A SINGULAR PERTURBATION ANALYSIS OF DISCRETISATION METHODS FOR THE PARABOLIC SEMICONDUCTOR DEVICE EQUATIONS

WOLFGANG AGLER *
PETER A. MARKOWICH *
SIEGFRIED SELBERHERR **

* Institut für Angewandte und Numerische Mathematik, Technische Univesität Wien, Gusshausstraße 25 - 29, A - 1040 Wien, Austria

** Institut für Allgemeine Elektrotechnik,Technische Univesität Wien, Gusshausstraße 25 - 29, A - 1040 Wien, Austria

ABSTRACT

The fundamental transient semiconductor device equations are scaled appropriately such that a singular perturbation problem is obtained. The singular perturbation parameter λ^2 (which is physically identified as the square of the minimal Debye length of the device under consideration) appears as multiplier of the Laplace-operator in Poisson's equation. We derive asymptotic expansions for the solutions of the semiconductor problem as $\lambda^2 \to 0$ by splitting up the solutions into (spatially) slowly varying and fast varying terms. The fast varying terms (layer terms) dominate the behaviour of the solution in thin regions (layer regions) about junctions between differently doped areas of the device (for example pn-junctions) and decay exponentially away from the junctions. The slowly varying terms (reduced solutions) solve the 'zero - space - charge - approximation' and are independent of λ.

The local transient behaviour of the potential and the carrier densities is used to construct efficient spatial grids and time step-sequences for discretisation methods by equidistributing the local discretisation error. It is demonstrated that equidistributing meshes can be chosen such that the number of grid points is independent of the perturbation parameter λ (and only depends on the bias and the prescribed error tolerance). We illustrate the grid construction by presenting numerical results for the transient Gummel - algorithm applied to a one - dimensional diode.

1. THE SINGULARLY PERTURBED SEMICONDUCTOR DEVICE EQUATIONS

The basic transient semiconductor device equations are (using standard notation) (see Van Roosbroeck (1950)) :

$$(1.1)\ (a) \qquad \varepsilon \Delta\psi = q\,(n - p - C)$$

$$(1.1)\ (b) \qquad n_t = \operatorname{div}\,(D_n \nabla n - \mu_n n \nabla\psi) - R\,(n,p)$$

$$(1.1)\ (c) \qquad p_t = \operatorname{div}\,(D_p \nabla p + \mu_p p \nabla\psi) - R\,(n,p)$$

in some bounded spatial domain Ω representing the device geometry and $t > 0$. (1.1) is supplemented by mixed boundary conditions. ψ, n and p are prescribed at Ohmic contacts, the Dirichlet - boundary condition for ψ is time - varying, those for n and p are time - independent and $\nabla\psi$, ∇n ∇p vanish at insulating segments. Also semiconductor - oxide interfaces and Schottky contacts can be included into the following analysis.
The detailed form of the boundary conditions and of the recombination - generation term R is insignificant for the following discussion. Also initial conditions $\psi\ (t = 0)$, $n\ (t = 0)$, and $p\ (t = 0)$, which fulfill (1.1) (a) are given.

We define (assuming that $D_{n,p}$ and $\mu_{n,p}$ are constant) :

$$(1.2) \qquad \bar{C} = \max_{\Omega} |C| \ , \quad \ell = \operatorname{diam}\,(\Omega) \ , \quad \tau = \frac{\ell^2}{D_n}$$

and employ the following scaling (see Markowich, Ringhofer, Selberherr and Langer (1982) and Markowich (1983)) :

$$(1.3)\ (a) \qquad \psi_s = \frac{\psi}{U_T} \ , \quad n_s = \frac{n}{\bar{C}} \ , \quad p_s = \frac{p}{\bar{C}} \ , \quad D = \frac{C}{\bar{C}}$$

$$(1.3)\ (b) \qquad \vec{x}_s = \frac{\vec{x}}{\ell} \ , \quad t_s = \frac{t}{\tau}$$

where $\vec{x}$ is the spatial (independent) variable and the subscript s denotes the scaled quantities. Then (1.1) transforms (after dropping the subscript s) to

$$\left.\begin{aligned} &(1.4)\ (a) \qquad \lambda^2 \Delta\psi = n - p - D \\ &(1.4)\ (b) \qquad n_t = \operatorname{div}\,(\nabla n - n \nabla\psi) - S\,(n,p) \\ &(1.4)\ (c) \qquad p_t = \operatorname{div}\,(\nabla p + p \nabla\psi) - S\,(n,p) \end{aligned}\right\} \quad \begin{matrix} \vec{x} \in \Omega \\ t > 0 \end{matrix}$$

with $S(n,p) = \frac{\tau}{\bar{C}} R\,(\bar{C}n, \bar{C}p)$ and

$$(1.5) \qquad \lambda^2 = \left(\frac{\lambda_D}{\ell}\right)^2 = \frac{\varepsilon U_T}{\ell^2 q \bar{C}}$$

where λ_D is the minimal Debye length of the device under consideration.

