INFORMATION TO USERS

Pseudo arc-length continuation method for multiple solutions in one-dimensional steady state semiconductor device simulation

Tso, Tai-Yih, Ph.D.

Iowa State University, 1991

Pseudo arc-length continuation method for multiple solutions in

one-dimensional steady state semiconductor device simulation

by

Tai-Yih Tso

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Mathematics
Major: Applied Mathematics

Approved:

In Charge of Major Work

For the Major Department

For the Graduate College

Iowa State University
Ames, Iowa
1991

TABLE OF CONTENTS

LIST OF FIGURES

v

ACKNOWLEDGEMENTS

1. INTRODUCTION

1.1 Overview

Since a research team at the Bell Telephone Laboratories built the first bipolar transistor and developed the theory of semiconductor physics in the late 1940s, semiconductor devices have formed the foundation of modern electronics. The theoretical foundation for the p-n junction diode and the transistor was established by Shockley in his 1949 paper [1]. Van Roosbroeck had formulated a system of basic semiconductor device equations which is the most commonly used in numerical device simulations [2]. These basic semiconductor device equations are a coupled system of nonlinear second order partial differential equations. They describe the distributions of electrostatic potential and carrier concentration and current flow within semiconductor devices.

Prior to semiconductor device simulation using computer-aided design, the analyses of semiconductor devices were based on the regional approximation methods to obtain closed-form approximate analytical solutions [3] [4]. Although this regional approach allows for simplifying the model and getting rapid analysis, it is understood that this approach is not suitable when a unified device model is desired, especially in modeling of sub-micron VLSI. And the traditional experimental approach is expensive for developing new complex integrated circuits. Therefore, numerical simulation

for semiconductor devices was introduced.

The first work in semiconductor modeling using a numerical method instead of the regional analytical method was proposed by Gummel. In his 1964 paper [5], Gummel successfully demonstrated a numerical method in simulating the one-dimensional steady state transistor. This method was applied to the p-n junction by De Mari [6] [7] and to the silicon Read diode by Scharfetter and Gummel [8]. It was further developed in two-dimensions by Slotboom [9] and others. Regarding the historical development of numerical device modeling, see the books by Selberherr [10] and Snowden [11].

The basic interest in this dissertation is numerical simulation of one-dimensional steady state thyristors. A thyristor is a semiconductor device with four layer p-n-p-n structure, used to control the switching of *dc* and *ac* power. Traditional analysis for thyristors is with reference to the two bipolar transistors analogy for a thyristor [4]. In this two-transistor analogy, the collector of each transistor is connected to the base of the complementary transistor. This is shown in Figure 1.1.

Numerical computation for thyristors is quite difficult because they have multiple steady solutions under certain biasing conditions. The snap-back phenomenon in the current-voltage characteristic of thyristors has caused computational problems. To overcome the snap-back and multi-solution problems, we have applied the arc-length continuation method. We successfully obtain the current-voltage characteristic of thyristors by this method.

This dissertation is organized as follows. In Chapter 1, the basic semiconductor device equations are introduced. Then we explain briefly the concepts of singular perturbation analysis and numerical methods for solving these equations. The second chapter is devoted to the review of the analyses of existence and uniqueness, regional
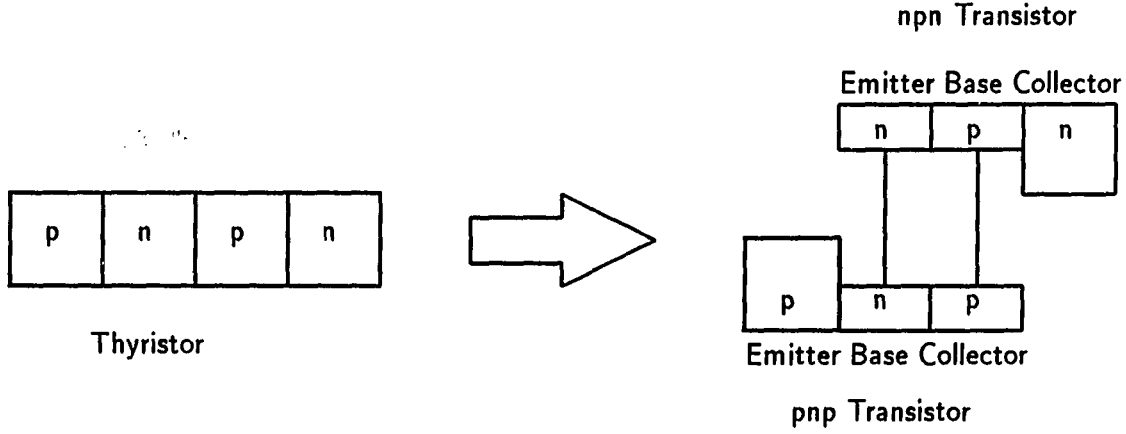
Figure 1.1: Two transistors analogy

approximation and singular perturbation for the semiconductor equations. Chapter 3 discusses the discretization of the semiconductor equations and mesh generation. In Chapter 4, we investigate numerical schemes for solving the discrete semiconductor equations and the implementation of various linear system solvers. In the final chapter, we present some computational results on the diode and thyristor from our semiconductor simulator.

1.2 Basic semiconductor device equations

In order to analyze the semiconductor device characteristics, we require a suitable mathematical model describing *electrical* and *physical* processes. The basic semiconductor device equations are the commonly adopted model for numerical simulations. They are the Poisson equation, two continuity equations and two current equations. To simulate the behavior of a semiconductor device, the basic semicon-

ductor device equations have to be solved for the electrostatic potential ψ, electron concentration n and hole concentration p.

The electrostatic potential ψ is determined by the Poisson equation:

$$\nabla^2 \psi = -\frac{\rho}{\varepsilon}, \tag{1.1}$$

where ε is the material permittivity and ρ is the space charge density. The space charge density ρ can be expressed as $\rho = q(p - n + C(x))$, where q is the elementary charge. The unknowns n and p are charge densities for electrons and holes, respectively. The function C is called the doping concentration or the doping profile, which is explained in Section 1.3:

From the conservation of charge, the electron and hole densities can be found by the continuity equations:

$$\frac{\partial n}{\partial t} = \frac{1}{q}(\nabla \cdot J_n - qR), \tag{1.2}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}(\nabla \cdot J_p + qR), \tag{1.3}$$

where J_n is electron the current density and J_p is the hole current density. The function R is the net recombination-generation rate. It accounts for a number of physical processes that result in the creation or annihilation of electron-hole pairs. In our device simulator, the total rate R is the sum of Shockley-Read-Hall and Auger recombination rates and impact ionization generation rate which are explained in Section 1.3.

$$R = R_{SRH} + R_{AU} + R_{II}. \tag{1.4}$$

If an electric field E is applied to the semiconductor sample, the carriers will drift in the electric field. This transport of carriers produces a current called the

drift current. The influence of the concentration gradient ∇n, ∇p causes a diffusion carrier flow called the diffusion current. Therefore, the models of current densities of electron and hole and total current densities are given by the drift-diffusion equations:

$$J_n = q(D_n\nabla n - \mu_n n E), \tag{1.5}$$

$$J_p = -q(D_p\nabla p + \mu_p p E), \tag{1.6}$$

$$J = J_n + J_p, \tag{1.7}$$

where μ_n and μ_p are the mobilities of holes and electrons, D_n and D_p are diffusion constants of electrons and holes which are explained in the Section 1.3. The electric field E is, as usual, given by

$$E = -\nabla\psi. \tag{1.8}$$

Mobilities and diffusion constants are often taken to be connected by the Einstein relations:

$$D_n = \mu_n U_T, \tag{1.9}$$

$$D_p = \mu_p U_T. \tag{1.10}$$

Here U_T, called the thermal voltage, is defined by

$$U_T = \frac{\kappa T}{q}, \tag{1.11}$$

where κ is Boltzmann's constant and T is the absolute temperature. Under the Einstein relations, the current equations can be rewritten in the well known form:

$$J_n = q\mu_n(U_T\nabla n - nE), \tag{1.12}$$

$$J_p = -q\mu_p(U_T\nabla p + pE). \tag{1.13}$$

Now we summarize the equations forming the basic semiconductor device equations for the quantities ψ, p and n.

