x \tag{5.27}$$

such that $t(x,y)=(\sqrt{2}/2)(x-y)$ holds. Equations (5.16) and (5.21) give bounds for the mesh sizes

$$\text{(a)}\quad h_i\leqslant c_1\lambda\delta\exp\left(\frac{\sqrt{\underline{\alpha}}}{\lambda}\frac{\sqrt{2}}{2}|x_i-y_j|\right)$$

$$\text{(b)}\quad k_j\leqslant c_2\lambda\delta\exp\left(\frac{\sqrt{\underline{\alpha}}}{\lambda}\frac{\sqrt{2}}{2}|x_i-y_j|\right) \tag{5.28}$$

for $|x_i-y_j|\leqslant(2/\sqrt{2\underline{\alpha}})\lambda|\ln\lambda|$. The number of meshpoints inside the layer is at least

$$N_{\text{layer}}=0\left(\frac{1}{\lambda\delta^2}\right). \tag{5.29}$$

Since the x- *and* y-derivatives inside the layer are large the mesh sizes in x- *and* y-direction depend on λ and therefore the number of meshpoints inside the layer increases for constant δ as λ decreases. Even assuming that gridlines terminate outside layer regions, this leads to storage requirements which virtually cannot be met. Normally $\lambda\leqslant 10^{-3}$, then an accuracy bound $\delta=10^{-2}$ implies that (in the order of magnitude) 10^7 gridpoints have to be placed inside the layer. Including the continuity equations linear systems of approximate dimension 3×10^7 would have to be solved in every Newton step.

Therefore equidistribution has to be relaxed and we have to check whether larger mesh sizes also lead to acceptable discrete solutions. To do so we split the matrix $L_{h,k}(\lambda)$ into

$$L_{h,k}(\lambda)=\lambda^2L^{(1)}_{h,k}+L^{(2)}_{h,k} \tag{5.30}$$

where $L^{(1)}_{h,k}$ represents the discrete Laplace operator (5.8) (the rows corresponding to exterior gridpoints have only zero entries) and $L^{(2)}_{h,k}$ has diagonal entries α_{ij} at rows corresponding to interior gridpoints, it has diagonal entries 1 at rows corresponding to Dirichlet boundary points and the rows corresponding to Neumann boundary points have two nonvanishing entries representing the discrete boundary condition. Obviously $L^{(2)}_{h,k}$ is nonsingular and

$$\|L^{(2)-1}_{h,k}\|\leqslant\max\left(1,\frac{1}{\underline{\alpha}}+\bar{h}+\bar{k}\right) \tag{5.31}$$

holds. Also we obtain

$$\|L^{(1)}_{h,k}\| \leqslant 4\left(\frac{1}{\underline{h}^2}+\frac{1}{\underline{k}^2}\right) \tag{5.32}$$

where $\underline{h}=\min h_i, \underline{k}=\min k_j$. Assume that

$$\left(\frac{\lambda}{\underline{h}}\right)^2+\left(\frac{\lambda}{\underline{k}}\right)^2 < \frac{\max\left(1,\frac{1}{\alpha}+\bar{h}+\bar{k}\right)}{4} \tag{5.33}$$

holds. Then the solution of (5.9) fulfills

$$w_{h,k}=L^{(2)-1}_{h,k}f_{h,k}+0\left(\left(\frac{\lambda}{\underline{h}}\right)^2+\left(\frac{\lambda}{\underline{k}}\right)^2\right) \tag{5.34}$$

or component wise

$$w_{ij}=\bar{w}(x_i,y_j)+0(h_0+h_N+k_0+k_M)+0\left(\left(\frac{\lambda}{\underline{h}}\right)^2+\left(\frac{\lambda}{\underline{k}}\right)^2\right) \tag{5.35}$$

where $\bar{w}$ is the reduced solution of (5.5), given by (5.18). If $\lambda/\underline{h}, \lambda/\underline{k}$ are sufficiently small, the discrete solution approximates the reduced solution. Since $\lambda < h_i, \lambda < k_j$ for all i, j is required the mesh is coarse inside the layer, however discretization errors which occur inside layer regions do not spread out. Of course layers are not resolved at all by this mesh. By choosing $\bar{h}=0(\delta), \bar{k}=0(\delta)$ or by equidistributing the local error of the continuity equation outside layer regions with error bound δ the reduced solution of the semiconductor equations can be resolved accurately for small λ on such a grid. The bound for the global error then is $0(\delta)+0(\lambda)+0((\lambda/\underline{h})^2+(\lambda/\underline{k})^2)$.