The scaling described above and the scaled problem (1.4) change slightly when $D_{n,p}$ and $\mu_{n,p}$ are not constant but it is completely sufficient to use (1.4) for conceptual considerations.

Note that (for modern Silicon devices) τ is of the same order of magnitude as the electron - and hole - lifetime.

Typically, $\lambda^2 << 1$ (e.g. $\lambda^2 \approx 10^{-6}$ for a Silicon device with $\ell = 5 \cdot 10^{-3}$ cm and a maximal doping of $\bar{C} = 10^{17}$ cm^{-3}), therefore (1.4) constitutes a singularly perturbed boundary value problem. For simplicity we assume that the device we consider has only one abrupt pn - junction Γ. That means that the scaled doping profile D has a jump - discontinuity across Γ. Therefore the solutions $\bar{n}$ and $\bar{p}$ of the reduced problem (that is (1.4) with $\lambda = 0$, also referred to as 'vanishing - space - - charge - approximation') are discontinuous across Γ and (1.4) (b) (c) imply that the reduced potential $\bar{\psi}$ has a jump - discontinuity across Γ, too.

Singular perturbation theory can be used to show that the reduced solutions $\bar{n}$, $\bar{p}$ and $\bar{\psi}$, subject to appropriate boundary conditions, interface conditions at Γ and the reduced initial conditions $\bar{n}$ $(t = 0)$, $\bar{p}$ $(t = 0)$ and $\bar{\psi}$ $(t = 0)$ are close to the 'full solutions' n, p and ψ (the solution of (1.4)) away from Γ and from $\partial\Omega$ when λ is small and that an internal layer (a region of fast variation) occurs at Γ. It also can be shown that no layer occurs at Ohmic contacts and at insulating segments (a boundary layer generally occurs at Schottky contacts and at oxide - semiconductor interfaces) (see Markowich (1983)).

In order to obtain a uniform approximation to ψ, n and p correction terms $\hat{\psi}$, $\hat{n}$ and $\hat{p}$ have to be added to $\bar{\psi}$, $\bar{n}$ and $\bar{p}$. These correction terms fulfill

$$(1.6) \qquad \left|\hat{\psi}\left(\frac{r}{\lambda}, s, t\right)\right|, \left|\hat{n}\left(\frac{r}{\lambda}, s, t\right)\right|, \left|\hat{p}\left(\frac{r}{\lambda}, s, t\right)\right| \leq$$

$$\leq C_1(s,t) \exp\left(-C_2(s,t)\frac{r}{\lambda}\right)$$

where $r = r(\vec{x})$ denotes the distance from $\vec{x}$ to Γ, $s = s(\vec{x})$ is the point on Γ closest to $\vec{x}$ and C_1, $C_2 > 0$ depend on the potential drop across Γ (at s) at time t (see Markowich (1983)). The layer terms decay exponentially away from Γ. Estimates for the derivatives of the layer terms with respect to r and s are obtained by differentiating the right hand side of (1.6). Their derivatives with respect to t are uniformly bounded as $\lambda \to 0+$.

It can be shown that $\bar{\psi} + \hat{\psi}$, $\bar{n} + \hat{n}$ and $\bar{p} + \hat{p}$ approximate the solution of (1.4) ψ, n and p resp. uniformly (in λ) throughout Ω for $t > 0$ (if the potentials applied to the Ohmic contacts approach stationary limits as $t \to \infty$).

The derivatives of the solutions ψ, n, p in perpendicular direction to the junction are large (at most $O(\lambda^{-i})$ for the i-th derivative), those in tangential direction and in t - direction are uniformly bounded in λ.

Qualitatively, the results remain valid for doping profiles which are exponentially graded at junctions and analogous asymptotic results hold close to oxide - semi - conductor interfaces (inversion layers) and Schottky contacts.