$$\nabla^2 \psi = \frac{q}{\varepsilon}(n - p - C(x)),$$ (1.14)

$$\frac{\partial n}{\partial t} = \frac{1}{q}(\nabla \cdot J_n - qR),$$ (1.15)

$$\frac{\partial p}{\partial t} = -\frac{1}{q}(\nabla \cdot J_p + qR),$$ (1.16)

$$J_n = q\mu_n(U_T \nabla n - nE),$$ (1.17)

$$J_p = -q\mu_p(U_T \nabla p + pE),$$ (1.18)

$$J = J_n + J_p.$$ (1.19)

The bounded domain $\Omega \in R^N(N = 1, 2, 3)$ of the basic semiconductor equations is the physical extent of the actual device. The boundary conditions for the basic semiconductor equations under Ohmic contacts are derived from the following three physical requirements which are explained in the next section:

$$pn = n_i^2,$$ (1.20)

$$n - p - C = 0,$$ (1.21)

$$\psi = \psi_{bi} + V_o,$$ (1.22)

where n_i is called the intrinsic carrier density and V_o is the externally applied voltage. The so-called built-in voltage ψ_{bi} is defined by

$$\psi_{bi} = U_T \ln\left[\frac{C + \sqrt{C^2 + 4n_i^2}}{2n_i}\right].$$ (1.23)

1.3 Physical basis

A solid-state material can be classified as a conductor, semiconductor or insulator. This classification is based on the ability of the material to produce free charge carriers since the movement of carriers constitutes a current flow. The intrinsic semiconductor is just a solid material containing fewer charge carriers than a conductor but more than an insulator. The intrinsic carrier density n_i(i.e. the number of electrons or holes per cubic centimeter) can be determined by the theory of energy band and density of states in quantum mechanics [10]:

$$n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2\kappa T}\right), \qquad (1.24)$$

where $\kappa = 1.38066 \times 10^{-23} Joule/K$ is Boltzmann's constant and T is the absolute temperature. The quantities N_c and N_v are the effective densities of states in the conduction and valence band, respectively. At room temperature($300K$), N_c is $2.8 \times 10^{19} cm^{-3}$ and N_v is $1.04 \times 10^{19} cm^{-3}$ for silicon. The value of the bandgap E_g is $1.12eV$ at room temperature. If we substitute the quantities N_c, N_v, E_g and κ into the equation (1.24), at room temperature, then the intrinsic density n_i is $1.45 \times 10^{10}/cm^3$ for silicon.

To make a useful semiconductor device, certain impurities are added to the semiconductor in very carefully controlled amounts. This process is called doping. The concentration of added impurities, or *doping concentration* essentially determines the device function. When a semiconductor is doped with impurities, the impurities create two types of extrinsic semiconductor. One, with an excess of electrons, is called n-type and impurities are called *donors*. The other with an excess of holes is called p-type and impurities are called *acceptors*. The doping profile is the difference

of the donor density and the acceptor density. The semiconductor containing both p-type and n-type regions forms a p-n junction.

The density of electrons and holes will keep a dynamic balance between the generation and recombination rates under a thermal equilibrium condition. An important relationship between electron density n and hole density p is the mass action law [4].

$$np = n_i^2$$

This mass action law always holds for both intrinsic and extrinsic semiconductors at thermal equilibrium. The electron and hole densities in thermal equilibrium are given by [10]

$$n = n_i \exp\left(\frac{q\psi}{\kappa T}\right), \tag{1.25}$$

$$p = n_i \exp\left(\frac{-q\psi}{\kappa T}\right). \tag{1.26}$$

Most semiconductor devices operate under non-equilibrium conditions(i.e. $np \neq n_i^2$) by such means as electrical or optical excitation. Whenever the thermal equilibrium condition(i.e. $np = n_i^2$) is disturbed certain physical processes act to restore it. These processes are performed by carrier recombination and generation. Several physical mechanisms describe the recombination-generation phenomenon [10]. Three types of recombination-generation rates are considered in our semiconductor device simulator: Shockley-Read-Hall, Auger and impact ionization. An indirect recombination process dominates in materials such as germanium and silicon [4]. This process may be treated by the well-known Shockley-Read-Hall expression R_{SRH}. It is expressed as

$$R_{SRH} = \frac{np - n_i^2}{\tau_p(n + n_i) + \tau_n(p + n_i)}$$

where n_i is the intrinsic density, τ_n and τ_p are the life times of electrons and holes. respectively. Auger recombination describes the process of three particle transitions. This means that the recombination of an electron and hole releases energy, exciting a third carrier to some higher energy. The Auger recombination is significant on high power devices. It is expressed as

$$R_{AU} = (np - n_i^2)(C_p p + C_n n),$$

where C_p and C_n are the Auger capture coefficients for holes and electrons, respectively. When the electric fields are high enough, the carriers gain enough kinetic energy to break valence bonds, generating electron-hole pairs. These generated pairs accelerate in the high electric fields and collide with the atoms in the crystal lattice generating other electron-hole pairs. This is called the avalanche or impact ionization process. This impact ionization process is treated by the impact ionization rate R_{II}. It is expressed as

$$R_{II} = -\frac{1}{q}(\alpha_n \mid J_n \mid + \alpha_p \mid J_p \mid),$$

where α_n and α_p are impact ionization coefficients of electrons and holes, respectively. They are strongly dependent on electric field components in the direction of current flow. Commonly used models(see [10]) are as follows:

$$\alpha_n = \alpha_n^\infty \exp\left(-\frac{E_n^c}{|E|}\right), \tag{1.27}$$

$$\alpha_p = \alpha_p^\infty \exp\left(-\frac{E_n^p}{|E|}\right). \tag{1.28}$$

At room temperature, α_n^∞ is $10^6 cm^{-1}$, α_p^∞ is $2\times10^6 cm^{-1}$, E_n^c is $1.66\times10^6 Volts/cm$, and E_n^c is $2 \times 10^6 Volts/cm$ for silicon. When an electric field E is applied to the

semiconductor sample, the electrons drift with mean velocity v_n proportional to the field. The proportionally factor μ_n is called the electron mobility. That is,

$$v_n = -\mu_n E, \qquad (1.29)$$

where the negative sign means the electrons drift against the direction of the electric field. The acceleration in the direction of the electric field E is $-qE/m_n^*$, where m_n^* is the effective electron mass. Let t_n be the electron relaxation time which is the average time between electron collisions in the crystal lattices; then by equating the momentum($force \times time$) we have

$$v_n = (\frac{-qE}{m_n^*})t_n = -\mu_n E. \qquad (1.30)$$

Thus,

$$\mu_n = \frac{qt_n}{m_n^*}. \qquad (1.31)$$

By similar argument, we can express the hole mobility μ_p:

$$\mu_p = \frac{qt_p}{m_p^*} \qquad (1.32)$$

where m_p^* is the effective hole mass. These carrier mobilities are complicated functions because of the relaxation times between the collisions. The relaxation times are determined by the various scattering mechanisms.

A semiconductor contact which has a negligible resistance regardless of the polarity of the externally applied voltage is called an Ohmic contact. Usually, the zero-space-charge condition(i.e. $n - p - C'(x) = 0$) is assumed to hold at Ohmic contacts. Nonzero space charge in a highly doped semiconductor contact causes an extremely high electric field, resulting in a breakdown condition at the contact. The

thermal-equilibrium condition(i.e. $np = n_i^2$) is also assumed to hold at Ohmic contacts. This means that the excess carriers vanish immediately. The zero-space-charge condition(i.e. $n - p - C(x) = 0$) and thermal-equilibrium condition(i.e. $np = n_i^2$) determine the boundary conditions for n and p at Ohmic contacts.