Therefore even a constant coarse grid $h_i \equiv k_j \equiv h \gg \lambda$ can be chosen as initial grid for the iteration procedure and the computed initial solution will be close to the reduced solution.

We now investigate the behavior of the discrete solutions when the condition $\lambda/\underline{h}, \lambda/\underline{k}$ small is neglected.

It suffices to investigate the one-dimensional constant coefficient version of (5.5)

$$\begin{aligned} &\text{(a)}\quad \lambda^2 w'' = \alpha w + E(x), \quad -1 \leqslant x \leqslant 1 \\ &\text{(b)}\quad w(-1)=\bar{w}(-1), w(1)=\bar{w}(1) \end{aligned} \tag{5.36}$$

where E has a jump discontinuity at $x=0$ and $\alpha > 0$. $\bar{w} = -E/\alpha$ is the reduced solution. Therefore no boundary layers occur, only an internal layer at $x=0$.

The one-dimensional finite difference analogue to (5.7), (5.8) on the constant grid with $h_i \equiv h = 1/N$ is

$$\begin{aligned} &\text{(a)}\quad \phi^2(w_{i+1}-2w_i+w_{i-1}) = w_i - E(x_i) \\ &\text{(b)}\quad w_0=\bar{w}(-1), w_N=\bar{w}(1) \end{aligned} \tag{5.37}$$

with

$$\phi = \frac{\lambda}{h}. \tag{5.38}$$

We regard (5.37) as two-step recursion. The characteristic equation is

$$\phi^2(r^2-2r+1)-\alpha r = 0. \tag{5.39}$$

The roots of (5.39) are

$$\begin{aligned} &\text{(a)}\quad r_+(\phi)=\frac{2\phi^2+\alpha+\sqrt{4\alpha\phi^2+\alpha^2}}{2\phi^2} > 1 \\ &\text{(b)}\quad r_-(\phi)=\frac{2\phi^2+\alpha-\sqrt{4\alpha\phi^2+\alpha^2}}{2\phi^2} < 1. \end{aligned} \tag{5.40}$$

An easy calculation gives

$$\begin{aligned} &\text{(a)}\quad \lim_{\phi\to 0+} r_+(\phi) = +\infty, \\ &\qquad \lim_{\phi\to\infty} r_+(\phi)=1, r_+(\phi)\downarrow \text{ monotonically as } \phi\to\infty \\ &\text{(b)}\quad \lim_{\phi\to 0+} r_-(\phi)=0, \\ &\qquad \lim_{\phi\to\infty} r_+(\phi)=1, r_-(\phi)\uparrow \text{ monotonically as } \phi\to\infty. \end{aligned} \tag{5.41}$$

The difference scheme (5.37) can be solved explicitly

$$w_i = r_+^{i-N}a(N)+r_-^i b(N)-\frac{1}{\sqrt{4\alpha\phi^2+\alpha^2}} \cdot\left(\sum_{j=i}^{N-1} r_+^{i-j}E(x_j)+\sum_{j=0}^{i-1} r_-^{i-j}E(x_j)\right) \tag{5.42}$$

where

$$a(N)=\frac{\hat{w}_N - r_-^N\hat{w}_0}{1-\left(\frac{r_-}{r_+}\right)^N}, \quad b(N)=\frac{\hat{w}_0-\hat{w}_N r_+^{-N}}{1-\left(\frac{r_-}{r_+}\right)^N} \tag{5.43}$$

with

$$\begin{aligned} &\text{(a)}\quad \hat{w}_0=\bar{w}(-1)+\frac{1}{\sqrt{4\alpha\phi^2+\alpha^2}}\sum_{j=0}^{N-1} r_+^{-j}E(x_j) \\ &\text{(b)}\quad \hat{w}_N=\bar{w}(1)+\frac{1}{\sqrt{4\alpha\phi^2+\alpha^2}}\sum_{j=0}^{N-1} r_-^{N-j}E(x_j) \end{aligned} \tag{5.44}$$

holds. We take $N=2L$ such that $x_L=0$ holds. For simplicity we assume

$$\text{(a)}\quad E(x)=\begin{cases} -1, & x<0 \\ 1, & x\geqslant 0 \end{cases} \tag{5.45}$$