2. DIFFERENCE METHODS

We assume that Gummel's method is used for the time discretisation of (1.4) (see Mock (1976)), that the Laplacian is discretised with the usual second order accurate scheme (7 points in three spatial dimensions, 5 points in two spatial dimensions and 3 points in one spatial dimension) and that the spatial operator in the continuity equations is discretised with the Scharfetter - Gummel - scheme (see Markowich, Ringhofer and Selberherr (1982)).

As basic concept for the grid generation we use equidistribution of the local discretisation error combined with the finite - boxes - approach (see G. Franz et al (1982)). Since the Scharfetter - Gummel - method is uniformly (in λ) convergent, it suffices to equidistribute the functional

$$(2.1) \qquad E^m_{ij\ell}(\psi) =$$

$$= \lambda^2 h_i^{xm} \left| \psi_{xxx}(x_i, y_j, z_\ell, t_m) \right| + \lambda^2 h_j^{ym} \left| \psi_{yyy}(x_i, y_j, z_\ell, t_m) \right|$$

$$+ \lambda^2 h_\ell^{zm} \left| \psi_{zzz}(x_i, y_j, z_\ell, t_m) \right| + \Delta t_m \left| \psi_t (x_i, y_j, z_\ell, t_m) \right|$$

where h_i^{xm}, h_j^{ym} and h_ℓ^{zm} denote the spatial mesh - sizes in corresponding direction and Δt_m is the m-th temporal mesh - size ($t_{m+1} = t_m + \Delta t_m$; $x_{i+1} = x_i + h_i^{xm}$ in the m-th time step & analogously for y_{j+1} and $z_{\ell+1}$).

$E^m_{ij\ell}(\psi)$ is (approximately) the local discretisation error of Poisson's equation. Equidistribution of this term means that the mesh is chosen such that $E^m_{ij\ell}(\psi) = O(\delta)$ on $\Omega \times [0,T]$, where δ is the accuracy parameter and $[0,T]$ is the desired simulation time - period. (1.6) implies that equidistributing spatial grids are exponentially graded near junctions. Away from junctions the spatial grid - sizes should be limited by $O(\delta)$ (since the continuity equations have no multiplying factor λ^2).

Since ψ_t is uniformly bounded (in λ) time steps of width at least $O(\delta)$ can be chosen. Note that the spatial mesh - sizes generally depend on the time step (moving grid). This is due to the fact that the widths of the junction - layers depend on the applied potentials (the layers can shrink and extend, but do not move away from the junctions as t increases). Therefore spatially intermediate values for the discrete solution have to be obtained by interpolation. Of course time steps are allowed to increase when $\psi(\vec{x},t)$ gets close to its stationary limit $\psi(\vec{x},\infty)$.

To illustrate the grid construction we take a one - dimensional diode (with the pn - junction at $x = 0$) and obtain :

$$(2.2) \qquad E^m_i(\psi) = \lambda^2 h_i^m \left| \psi_{xxx}(x_i, t_m) \right| + \Delta t_m \left| \psi_t (x_i, t_m) \right|$$

(1.6) yields

$$(2.3) \qquad E^m_i(\psi) \approx \frac{h_i^m}{\lambda} e^{-C_m \frac{|x_i|}{\lambda}} + \Delta t_m \left| \psi_t (x_i, t_m) \right|$$

inside the layer, where C_m only depends on the bias applied at $t = t_m$.

The optimal equidistributing spatial grid therefore is

$$(2.4) \qquad h_i^m = \delta \lambda e^{C_m \frac{|x_i|}{\lambda}}$$

(see Ascher and Weiss (1981)). The number of grid points inside the layer is $O(\delta^{-1})$ (and is therefore asymptotically independent of λ !)

Spatial grids for two dimensional problems are constructed in Markowich, Ringhofer and Selberherr (1982).

The time - grid is chosen such that

$$(2.5) \qquad \Delta t_m \max_{x_i} | \psi_t (x_i, t_m) | = \delta$$

holds (see Chong (1978)).

Of course, the equidistribution technique can be applied to other time - discretisation - methods by modifying the local error functional (2.1) (see Mock (1976) for other discretisation schemes).

The equidistribution technique was used for the simulation of a one - dimensional diode (with $\ell = 2.5 \times 10^{-3}$; $C = -10^{17}$ in $[-\ell, -\frac{1}{2}\ell]$, $C = 10^{15}$ in $(-\frac{1}{2}\ell, \ell]$; that gives $\lambda^2 = 2.6744 \times 10^{-7}$) with the applied bias shown in Figure 1.