The boundary conditions for ψ are derived as follows. At thermal equilibrium without externally applied voltage, the built-in voltage can be derived by substituting equations (1.25), (1.26) into $n - p - C(x) = 0$. That is,

$$n_i \exp\left(\frac{q\psi_{bi}}{\kappa T}\right) - n_i \exp\left(\frac{q\psi_{bi}}{\kappa T}\right) - C = 0. \tag{1.33}$$

Thus,

$$\psi_{bi} = U_T \ln\left[\frac{C + \sqrt{C^2 + 4n_i^2}}{2n_i}\right]. \tag{1.34}$$

Therefore, the boundary conditions for ψ under the applied voltage V_o is given by $\psi = \psi_{bi} + V_o$ at Ohmic contacts.

1.4 Scaled form of the equations

The dependent variables in the semiconductor device equations: electrostatic potential ψ, electron density n, and hole density p, are naturally chosen from the physical view point. However, some works in device modeling use different sets of dependent variables instead of (ψ, n, p) for analytical and computational purposes. One set of variables which is used in some computational works is the electrostatic potential ψ, electron quasi-Fermi level φ_n and hole quasi-Fermi level φ_p. These quasi-Fermi levels are defined by

$$\varphi_n = \psi - U_T \ln\left(\frac{n}{n_i}\right), \tag{1.35}$$

$$\varphi_p = \psi + U_T \ln\left(\frac{p}{n_i}\right). \tag{1.36}$$

Mathematically, these state variables $(\psi, \varphi_n, \varphi_p) \in R^3$ are related in a one-to-one way to the set $(\psi, n, p) \in R \times (0, \infty)^2$. The range of values of φ_n and φ_p is smaller but the equations become more non-linear. Another set of variables (ψ, u, v), which is most used in analytical works, is directly derived from the quasi-Fermi levels set $(\psi, \varphi_n, \varphi_p)$. These are defined by

$$u = \frac{n}{n_i} \exp\left(\frac{-\psi}{U_T}\right), \tag{1.37}$$

$$v = \frac{p}{n_i} \exp\left(\frac{\psi}{U_T}\right). \tag{1.38}$$

The equations for these variables (ψ, u, v) are more linear and self-adjoint but the range of values of u and v is too large to use in practical computations. Kurata used the electric field E instead of electrostatic potential ψ in his current-control type computation for thyristors [12]. The choice of variables (E, n, p) simplifies equations but encounters the problem of numerical divergence. The state variables (ψ, n, p) are used in our numerical computations.

The variables in the basic semiconductor device equations have greatly different magnitudes in different space regions. Therefore, the variables and equations must be scaled to obtain dimensionless equations. The scaling based on singular perturbation analysis [13] is summarized as follows.

1. The scaling factor for length is the maximal diameter of the device domain.

2. The scaling factor for all potentials is the thermal voltage U_T.

3. The scaling factor for all densities is the maximum doping profile C.

4. The scaling factor for mobilities is a reference carrier mobility.

After scaling, the device equations in one-dimensional steady state under Ohmic contacts take the form:

$$\lambda^2 \frac{d^2\psi}{dx^2} = (n - p - C(x)), \tag{1.39}$$

$$\frac{dJn}{dx} = R(x, \psi, n, p), \tag{1.40}$$

$$\frac{dJp}{dx} = -R(x, \psi, n, p), \tag{1.41}$$

$$Jn = \mu n \left(\frac{dn}{dx} - n \frac{d\psi}{dx} \right), \tag{1.42}$$

$$Jp = -\mu p \left(\frac{dp}{dx} + p \frac{d\psi}{dx} \right), \tag{1.43}$$

The bounded doman $[-1, 1]$ is the scaled domain. The parameter λ is given by

$$\lambda = \frac{\lambda_D}{l} = \frac{\sqrt{\frac{\varepsilon U_T}{qC_o}}}{l}, \tag{1.44}$$

where l is the scaling factor for length and C_o is the scaling factor for density. λ_D the so-called Debye length, so λ is called the normed Debye length. It is very small; typically, its order of magnitude is 10^{-3}–10^{-5}. The scaled boundary conditions at Ohmic contacts are as follows:

$$\psi(\pm 1) = \psi_{bi}(\pm 1) + V_{0\pm 1}, \tag{1.45}$$

$$n(\pm 1) = \frac{1}{2} \left(C(\pm 1) + \sqrt{C^2(\pm 1) + 4\delta^4} \right), \tag{1.46}$$

$$p(\pm 1) = \frac{1}{2} \left(-C(\pm 1) + \sqrt{C^2(\pm 1) + 4\delta^4} \right), \tag{1.47}$$

where ψ_{bi}, the scaled built-in voltage, is given by

$$\psi_{bi}(x) = \ln \left[\frac{C(x) + \sqrt{C^2(x) + 4\delta^4}}{2\delta^2} \right]. \tag{1.48}$$

The symbol δ^2 is the so-called scaled intrinsic density defined by

$$\delta^2 = \frac{n_i}{C_o}.$$

(1.49)

1.5 Mathematical analysis

A review of mathematical analyses for the basic semiconductor device equations in steady state is considered in this section. We will be concerned with three aspects of the basic equations: the existence and uniqueness of solutions, the singular-perturbation character, and the properties of solutions. This analytical knowledge is essential for successful numerical computation.

The scaled system of the basic semiconductor device equations in steady state has a small parameter λ^2 multiplying the second derivative in the Poisson equation. This makes the boundary value problem singularly perturbed and causes the solutions to exhibit layer-type behavior. The so-called reduced equations are derived by setting $\lambda = 0$, i.e. $n - p - C(x) = 0$. This is called the charge-neutral approximation in the physics literature. These reduced equations with the reduced solutions $\bar{\psi}$, $\bar{n}$, $\bar{p}$, $\bar{J}_n$, $\bar{J}_p$ are written in one-dimensional scaled form as

$$0 = \bar{n} - \bar{p} - C(x)$$

(1.50)

$$\frac{d\bar{J}_n}{dx} = \bar{R}(x, \bar{\psi}, \bar{n}, \bar{p})$$

(1.51)

$$\frac{d\bar{J}_p}{dx} = -\bar{R}(x, \bar{\psi}, \bar{n}, \bar{p})$$

(1.52)

$$\bar{J}_n = \mu_n \left(\frac{d\bar{n}}{dx} - \bar{n} \frac{d\bar{\psi}}{dx} \right)$$

(1.53)

$$\bar{J}_p = -\mu_p \left(\frac{d\bar{p}}{dx} + \bar{p} \frac{d\bar{\psi}}{dx} \right)$$

(1.54)

for $x \in [-1,1]$. To simplify the analysis we consider a diode and assume the scaled doping profile C(x) to be piecewise constant:

$$C(x) = \begin{cases} C_p & -1 \leq x < x_j, \\ C_n & x_j < x \leq 1, \end{cases}$$

where the point x_j where C changes sign is called the p-n junction. These reduced equations are compatible with the boundary conditions under Ohmic contact since $\bar{n} - \bar{p} - C(x) = 0$. Therefore, the solution of the basic device equations under Ohmic contact has only a junction layer at the junction x_j where C(x) is discontinuous. This means that the solution varies slowly in the region far away from the junction, and rapidly inside the junction layer. A junction correction term added to the reduced solution is required to make the approximation for the full equations. Markowich applied singular perturbation theory to construct the approximation in the form [13]:

$$\psi(x,\lambda) \sim \bar{\psi}(x) + \hat{\psi}(\frac{x - x_j}{\lambda}) \tag{1.55}$$

where $\bar{\psi}$ is the reduced solution and $\hat{\psi}$ is a correction term which decays exponentially to zero as $(\frac{x-x_j}{\lambda}) \longrightarrow \infty$, i.e. x far away from junctions. He has proved that this asymptotic expression is valid under zero bias, i.e., the recombination-generation term $R = 0$, and the full system has a unique solution at thermal equilibrium. This result can be stated as the following theorem.

Theorem 1.1 *Let the mobilities μ_n, μ_p be constant, and the doping profile $C(x)$ be piecewise constant. Then the one-dimensional steady state basic semiconductor device equation has a unique solution ψ_e under zero-bias and*

$$\|\psi_e(x,\lambda) - \psi_{bi}(x) - \hat{\psi}(\frac{x - x_j}{\lambda})\|_\infty = O(\lambda) \tag{1.56}$$

where $\psi_{bi}(x)$ is the built-in potential.