Then, using (5.40), (5.41) we obtain the asymptotics

$$\text{(b)}\quad w_i=\begin{cases} \dfrac{1}{\alpha}-\dfrac{2}{(r_+-1)\sqrt{4\alpha\phi^2+\alpha^2}}r_+^{i-L} \\ \quad +0(r_+^{-L})+0(r_-^L), & i<L \\ -\dfrac{1}{\alpha}+\dfrac{2}{(1-r_-)\sqrt{4\alpha\phi^2+\alpha^2}}r_-^{i-L} \\ \quad +0(r_+^{-L})+0(r_-^L), & i\geqslant L \end{cases} \tag{5.45}$$

The first term on the right-hand side of (5.45) is the

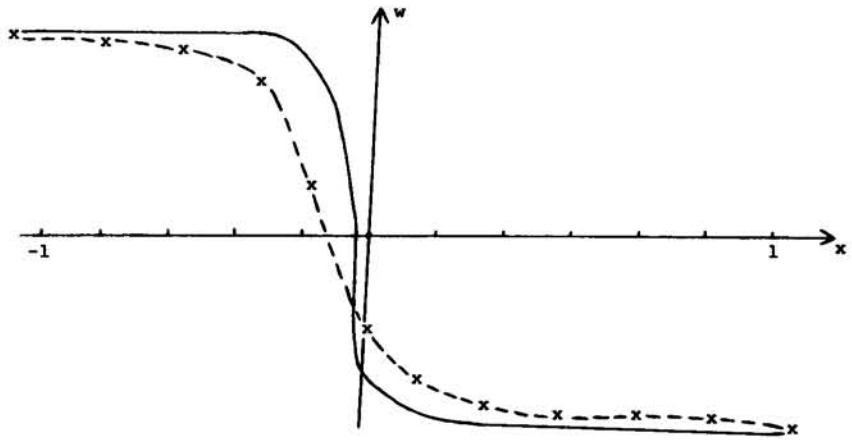

Fig. 3. Exact and discrete solution of the one-dimensional linearized Poisson equation.

reduced solution

$$\bar{w}(x) = \begin{cases} \frac{1}{\alpha}, & x < 0 \\ -\frac{1}{\alpha}, & x > 0 \end{cases}$$

and the second is the discrete internal layer term which decays away from $x = 0$ to both sides.

Therefore, the structure of the discrete solution is *analogous* to the structure of the continuous solution no matter what value $\phi = \lambda/h$ takes. However the "discrete internal layer terms" are not close to the "continuous internal layer terms" as given by (5.19) unless λ/h is small or large. Therefore the five-point scheme is *not* uniformly (i.e., independently of λ) convergent since the choice $h \approx \lambda$ creates a large global error. The behavior of the continuous and discrete solution for $h \approx \lambda$ is illustrated by Fig. 3.

The discrete solution displays internal layer structure. Large errors only occur within the internal layer. The layer terms decay exponentially away from the layer. Therefore the solution may be regarded as acceptable.

The situation for the two-dimensional problem is completely analogous.

The following grid construction strategy can be employed:

(a) equidistribution outside layers
(b) "limited" equidistribution within layers ("limited" means that not more than an *a priori* defined number of gridpoints depending on storage restrictions is allowed within each layer).

For the automatization of this strategy it is necessary to determine for each gridpoint whether it lies inside or outside the layer. This can be done by checking the norm of the gradient of E at that point. A discrete solution of the five-point scheme, which qualitatively agrees with the solution of the continuous problem and which agrees quantitatively with the continuous solution outside layers is obtained using the strategy (a), (b).