The accuracy parameter was set to 10^{-2} and between 120 and 135 spatial grid points were used (the exact number depends on the time step). 305 time steps were used on the simulation interval [0,2.3]. Figure 2 shows the potential distribution and Figure 3 the electron density. The accumulation of electrons during the high - injection time interval [0.5,1] is clearly visible. It can be seen that the layer width depends on time (it is largest in the high - injection time interval).

A discussion of the convergence properties of discretisation schemes for Poisson's equation when equidistribution is relaxed , is given in Markowich, Ringhofer and Selberherr (1982) for the stationary case.The results carry over to the parabolic semiconductor problem.

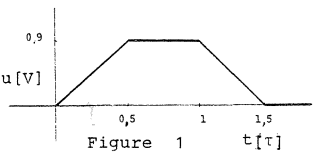

Figure 1

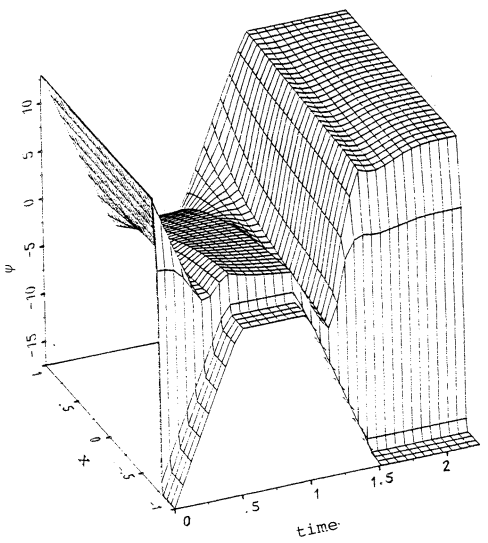

Figure 2

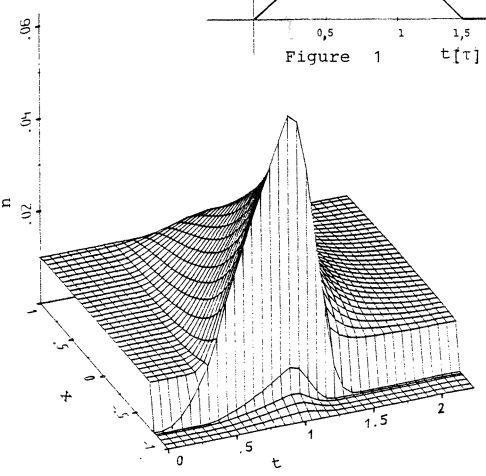

Figure 3

3. REFERENCES

U. Ascher and R. Weiss (1981)

'Singular Perturbation Problems I ;
First Order Systems with Constant Coefficients'
to appear in SIAM J. Numer. Anal.

A. F. Franz , G. A. Franz , S. Selberherr , C. A. Ringhofer and P. A. Markowich (1982)

'Finite Boxes : A Generalisation of the Finite Difference Method Suitable for Semiconductor Device Simulation'
presented at the conference on 'Numerical Simulation of VLSI Devices'. Boston, 1982

P. A. Markowich , C. A. Ringhofer , S. Selberherr and E. Langer (1982)

'A Singulary Perturbed Boundary Value Problem Modelling A Semiconductor Device'
to appear in SIAM J. Appl. Math.

P. A. Markowich , C. A. Ringhofer and S. Selberherr (1982)

'A Singular Perturbation Approach for the Analysis of the Fundamental Semiconductor Equations'
presented at the conference on 'Numerical Simulation of VLSI Devices'. Boston, 1982

P. A. Markowich (1983)

'Asymptotic Analysis of the Fundamental Semiconductor Device Equation'
to appear as MRC - TSR

M. Mock (1976)

'Time Discretisation of a Nonlinear Initial Value Problem'
Journal of Comp. Physics, Vol 21, pp 20 - 37

W. V. Van Roosbroeck (1950)

'Theory of Flow of Electrones and Holes in Germanium and other Semiconductors'
Bell Syst. Techn. Journal, Vol 29, pp 560 - 607

T. H. Chong (1978)

'A Variable Mesh Finite Difference Method for Solving a Class of Parabolic Differential Equations in One Space Variable'
SIAM J. Numer. Anal, Vol 15, No 4, pp 835 - 857