There is at this time no proof of existence of solutions for the basic semiconductor device equations without any restrictions. A general proof about the uniqueness of solutions for arbitrary applied voltage can not be expected. In fact, semiconductor device physicists tell us that certain devices have multiple steady states for certain biasing conditions. For example, the four-layer p-n-p-n thyristors can have three states.

1.6 Numerical methods

The usual procedure of the numerical methods in solving differential equations contains three steps. First, the domain of the differential equations has to be partitioned into subdomains to generate the mesh points at which the solutions are computed. Secondly, the differential equations are discretized by suitable discretization schemes yielding a system of algebraic equations. Finally, these algebraic equations are solved by some numerical method to obtain the approximate solutions at the mesh points. This introductory chapter only describes the discretization scheme and the methods for solving the discrete equations for our problem. The details of numerical analysis — accuracy, convergence and efficiency — will be discussed in Chapters 3 and 4.

The mesh generation is based on the solution character and the idea of equidistributing the local truncation error. Since the solution varies strongly in the junction layers but slowly in the region far from junctions, more mesh points are required near the junctions. This means that the usual uniform mesh is not appropriate. To achieve a specified accuracy with as few mesh points as possible, we first set up a coarse mesh refined at the junctions, and then refine the mesh further by equidistributing the local

truncation error to achieve the specified accuracy.

The finite difference and the finite element approaches are two basic numerical schemes to discretize the differential equations. In our semiconductor device simulation programs, the so-called box method [14] of the finite difference scheme is used for discretization of the semiconductor device equations. The box method is very suitable for equations in divergence form. To explain this method, a two dimensional Poisson's equation

$$div(\nabla f) = r \qquad (1.57)$$

is discussed for simplicity. By the divergence theorem, equation (1.57) can be replaced by

$$\int_C \nabla f \, ds = \int \int_B r \, dA \qquad (1.58)$$

where B is the box region containing a mesh point P_0 and C is its boundary curve (see Figure 1.2). The right-hand side of equation (1.58) is approximated by

$$\int \int_B r \, dA \approx r(P_0) area(B) \qquad (1.59)$$

and the left-hand side of equation (1.58) is approximated by

$$\int_C \nabla f \, ds \approx \sum_{i=1}^{4} [\nabla f(M_i)\hat{n}] l_i, \quad l_i = | Q_{i+1} - Q_i | . \qquad (1.60)$$

In equation (1.60) $\nabla f(M_i)\hat{n}$ can be approximated by the central difference,

$$\nabla f(M_i)\hat{n} \approx \frac{f(P_i) - f(P_0)}{| P_i - P_0 |}. \qquad (1.61)$$

Using the above approximation, the discretization of Poisson's equation (1.57) leads to the five-point formula

$$\sum_{i=1}^{4} \frac{f(P_i) - f(P_0)}{| P_i - P_0 |} \approx r(P_0) area(B). \qquad (1.62)$$

P4

Q1 M4 Q4

P1 M1 P3

P0 M3

Q2 Q3

M2

P2
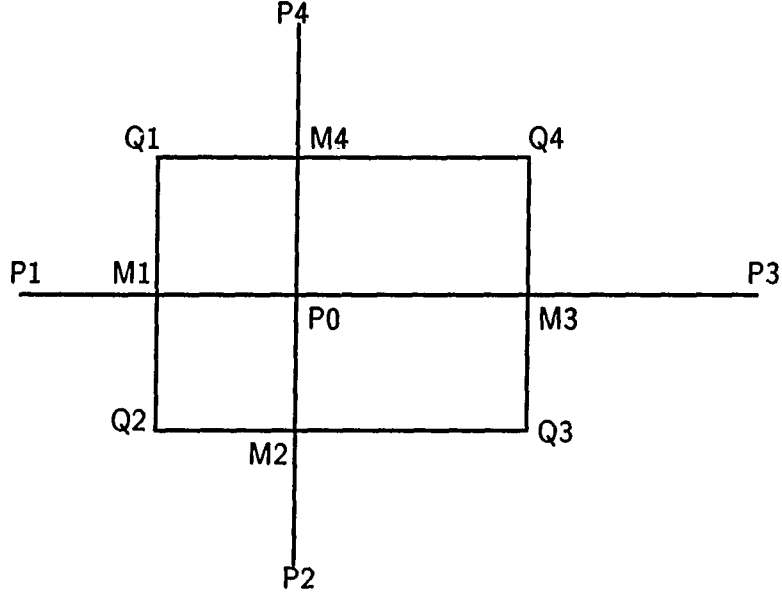
Figure 1.2: Box scheme

This box method can also be applied to the continuity equation and yields the Scharfetter-Gummel discretization [8] for carefully dealing with the current relations by exponential fitting.

The basic semiconductor device equations are a nonlinear system of differential equations, so the discretization of these equations is also a nonlinear system of difference equations. For convenience, the following symbolical notation is used to denote the nonlinear system of discrete equations. By substituting the current relations into the continuity equations, we write the discrete equations symbolically as

$$F(z) = \begin{pmatrix} f_\psi(z) \\ f_n(z) \\ f_p(z) \end{pmatrix} = 0 \qquad (1.63)$$

where $z = (\psi, n, p)^T$, the state variables of the discrete semiconductor equations

at mesh points. The discrete Poisson equation is f_ψ, and f_n, f_p are the discrete continuity equations for electrons and holes, respectively.

Now to solve this nonlinear system of difference equations, the most frequently used algorithms are iterative methods. Two basic iterative methods are used in semiconductor device simulations. A nonlinear Gauss-Seidel procedure(the decoupled algorithm) first proposed by Gummel, is called the Gummel method [5]. The Newton iteration is the so-called the coupled algorithm.

The idea of the original Gummel's method is to solve the electron and hole continuity equations for n and p with ψ held fixed, and then solve Poisson's equation for electrostatic potential ψ, using only one step of Newton's method. The nonlinear Gauss-Seidel iteration converges well for small applied voltage, but slowly for the high-current situations. Computer times in each iteration are saved in this method because only one equation in the full system is solved on each step.

Newton's method, on the other hand, converges quadratically from a good initial approximation but each iteration is expensive, because it requires computing the Jacobian and solving a full system simultaneously. The trade-offs between these two methods have to be considered. One way is to use Gummel's method first to get a good initial guess for Newton's method and then switch to Newton's method.

1.7 Computation of current-voltage charcteristics in the one-dimensional thyristor model

The major goal in this paper is to compute the current-voltage characteristics for the one-dimensional static thyristor (see Figure 1.3). There are three different current states in a thyristor under certain biases. From a mathematical point of
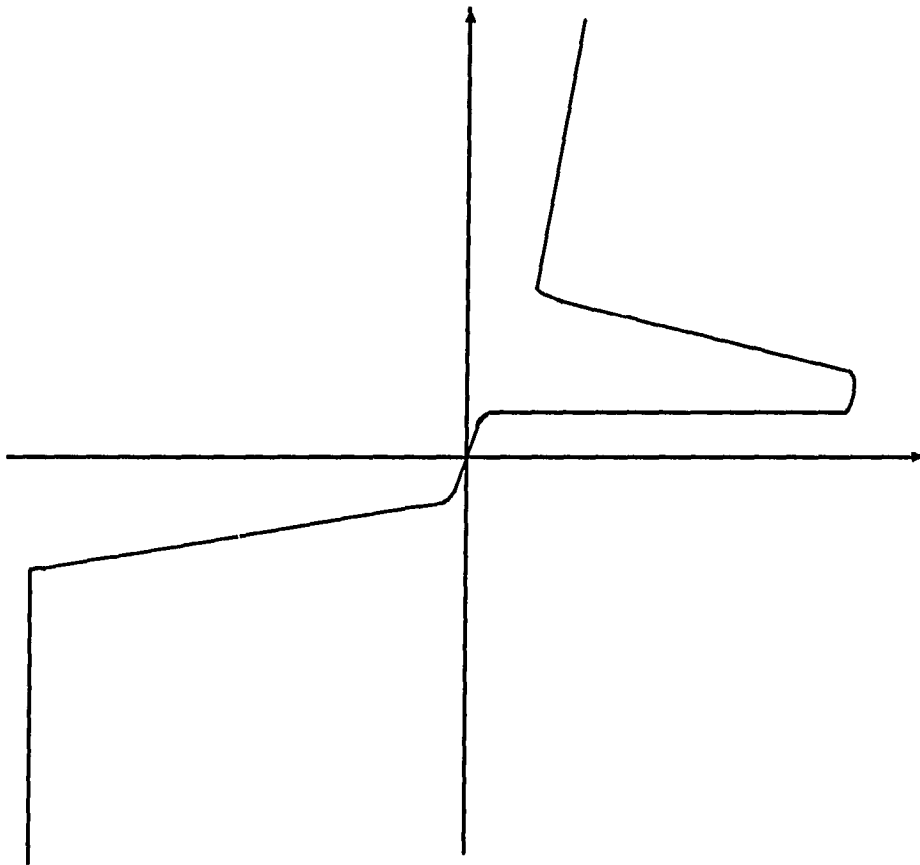
Figure 1.3:   Current-voltage characteristics of a thyristor