This however requires that gridlines are allowed to terminate outside layers. We now discuss the situation

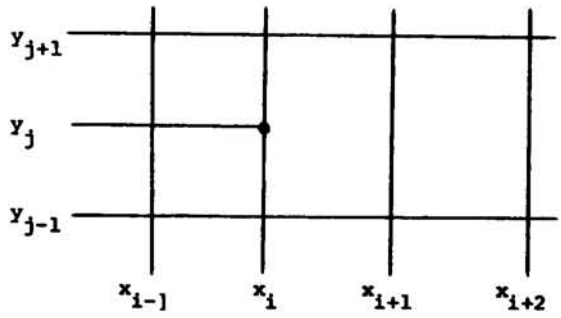

Fig. 4. A terminating line.

depicted in Fig. 4. The y_j-gridline terminates at x_i, the point (x_{i+1}, y_j) is no grid point.

We approximate the "missing" x-difference quotient by linear interpolation between the $(j+1)$st and $(j-1)$st y level

$$\frac{w_{i+1,j} - w_{ij}}{h_i} \approx \frac{k_{j-1}}{k_j + k_{j-1}} \frac{w_{i+1,j+1} - w_{i,j+1}}{h_i} + \frac{k_j}{k_j + k_{j-1}} \frac{w_{i+1,j-1} - w_{i,j-1}}{h_i}. \quad (5.46)$$

The right-hand side of (5.46) is then substituted into $L_{ij}w_{ij}$ as given (5.8). The local error l_{ij}^{II} introduced by (5.46) fulfills

$$|l_{ij}^{II}| \leq \frac{1}{h_i(h_i + h_{i-1})} \left(\frac{h_i^3}{3} + \frac{h_i^2}{2}(k_j + k_{j-1}) + \frac{h_i}{2}(k_j^2 + k_{j-1}^2) \right) \cdot \|D_3^x w\|_{[x_{i-1}, x_{i+1}] \times [y_{j-1}, y_{j+1}]} \quad (5.47)$$

where

$$D_3^x w = |w_{xxx}| + |w_{xxy}| + |w_{xyy}|. \quad (5.48)$$

Missing y-derivatives are approximated analogously. The local error contribution then is obtained by interchanging h_i, k_j in (5.47) and $D_3^y u$ is obtained from $D_3^x u$ by interchanging x and y.

Equations (5.47) implies that grid-size ratio restrictions have to be assumed in order to get consistency.

For a terminating y-gridline we require

$$\frac{k_j}{h_i} \leqslant c, \frac{k_{j-1}}{k_j} \leqslant c \tag{5.49}$$

and for a terminating x-gridline

$$\frac{h_i}{k_j} \leqslant c, \frac{h_{i-1}}{h_i} \leqslant c \tag{5.50}$$

where c is a moderate constant. Then (5.47) simplifies to

$$|l_{ij}^{(l)}| \leqslant \left(\frac{h_i}{3} + \frac{k_j}{2} + \frac{k_{j-1}}{2}\right) + \frac{c}{2}(1+c^2)k_j\Big) \cdot \|D_3^x u\|_{[x_{i-1}, x_{i+1}]\times[y_{i-1}, y_{i+1}]} \tag{5.51}$$

and analogously for a terminating x-gridline. The local error introduced by a terminating line is of first order. Similarly the SG scheme for the continuity equations is changed at gridpoints where a line terminates.

Generally, lines are only allowed to terminate outside layer regions, since there $D_3^x u$, $D_3^y u$ have moderate values. The decision if and where a x- or y-gridline terminates is explained in Franz *et al.* [7].

At last we show how the five point scheme can be adjusted such that the resulting difference scheme is uniformly convergent. We modify the five point scheme as follows:

$$\lambda^2\left[\frac{2\sigma_{ij}^x}{h_i + h_{i-1}}\left(\frac{w_{i+1,j} - w_{ij}}{h_i} - \frac{w_{ij} - w_{i-1,j}}{h_{i-1}}\right) + \frac{2\sigma_{ij}^y}{k_j + k_{j-1}}\left(\frac{w_{i,j+1} - w_{ij}}{k_j} - \frac{w_{ij} - w_{i,j-1}}{k_{j-1}}\right)\right] = \alpha_{ij} w_{ij} - E(x_i, y_j) \tag{5.52}$$

where $\sigma_{ij}^x, \sigma_{ij}^y$ will be determined by exponential fitting (see Doolan, Miller and Schilders [5]).