view, this means that the static basic semiconductor device equations for a thyristor have nonunique solutions. This multi-solution characteristic creates difficulties for numerical computation. To overcome this problem, we use the pseudo arc-length continuation method [15]. Newton's method and Gummel's method can be treated as basic tools for this continuation method.

With the notation of (1.63), the discrete system of the boundary value problem is

$$F(z, V) = 0 \qquad (1.64)$$

where $F : R^N \times R \longrightarrow R^N$, and $V \in R$, the applied voltage on a contact, is the natural parameter in this system.

The solution path $\Gamma = \{(z, V) \mid F(z, V) = 0\}$ of equation (1.64) must be computed to get the current-voltage curve(see Figure 1.3). There is a way for solving this problem just by Newton's method. However, Newton's method fails to converge when a poor initial guess is given. The natural continuation method is introduced to overcome this problem. The natural continuation method uses a previous known solution and the tangent vector to Γ to construct the initial guess for Newton's method. This algorithm is written as follows.

Algorithm 1 (Natural Contiuation)   *1. Start at a known solution (z_0, V_0) on the solution path.*

2. Compute tangent vector z_v by $F_z z_v = -F_v$.

3. Predict an initial approximation by $z^0 = z_0 + (v_1 - v_0)z_v$.

4. Use z^0 as an initial approximation for Newton's step $F_z(z^i - z^{i-1}) = -F(z^{i-1}, V_1)$.

5. Use the solution (z_1, V_1) from step 4 as new (zo, Vo) and go to step 1.

This natural continuation procedure works well for sufficiently small step size ($V_1 - V_0$) but needs some modification for dealing with multi-solution problems. A curve with multiple solutions will usually have turning points. The Jacobian F_z is singular at the turning points. To pass these turning points, the so-called pseudo arc-length continuation method is applied for snap-back problems. The pseudo arc-length continuation method uses an approximate arc-length parameter s instead of the "natural" parameter V. By adding an arc-length equation, the system is expressed as an augmented system.

$$A(z, V, s) = \begin{pmatrix} F(z(s), V(s)) \\ \|\dot{z}(s)\|^2 + |\dot{v}(s)|^2 - 1 \end{pmatrix} = 0 \qquad (1.65)$$

where $\dot{z}$ and $\dot{V}$ are derivatives of z and V with respect to the arc-length parameter s. Keller [23] has shown that this augmented system $A(z, V, s) = 0$ is nonsingular at turning points, even when the subsystem $F(z(s), V(s)) = 0$ is singular. This leads to the pseudo arc-length continuation algorithm.

Algorithm 2 (Pseudo Arc-Length Continuation) *1. Start at a known solution $(z(s_0), V(s_0)) \equiv (z_0, V_0)$.*

2. Compute the tangent vector $(\dot{z}(s_0), \dot{V}(s_0)) \equiv (\dot{z}_0, \dot{V}_0)$ by solving

$$F_z \dot{z}_0 + \dot{V}_0 F_V = 0$$

$$\|\dot{z}_0\|^2 + |\dot{V}_0|^2 = 1$$

3. Predict an initial approximation by

$$z^o = z_0 + (s_1 - s_0)\dot{z}_0$$

$$V^O = V_0 + (s_1 - s_0)\dot{V}_0$$

4. *Use* (z^O, V^O) *as an initial approximation in Newton's method for solving the augmented system* $A(z, V, s) = 0$.

5. *Use the solution* (z_1, V_1) *from step 4 as new* (z_0, V_0), *go to step 1.*

In our semiconductor simulation programs, we successfully pass the turning point at the holding voltage and obtain the current-voltage characteristic by this algorithm.

2. ANALYTICAL INVESTIGATIONS OF THE CHARACTER OF SOLUTIONS

This chapter provides a survey of some existing analytical results about the nature of solutions of the steady state semiconductor equations. We review some results about existence and uniqueness of solutions. We next present the regional approximation for a $p - n$ junction which is a basic unit in studying semiconductor devices. The chapter closes with a discussion of singular perturbation analysis which yields insight into the qualitative and quantitative structure of solutions. The knowledge of these characteristics of solutions is essential for successful numerical computation.

2.1 Existence and uniqueness of solutions

So far, the proofs of existence of solutions for the basic semiconductor device equations have not been achieved without restrictions. These restrictions are due basically to the recombination-generation rate and the carrier mobilities. The existence proof under the assumptions of constant mobilities and bounded recombination-generation rate has been given by Mock [28]. Seidman [17] has obtained similar results for the Shockley-Read-Hall recombination-generation rate. Jerome [18] has an existence analysis for considering Shockley-Read-Hall and Auger recombination-generation rate and its discrete analogue. Similar results have been obtained by

Markowich [13].

If we use the state variables (ψ, u, v) and insert the current relations into the continuity equations, the one-dimensional steady state basic semiconductor device equations under Ohmic contacts take the form:

$$\lambda^2 \frac{d^2\psi}{dx^2} = \delta^2 e^{\psi} u - \delta^2 e^{-\psi} v - C(x) \tag{2.1}$$

$$\frac{d}{dx}(\delta^2 \mu_n e^{\psi} \frac{du}{dx}) = R(x, \psi, u, v) \tag{2.2}$$

$$\frac{d}{dx}(\delta^2 \mu_p e^{-\psi} \frac{dv}{dx}) = R(x, \psi, u, v) \tag{2.3}$$

for $x \in [-1, 1]$. The scaled boundary conditions at Ohmic contacts are:

$$\psi(\pm 1) = \psi_{bi}(\pm 1) + V_{o\pm 1} \tag{2.4}$$

$$u(\pm 1) = e^{-V_o(\pm 1)} \tag{2.5}$$

$$v(\pm 1) = e^{V_o(\pm 1)} \tag{2.6}$$

where ψ_{bi} is the scaled built-in voltage.

The basic ideas of Jerome and Markowich for proving existence of solutions use the Gummel decoupling algorithm and the Schauder fixed point theorem. The Gummel decoupling algorithm is as follows.

Algorithm 3 (Gummel Decoupling Algorithm) *Starting with a given pair* $(u, v) =$ (u_0, v_0), *repeat the following steps:*

1. Solve Poisson equation :

$$\lambda^2 \frac{d^2\psi_k}{dx^2} = \delta^2 e^{\psi_k} u_{k-1} - \delta^2 e^{-\psi_k} v_{k-1} - C(x)$$

for $x \in [-1,1]$, *with boundary condition*

$$\psi(\pm1) = \psi_{bi}(\pm1) + V_{o\pm1}$$

for ψ_k.

2. *Solve the electron continuity equation:*

$$\frac{d}{dx}(\delta^2 \mu_n e^{\psi_k} \frac{du_k}{dx}) = R(x, \psi_k, u_k, v_{k-1})$$

for $x \in [-1,1]$, *with boundary condition*

$$u_k(\pm1) = e^{-V_o(\pm1)}$$

for u_k.

3. *Solve the hole continuity equation*

$$\frac{d}{dx}(\delta^2 \mu_p e^{-\psi_k} \frac{dv_k}{dx}) = R(x, \psi_k, u_k, v_k)$$

for $x \in [-1,1]$, *with boundary condition*

$$v_k(\pm1) = e^{V_o(\pm1)}$$

for v_k.

until accuracy is achieved.

The fixed point operator T is then constructed by

$$T(u_{k-1}, v_{k-1}) = (u_k, v_k).$$

A solution $(\psi, u, v) = (\psi(u,v), u, v)$ of the basic device equations is then a fixed point of the operator T:

$$T(u, v) = (u, v).$$

The techniques of this proof of existence are much more complex and difficult than the concept of fixed point argument. We state existence results in the one-dimensional case here without proofs. Details and proofs can be found in [18] [13].

To state the existence theorems, we need the assumptions:

(A1) the doping profile $C \in L^\infty[-1,1]$, denoting

$$C_m = ess \inf_{-1 \leq x \leq 1} C(x),$$

$$C_M = ess \sup_{-1 \leq x \leq 1} C(x);$$

(A2) the mobilities are constant,

(A3) the Shockley-Read-Hall and Auger recombination-generation rates satisfy

$$R = P(x, \psi, u, v)(uv - 1),$$

with

$$P(x, \psi, u, v) \geq 0.$$

Theorem 2.1 *Let the assumption (A1) hold and let* $(u,v) \in (L^\infty[-1,1])^2$ *satisfy* $0 < u_m \leq u(x) \leq u_M$, $0 < v_m \leq v(x) \leq v_M$ *a.e. in* $[-1,1]$. *Then the Poisson equation (2.1), (2.4) has a unique weak solution* $\psi \in H^1[-1,1] \cap L^\infty[-1,1]$, *which satisfies the estimate:*

$$\min\left(\min(\psi(-1), \psi(1)), \ln\left[\frac{C_m + \sqrt{C_m^2 + 4\delta^4 u_M v_m}}{2\delta^2 u_M}\right]\right) \leq \psi(x)$$

$$\leq \max\left(\max(\psi(-1), \psi(1)), \ln\left[\frac{C_M + \sqrt{C_M^2 + 4\delta^4 u_m v_M}}{2\delta^2 u_m}\right]\right) \quad a.e. \ in \ [-1,1].$$

Theorem 2.2 *Let the assumptions* $(A1)$, $(A2)$ *and* $(A3)$ *hold. Suppose that the boundary data satisfies* $|V_{0\pm1}| \leq V$. *Then the system of the basic semiconductor equations (2.1), (2.2), (2.3), (2.4), (2.5), (2.6) has a weak solution* $(\psi, u, v) \in$ $(H^1[-1,1] \cap L^{\infty}[-1,1])^3$, *which satisfies the estimate:*

$$e^{-V} \leq u(x) \leq e^{V} \quad a.e. \ in \ [-1,1]$$

$$e^{-V} \leq v(x) \leq e^{V} \quad a.e. \ in \ [-1,1]$$

$$\min\left(\min(\psi(-1), \psi(1)), \ \ln\left[\frac{C_m + \sqrt{C_m^2 + 4\delta^4}}{2\delta^2}\right] - V\right) \leq \psi(x)$$

$$\leq \max\left(\max(\psi(-1), \psi(1)), \ \ln\left[\frac{C_M + \sqrt{C_M^2 + 4\delta^4}}{2\delta^2}\right] + V\right) \quad a.e. \ in \ [-1,1].$$

The above existence theorem does not yield any assertion on uniqueness of solutions for the semiconductor equations. Mock [19] has shown that there exists a unique solution for small external applied voltage V_0. But in general, for arbitrary applied voltage the solution is not unique.

In fact, the four layer p-n-p-n thyristor has three steady state solutions under certain biasing. Rubinstein [20] has constructed three solution branches of the reduced equations obtained by setting $\lambda^2 = 0$ in the basic semiconductor device equations for a p-n-p-n thyristor with piecewise constant doping profile. He used the Newton-Kantorovich theorem to prove the existence of the multiple solutions under the assumptions of constant and equal carrier mobilitites and zero recombination-generation rate. Steinrück [21] used the singular perturbation approach and bifurcation theory to determine the structure of the current voltage characteristic for a thyristor with piecewise constant doping profile. He allows the recombination-generation to be the Auger or the Shockley-Read-Hall model in his S-shape current

voltage structure. Ward et al. in their recent paper [22] constructed the multiple steady state solutions for a multijunction semiconductor device with Shockley-Read-Hall recombination rate using a combination of asymptotic and numerical techniques.

2.2 The p-n diode at thermal equilibrium

If all external potentials applied to contacts of a device are zero at a given temperature, then the device is in thermal equilibrium. Device physics [4] tells us that, at thermal equilibrium, the drift current due to the electric field cancels the diffusion current due to the concentration gradient. Therefore, the electron current J_n, the hole current J_p, and the recombination-generation rate R vanish within the device.

Now we consider the current equations using the state variables (ψ, u, v) at thermal equilibrium,

$$J_n = \mu_n \delta^2 e^\psi \frac{du}{dx} = 0$$

$$J_p = \mu_p \delta^2 e^{-\psi} \frac{dv}{dx} = 0$$

for $x \in [-1, 1]$, with boundary conditions at Ohmic contact:

$$u(\pm 1) = v(\pm 1) = 1.$$

The solutions for these two boundary value problems are $u \equiv v \equiv 1$. This implies that there is no need to solve the full set of the basic semiconductor equations. We need solve only the Poisson equation to obtain the so-called equilibrium potential $\psi = \psi_e$. Then, the solution set $(\psi_e, 1, 1) \equiv (\psi, u, v)$ constitutes a solution of the basic semiconductor equations in thermal equilibrium.

At thermal equilibrium, Poisson's equation reduces to

$$\lambda^2 \frac{d^2\psi}{dx^2} = \delta^2 e^\psi - \delta^2 e^{-\psi} - C(x)$$

for $x \in [-1, 1]$, with Ohmic contact boundary conditions:

$$\psi(\pm 1) = \ln\left[\frac{C(\pm 1) + \sqrt{C^2(\pm 1) + 4\delta^4}}{2\delta^2}\right].$$

Note that the boundary values are just the built-in voltages $\psi_{bi}(\pm 1)$. The regional approximation method for solving this problem works by separating the domain into three regions. It is illustrated in Figure 2.1 for the doping profile in an abrupt junction.

On physical grounds, charge neutrality (i.e., $\delta^2 e^\psi - \delta^2 e^{-\psi} - C(x) = 0$) is assumed in the two bulk regions. The free carrier densities n and p are assumed to be zero in the depletion region, and the electric field $E = -\frac{d\psi}{dx}$ vanishes at the end points x_p, x_n of the depletion region.

Under these physical assumptions, the Poisson equation has the regional approximation solution (see [4] for calculation) at thermal equilibrium:

$$\psi(x) = \psi(-1) = \psi_{bi}(-1), \quad \text{for } -1 \leq x < -x_p,$$

$$\psi(x) = \psi(1) = \psi_{bi}(1), \quad \text{for } x_n < x \leq 1,$$

$$\psi(x) = \psi_{bi}(-1) + \frac{C_p}{2}\left(\frac{x+x_p}{\lambda}\right)^2, \quad \text{for } -x_p \leq x < x_j,$$

$$\psi(x) = \psi(x_n) - \frac{C_n}{2}\left(\frac{x-x_n}{\lambda}\right)^2, \quad \text{for } x_j < x \leq x_n.$$