We set

$$h_i = \rho_i^x \lambda, k_j = \rho_j^y \lambda \tag{5.53}$$

and since our previous analysis showed that only mesh sizes which are of the order of magnitude of λ can destroy convergence, we assume that ρ_i^x, ρ_j^y are independent of λ (later on we will get rid of this restriction).

In order to derive a necessary criterion for $\sigma_{ij}^x, \sigma_{ij}^y$ to make the scheme (5.52) uniformly convergent we assume

$$w_{ij} = w(x_i, y_j) + o(1) \text{ as } \lambda \to 0 \tag{5.54}$$

and compute $\sigma_{ij}^x, \sigma_{ij}^y$ from this assumption. Inserting the asymptotic expansion (5.17) into (5.52) and taking the limit for $\lambda \to 0$ gives

$$\lim_{\lambda \to 0}\left[\frac{2\sigma_{ij}}{\rho_i^x + \rho_{i1}^x}\left(\frac{\hat{w}(\tau_{i+1,j}, s_{ij}) - \hat{w}(\tau_{ij}, s_{ij})}{\rho_i^x} - \frac{\hat{w}(\tau_{ij}, s_{ij}) - \hat{w}(\tau_{i-1,j}, s_{ij})}{\rho_{i-1}^x}\right) + \frac{2\sigma_{ij}^y}{\rho_j^y + \rho_{j-1}^y}\left(\frac{\hat{w}(\tau_{i,j+1}, s_{ij}) - \hat{w}(\tau_{ij}, s_{ij})}{\rho_j^y} - \frac{\hat{w}(\tau_{ij}, s_{ij}) - w(\tau_{i,j-1}, s_{ij})}{\rho_{j-1}^y}\right)\right] = \lim_{\lambda \to 0} \alpha_{ij} \hat{w}(\tau_{ij}, s_{ij}) \tag{5.55}$$

where $\tau_{ij} = t(x_i, y_j)/\lambda, s_{ij} = s(x_i, y_j)$ holds. We now assume that the five-point star $\{(x_{i-1}, y_j), (x_i, y_j), (x_{i+1}, y_j), (x_i, y_{j-1}), (x_i, y_{j+1})\}$ is entirely on that side of the junction for which $t > 0$ holds.

Since

$$\tau_{i+1,j} = \tau(x_{i+1}, y_j) = \frac{t(x_i, y_j)}{\lambda} + \rho_i^x t_x(\xi_i, y_j), \xi_i \epsilon (x_i, x_{i+1}) \tag{5.56}$$

and similar expressions for $\tau_{i-1,j}, \tau_{i,j+1}, \tau_{i,j-1}$ hold, we set using (5.19e)

$$\frac{2\sigma_{ij}^x}{\rho_i^x + \rho_{i-1}^x}\left(\frac{e^{-\sqrt{\alpha_{ij}}\rho_i^x t_x(x_i, y_j)} - 1}{\rho_i^x} - \frac{1 - e^{\sqrt{\alpha_{ij}}\rho_{i-1}^x t_x(x_i, y_j)}}{\rho_{i-1}^x}\right) + \frac{2\sigma_{ij}^y}{\rho_j^y + \rho_{j-1}^y}\left(\frac{e^{-\sqrt{\alpha_{ij}}\rho_j^y t_y(x_i, y_j)} - 1}{\rho_j^y} - \frac{1 - e^{\sqrt{\alpha_{ij}}\rho_{j-1}^y t_y(x_i, y_j)}}{\rho_{j-1}^y}\right) = \alpha_{ij}. \tag{5.57}$$

For some $0 < c_{ij}^2 < 1$ we set

$$\text{(a)} \quad \sigma_{ij}^x = \frac{(\rho_i^x + \rho_{i-1}^x)c_{ij}^2 \alpha_{ij}}{2\left(\frac{\exp\left(-\sqrt{\alpha_{ij}}\rho_i^x t_x(x_i, y_j)\right) - 1}{\rho_i^x} - \frac{1 - \exp\left(\sqrt{\alpha_{ij}}\rho_{i-1}^x t_x(x_i, y_j)\right)}{\rho_{i-1}^x}\right)}$$

$$\text{(b)} \quad \sigma_{ij}^y = \frac{(\rho_j^y + \rho_{j-1}^y)(1 - c_{ij}^2)\alpha_{ij}}{2\left(\frac{\exp\left(-\sqrt{\alpha_{ij}}\rho_j^y t_y(x_i, y_j)\right) - 1}{\rho_j^y} - \frac{1 - \exp\left(\sqrt{\alpha_{ij}}\rho_{j-1}^y t_y(x_i, y_j)\right)}{\rho_{j-1}^y}\right)}. \tag{5.58}$$