To make a $C^1[-1, 1]$ solution, the junction width $w = x_p + x_n$ is determined to be

$$w = \lambda\sqrt{2\psi(1) - \psi(-1)}\frac{1}{C_n + C_p}\sqrt{\frac{C_n + C_p}{C_n C_p}}.$$
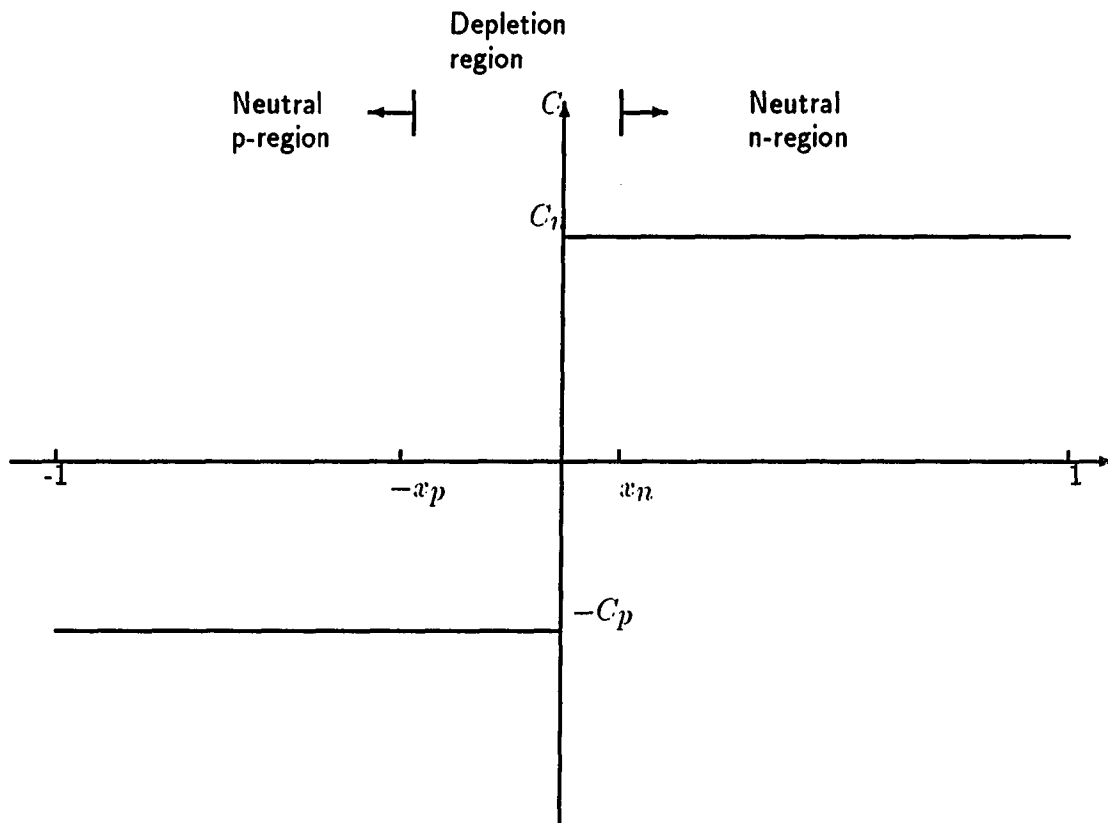
Figure 2.1:   An abrupt junction

This regional solution profile exhibits the junction layer character. This means that the electrostatic potential ψ varies strongly in the thin layer of width $O(\lambda)$, but slowly in the large regions far from the junction.

2.3 Singular perturbation analysis

In the preceding section we have considered the regional approximation method for p-n diodes. This analysis shows that the electrostatic potential varies greatly in the thin junction layer at thermal equilibrium. The order of magnitude of this layer is $O(\lambda)$. We suspect that the concentrations of electrons and holes have similar layer-type behavior at arbitrary biasing. The system of the scaled basic semiconductor equations is a singularly perturbed boundary value problem. In this section we summarize the singular perturbation approach of Markowich [24] to analyze the structure of solutions. This information about solution structure can be used to assess the discretization of the equations, design of the mesh and selection of iterative schemes.

To simplify the analysis, the mobilities μ_n, μ_p are assumed to be constant and the scaled doping profile C has a jump discontinuity at junction x_j, that is

$$\lim_{x \to x_j^+} C(x) \neq \lim_{x \to x_j^-} C(x).$$

At first, we try to solve the singularly perturbed semiconductor equations by a regular expansion in λ.

$$\psi(x,\lambda) \sim \bar{\psi}(x) + \sum_{i=1}^{\infty} \lambda_i \bar{\psi}_i(x),$$

$$n(x,\lambda) \sim \bar{n}(x) + \sum_{i=1}^{\infty} \lambda_i \bar{n}_i(x),$$

$$p(x,\lambda) \sim \bar{p}(x) + \sum_{i=1}^{\infty} \lambda_i \bar{p}_i(x),$$

$$J_n(x,\lambda) \sim \bar{J}_n(x) + \sum_{i=1}^{\infty} \lambda_i \bar{J}_{n_i}(x),$$

$$J_p(x,\lambda) \sim \bar{J}_p(x) + \sum_{i=1}^{\infty} \lambda_i \bar{J}_{p_i}(x).$$

Inserting these expansions into the equations and neglecting $O(\lambda)$ terms, we obtain the so-called reduced equations.

$$0 = \bar{n} - \bar{p} - C(x), \tag{2.7}$$

$$\frac{d\bar{J}_n}{dx} = \bar{R}(x, \bar{\psi}, \bar{n}, \bar{p}), \tag{2.8}$$

$$\frac{d\bar{J}_p}{dx} = -\bar{R}(x, \bar{\psi}, \bar{n}, \bar{p}), \tag{2.9}$$

$$\bar{J}_n = \mu_n \left(\frac{d\bar{n}}{dx} - \bar{n}\frac{d\bar{\psi}}{dx} \right), \tag{2.10}$$

$$\bar{J}_p = -\mu_p \left(\frac{d\bar{p}}{dx} + \bar{p}\frac{d\bar{\psi}}{dx} \right). \tag{2.11}$$

These reduced equations are called the charge-neutral approximation in device physics since space charge $n - p - C(x) = 0$.

The reduced solutions $\bar{\psi}, \bar{n}, \bar{p}$ are discontinuous since C has a jump discontinuity at junction x_j. They must be supplemented by the junction correction terms to make a continuous approximation for the full semiconductor equations. After the junction correction terms are included, we have the asymptotic expansion:

$$f(x,\lambda) \sim \bar{f}(x) + \hat{f}\left(\frac{x - x_j}{\lambda}\right) + \cdots, \tag{2.12}$$

where $f = (\psi, n, p, J_n, J_p)^T$ and the dots denote a power series in λ starting with the $O(\lambda)$ terms. The junction correction terms $\hat{\psi}, \hat{n}, \hat{p}, \hat{J}_p, \hat{J}_p$ are defined for $\sigma = \frac{x - x_j}{\lambda}$. They are supposed to decay exponentially to zero as $\sigma \to \infty$.

Now we insert the expansion (2.12) into the full basic semiconductor equations and consider the leading terms in λ. Evaluating the equations close to the junction x_j for $\sigma > 0$, we have right layer equations:

$$\frac{d^2\hat{\psi}}{d\sigma^2} = \hat{n} - \hat{p},$$

(2.13)

$$\frac{d\hat{n}}{d\sigma} = (\hat{n} + \bar{n}(x_j+))\frac{d\hat{\psi}}{d\sigma},$$

(2.14)

$$\frac{d\hat{p}}{d\sigma} = (\hat{p} + \bar{p}(x_j+))\frac{d\hat{\psi}}{d\sigma},$$

(2.15)

$$\frac{d\hat{J}_n}{d\sigma} = 0,$$

(2.16)

$$\frac{d\hat{J}_p}{d\sigma} = 0,$$

(2.17)

with

$$f(x_j\pm) = \lim_{x \to x_j\pm} f(x).$$

For $\sigma < 0$, replacing $\bar{n}(x_j+)$, $\bar{p}(x_j+)$ by $\bar{n}(x_j-)$, $\bar{p}(x_j-)$, we obtain the left layer problem:

$$\frac{d^2\hat{\psi}}{d\sigma^2} = \hat{n} - \hat{p},$$

(2.18)

$$\frac{d\hat{n}}{d\sigma} = (\hat{n} + \bar{n}(x_j-))\frac{d\hat{\psi}}{d\sigma},$$

(2.19)

$$\frac{d\hat{p}}{d\sigma} = (\hat{p} + \bar{p}(x_j-))\frac{d\hat{\psi}}{d\sigma},$$

(2.20)

$$\frac{d\hat{J}_n}{d\sigma} = 0,$$

(2.21)

$$\frac{d\hat{J}_p}{d\sigma} = 0.$$

(2.22)

We want the approximation

$$f(x, \lambda) \approx \bar{f}(x) + \hat{f}(\frac{x - x_j}{\lambda}) \qquad (2.23)$$

of the full semiconductor equations to satisfy the matching condition, that is,

$$\lim_{\lambda \to 0\pm} f(x, \lambda) = \lim_{\lambda \to 0\pm} (\bar{f}(x) + \hat{f}(\frac{x - x_j}{\lambda})) = \bar{f}(x).$$

This means that the asymptotic approximation $f(x, \lambda)$ (2.23) approaches the reduced solutions when λ is close to zero. The layer equations therefore have the asymptotic boundary conditions:

$$\hat{\psi}(\pm\infty) = \hat{n}(\pm\infty) = \hat{p}(\pm\infty) = \hat{J}_n(\pm\infty) = \hat{J}_p(\pm\infty) = 0. \qquad (2.24)$$

These asymptotic boundary conditions assure that the junction correction terms decay to zero as the variable σ approaches infinity.