Consistency requires that $\lim_{\rho_i^x, \rho_{i-1}^x \to 0} \sigma_{ij}^x = \lim_{\rho_j^y, \rho_{j-1}^y \to 0} \sigma_{ij}^y = 1$ and therefore we obtain by Taylor expansion

$$c_{ij}^2 = t_x^2(x_i, y_j), 1 - c_{ij}^2 = t_y^2(x_i, y_j) \tag{5.59}$$

since $t_x^2 + t_y^2 = 1$ holds.

Equation (5.58) holds for all i, j for which the five-point star centered at (x_i, y_j) is on that side of the junction for which $t > 0$ holds. On the other side $\sqrt{\alpha_{ij}}$ has to be substituted by $-\sqrt{\alpha_{ij}}$. In order to understand the influence of $\sigma_{ij}^x, \sigma_{ij}^y$ we take a constant grid $h_i \equiv h, k_j \equiv k$ such that $\rho_i^x \equiv \rho^x, \rho_j^y \equiv \rho^y$ holds. Equation (5.58) gives for $t < 0$ *and* $t > 0$

$$\sigma_{ij}^x = \sigma(u_{ij}), \sigma_{ij}^y = \sigma(v_{ij}) \tag{5.60}$$

where

$$\sigma(u) = \frac{u^2}{\sinh_u^2}, u_{ij} = \frac{\sqrt{\alpha(x_i, y_j)}\, t_x(x_i, y_j) h}{2\lambda}$$

and

$$v_{ij} = \frac{\sqrt{\alpha(x_i, y_j)}\, t_y(x_i, y_j) k}{2\lambda}.$$

We easily get

$$\sigma(u) = 1 + 0(u^2), u \to 0, \sigma(u) = 4u^2 e^{-2|u|} \text{ as } |u| \to \infty \tag{5.61}$$

Therefore the scheme (5.52) behaves (asymptotically) like the midpoint rule if $h \ll \lambda, k \ll \lambda$. If $h \gg \lambda, k \gg \lambda$ the scheme behaves like

$$\alpha_{ij} w_{ij} = E(x_i, x_j) + 0(e^{-e_{ij}h/\lambda} + e^{-f_{ij}k/\lambda}) \tag{5.62}$$

where $e_{ij} \geqslant 0, f_{ij} \geqslant 0, e_{ij} + f_{ij} > 0$. The reduced solution $\bar{w}$ is resolved much more accurately by (5.60) than by the standard five-point scheme.

These results hold locally for the variable mesh size case (5.58).

The modified scheme (5.52) is the five-point formula analogue to the SG method.

A problem which still has to be resolved is the modification of the five-point scheme at stars which are located at both sides of the junction (see Fig. 5).

In fact, it is for the abrupt junction formally not correct to write down the five-point scheme at (x_i, y_j) as depicted in Fig. 5 since at least one second-order partial derivative of the solution is not continuous at the junction. The correct approach is to substitute the five point formula at such an "interface" point by a discretized version of the interface condition

$$\text{grad}(w)\left(t_y^{tx}\right)|_{t=0+} = \text{grad}\, w\left(t_y^{tx}\right)|_{t=0-}. \tag{5.63}$$

at (x_i, y_j).

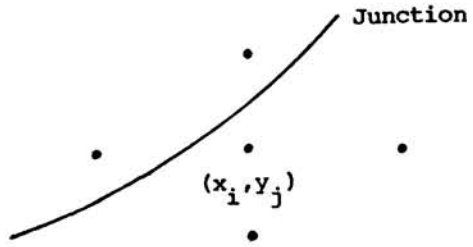

Fig. 5. A "crossing" star.

For the case of the vertical junction $x = 0$ exponential fitting gives the discrete interface condition

$$\frac{w_{L+1,j} - w_{L,j}}{1 - \exp\left(\frac{\sqrt{\alpha(0+, y_j)}}{\lambda} h_L\right)} = \frac{w_{L,j} - w_{L-1,j}}{1 - \exp\left(\frac{\sqrt{\alpha(0-, y_j)}}{\lambda} h_{L-1}\right)} \tag{5.64}$$

assuming that $x_L = 0$.

Then it can be shown that the modified midpoint rule is uniformly convergent and that the global error is $0(\bar{h}) + 0(\bar{k}) + 0(\lambda)$ for each grid (see Doolan, Miller and Schilders [5] for one-dimensional problem).

The disadvantage of the modified scheme (5.52) is that the junction Γ has to be known explicitly since $\sigma_{ij}^x, \sigma_{ij}^y$ depend on t_x, t_y. For most practical applications however the doping profile is graded and only given at discrete points. Then, since $\begin{pmatrix} t_x \\ t_y \end{pmatrix}$ is parallel to the direction of steepest descent (or growth) of the scaled doping profile $D(x, y, \lambda)$, the approximation

$$\begin{pmatrix} t_x \\ t_y \end{pmatrix} \approx \frac{\text{grad}\, D}{\|\text{grad}\, D\|} \tag{5.65}$$

can be used for the (numerical determination of the fitting factors (5.58).

VI. Conclusion

Summarizing the results obtained by the singular perturbation approach it can be stated that it generally provides an analysis of the qualitative behavior of the solution of the semiconductor equations. Its application to finite difference methods results in a convergence analysis of the Scharfetter–Gummel scheme which will be extended to the two-dimensional case in a succeeding paper. Using the asymptotic analysis of the exact solution we obtained a mesh selection criterion, which guarantees a good approximation of the solution away from the p-n junction with a reasonable amount of gridpoints whereas it was shown that the usual approach (the equidistribution of the "full" truncation error) does not lead to a suitable mesh. For the practical implementation of this mesh selection strategy see Franz *et al.* [7]. Finally a uniformly convergent finite-difference scheme for Poisson's equation was derived which—unlike the usual five point discretization—resolves the solution inside the p-n junction layer even if few gridpoints are located there. Finally it should be pointed

out that analytical results of Section II immediately carry over to the three-dimensional case.

APPENDIX

Assume that Λ is the interior of a rectangle. Let b_{ij} denote the (i, j) entry of the matrix $B_{h,k}(\lambda) = DA_{h,k}(\lambda)$ where D is the diagonal matrix which has -1 at every entry representing an interior gridpoint and 1 at every entry representing a boundary point. Then

(a) $b_{mn} \leqslant 0, \quad m \neq n$

(b) $$\sum_n b_{mn} = \begin{cases} 1, & \text{if the } m\text{th row stands for a Dirichlet boundary point} \\ \alpha_{ij}, & \text{if the } m\text{th row stands for the interior point } (x_i, y_j) \\ 0, & \text{if the } m\text{th row stands for a Neumann boundary point.} \end{cases}$$

(c) $B_{h,k}(\lambda)$ is irreducible.

We take the mesh function

$$\phi_{ij} = \frac{\alpha + x_i + y_j}{\beta}$$

where α, β are chosen such that $\phi_{ij} \geqslant c > 0$ and $\max_{i,j}|\phi_{ij}| = 1$ where i, j runs through all interior and boundary points. α, β only depend on Λ. Let $\phi_{h,k}$ denote the vector with components ϕ_{ij} (organized row wise). Then

$$(B_{h,k}(\lambda)\phi_{h,k})_{ij} = \begin{cases} \alpha_{ij}\phi_{ij}, & \text{if } (x_i, y_j) \text{ is an interior point} \\ \phi_{ij}, & \text{if } (x_i, y_j) \text{ is a Dirichlet boundary point} \\ \dfrac{1}{\beta}, & \text{if } (x_i, y_j) \text{ is a Neumann boundary point.} \end{cases}$$

From Doolan, Miller and Schilders [5], we derive that

$$\|B_{h,k}^{-1}(\lambda)\| \leqslant \frac{1}{\min_{i,j}(B_{h,k}(\lambda)\phi_{h,k})_{ij}} = \frac{1}{\min_{i,j}\left(\frac{1}{\beta}, \phi_{ij}\alpha_{ij}, \phi_{ij}\right)} \tag{A1}$$

holds.

The right-hand side of this estimate is independent of λ and of the grid and can be taken as the stability constant L since the inverse of D has norm one.

Assume here that the scaling of the independent variables is such that the rectangle $\partial\Lambda$ has the corner points $(-l^x, -l^y), (l^x, -l^y), (-l^x, l^y)$ and (l^x, l^y). Then a simple calculation shows that the constants α, β can be chosen such that

$$L = 2(l^x + l^y) + \frac{1}{\min(1, \underline{\alpha})}$$

holds. No better stability constant can be obtained from (A1). An analogous proof also holds for nonrectangular domains.

REFERENCES

[1] U. Ascher, J. Christiansen, and R. D. Russell, "A collocation solver for mixed order systems of boundary value problems," *Math. of Comp.*, vol. 33, pp. 659–679, 1979.

[2] U. Ascher and R. Weiss, "Singular perturbation problems I: First order systems with constant coefficients," *SIAM J. Numer. Anal.*

[3] ——, "Singular perturbation problems II: Linear first order systems without turning points," submitted to *SIAM J. Numer. Anal.*

[4] J. Barnes, "A two-dimensional simulation of MESFET's" Thesis, University of Michigan, 1976.

[5] E. P. Doolan, J. J. H. Miller, and W. H. A. Schilders, "Uniform numerical methods for problems with initial and boundary layers," in *Proc. Nasecode I Conf.*, Boole Press, Dublin, Ireland, 1980.

[6] P. C. Fife, "Semilinear elliptic boundary value problems with small parameters," *Arch. Rational Mech. and Anal.*, vol. 29, pp. 1–17, 1973.

[7] A. F. Franz, G. A. Franz, S. Selberherr, C. A. Ringhofer, and P. A. Markowich, "Finite boxes: A generalization of the finite difference method suitable for semiconductor device simulation," presented at the Conf. on Numerical Simulation of VLSI Devices, Boston, MA, 1982.

[8] P. A. Markowich, "A two-dimensional singular perturbation analysis of semiconductor devices," to appear as MRC Technical Summary Rep., 1982.

[9] P. A. Markowich, C. A. Ringhofer, S. Selberherr, and E. Langer, "A singularly perturbed boundary value problem modelling a semiconductor device," MRC Tech. Summary Rep. no. 2388, 1982.

[10] ——, "An asymptotic analysis of single junction semiconductor devices," to appear as MRC Tech. Summary Rep., 1982.

[11] P. A. Markowich and C. A. Ringhofer, "Collocation methods for the numerical solution of boundary value problems on 'Long' Intervals," to appear in *Math. Comp.*, Jan. 1983.

[12] M. S. Mock, "An initial value problem from semiconductor device theory," *SIAM J. Math. Anal.*, vol. 5, no. 4, pp. 587–612, 1974.

[13] W. V. Van Roosbroeck, "Theory of flows of electrons and holes in germanium and other semiconductors," *Bell Syst. Tech. J.*, vol. 29, pp. 560–607, 1950.

[14] D. L. Scharfetter and H. K. Gummel, "Large-scale analysis of a silicon read diode oscillator," *IEEE Trans. Electron Devices*, vol. ED-16, pp. 64–77, 1969.

[15] A. Schuetz, S. Selberherr, and H. W. Poetzl, "A two-dimensional model of the avalanche effect in MOS transistors," *Solid-State Electron.*, vol. 25, no. 3, pp. 177–183, 1981.

[16] S. Selberherr, (1981) "Zweidimensionale Modellierung von MOS-transistoren," Thesis, Technical University of Vienna, Austria, 1981.

[17] D. Smith, "On a singularly perturbed boundary value problem arising in the physical theory of semiconductors," Report Institut fuer Mathematik and Informatik, Technische Universitaet, Muenchen, Germany, 1980.

[18] A. B. Vasileva and V. G. Stelmakh, "Singularly distributed systems in the theory of semiconductor devices," *USSR Comput. Math. Phys.*, vol. 17, pp. 48–58, 1977.