By requiring continuity of the approximation (2.23), the layer equations also have the interface conditions:

$$\bar{\psi}(x_j+) + \hat{\psi}(0+) = \bar{\psi}(x_j-) + \hat{\psi}(0-), \qquad (2.25)$$

$$\bar{n}(x_j+) + \hat{n}(0+) = \bar{n}(x_j-) + \hat{n}(0-), \qquad (2.26)$$

$$\bar{p}(x_j+) + \hat{p}(0+) = \bar{p}(x_j-) + \hat{p}(0-), \qquad (2.27)$$

$$\bar{J}_n(x_j+) + \hat{J}_n(0+) = \bar{J}_n(x_j-) + \hat{J}_n(0-), \qquad (2.28)$$

$$\bar{J}_p(x_j+) + \hat{J}_p(0+) = \bar{J}_p(x_j-) + \hat{J}_p(0-), \qquad (2.29)$$

$$\hat{\psi}(0+) = \hat{\psi}(0-). \qquad (2.30)$$

From the equations (2.16), (2.17), (2.21), (2.22) and the asymptotic boundary conditions (2.24), we have

$$\hat{J}_n = \hat{J}_p = 0.$$

This means that electron and hole current densities have no junction layers. Integrating the equations (2.14), (2.15), (2.19), (2.20) and using the asymptotic boundary conditions (2.24), we obtain the carrier density correction terms $\hat{n}(\sigma)$, $\hat{p}(\sigma)$ in terms of $\hat{\psi}(\sigma)$ as follows:

$$\hat{n}(\sigma) = \bar{n}(x_j+)(e^{\hat{\psi}(\sigma)} - 1), \quad \text{for } \sigma > 0, \tag{2.31}$$

$$\hat{p}(\sigma) = \bar{p}(x_j+)(e^{-\hat{\psi}(\sigma)} - 1), \quad \text{for } \sigma > 0, \tag{2.32}$$

$$\hat{n}(\sigma) = \bar{n}(x_j-)(e^{\hat{\psi}(\sigma)} - 1), \quad \text{for } \sigma < 0, \tag{2.33}$$

$$\hat{p}(\sigma) = \bar{p}(x_j-)(e^{-\hat{\psi}(\sigma)} - 1), \quad \text{for } \sigma < 0. \tag{2.34}$$

The internal layer problem is derived by inserting the above equations (2.31)—(2.34) into equation (2.13).

$$\frac{d^2\hat{\psi}}{d\sigma^2} = \bar{n}(x_j+)e^{\hat{\psi}(\sigma)} - \bar{p}(x_j+)e^{-\hat{\psi}(\sigma)} - C(x_j^+), \quad \text{for}\sigma > 0, \tag{2.35}$$

$$\frac{d^2\hat{\psi}}{d\sigma^2} = \bar{n}(x_j-)e^{\hat{\psi}(\sigma)} - \bar{p}(x_j-)e^{-\hat{\psi}(\sigma)} - C(x_j^-), \quad \text{for}\sigma > 0, \tag{2.36}$$

with asymptotic boundary conditions

$$\hat{\psi}(\pm\infty) = 0, \tag{2.37}$$

and the interface conditions

$$\bar{\psi}(x_j+) + \hat{\psi}(0+) = \bar{\psi}(x_j-) + \hat{\psi}(0-), \tag{2.38}$$

$$\hat{\psi}(0+) = \hat{\psi}(0-). \tag{2.39}$$

Markowich [24] showed the intenal layer problem has a unique exponentially decaying solution $\hat{\psi}$. The ith order derivatives satisfy

$$\left|\frac{d^i\hat{\psi}(\sigma)}{d\sigma^i}\right| \leq K_1 e^{-K_2|\sigma|}$$

where $K_1, K_2 > 0$ only depend on i. The carrier density correction terms $\hat{n}$, $\hat{p}$ also decay exponentially as $\sigma \rightarrow \pm\infty$.

The equations (2.31)— (2.34) and the equations (2.25) — (2.27) give the interface conditions for the reduced carrier densities:

$$\bar{n}(x_j+) = \bar{n}(x_j-)e^{\bar{\psi}(x_j+)-\bar{\psi}(x_j-)}, \tag{2.40}$$

$$\bar{p}(x_j+) = \bar{p}(x_j-)e^{\bar{\psi}(x_j-)-\bar{\psi}(x_j+)}. \tag{2.41}$$

In physics language, (2.40), (2.41) represent the continuity of quasi-Fermi levels. The carrier densities have the layer jumps which depend exponentially on the voltage drop accross the junction. Since the electron and hole current densities have no junction layer (i.e. $\hat{J}_n = \hat{J}_p = 0$, we have the interface conditions of current densities for reduced equations:

$$\bar{J}_n(x_j+) = \bar{J}_n(x_j-), \tag{2.42}$$

$$\bar{J}_p(x_j+) = \bar{J}_p(x_j-). \tag{2.43}$$

Recent work of Ward, Reynal and Odeh [22] indicates that these conditions do not hold universally. The reduced equations (2.7)— (2.11) and the interface conditions (2.40)— (2.43) and the same boundary conditions as the full equations constitute the reduced boundary value problem.

In the same paper [24], Markowich showed that the reduced problem has a weak solution $(\bar{\psi}, \bar{n}, \bar{p})$ which approximates the solution of the full equations up to $O(\lambda)$ outside the layer region, and the width of layer is $O(\lambda|\ln \lambda|)$. He also showed that the asymptotic expansion (2.12) is valid at thermal equilibrium and satisfies

$$\|\psi(x,\lambda) - \bar{\psi}(x) - \hat{\psi}(\frac{x - x_j}{\lambda})\|_\infty = O(\lambda).$$

3. DISCRETIZATION AND MESH DESIGNING

In this chapter we shall derive the discretization of the basic semiconductor device equations for executing the numerical computations. We use a finite difference scheme to transform the continuous semiconductor equations into discrete forms in which the solutions are approximated at a finite number of mesh points. This finite difference discretization is based on the box method as discussed in chapter 1.

Designing the mesh plays a important role in the steady state semiconductor simulation program to achieve a specified accuracy with as few mesh points as possible. The normal uniform mesh is unsuitable for the discretization of the basic semiconductor equations because of the layer-type solutions, by the singular perturbation analysis. Our mesh construction is based on the solution character, and on the idea of equidistributing the local truncation error.

3.1 Discretization of the basic semiconductor device equations

In order to execute the numerical computation, the basic semiconductor device equations are solved in the discrete form in which the unknown variables are defined at mesh points. To denote the mesh points, we let h_j be the size of the jth interval

in the non-uniform mesh and we use the following notations:

$$x_{j+1} = x_j + h_j, \qquad j = 0, 1, \cdots, N,$$

$$x_{j+\frac{1}{2}} = \tfrac{1}{2}(x_j + x_{j+1}), \quad j = 0, 1, \cdots, N,$$

$$x_{j-\frac{1}{2}} = \tfrac{1}{2}(x_j + x_{j-1}), \quad j = 1, 2, \cdots, N+1,$$

$$\bar{x}_j = \tfrac{1}{2}(x_{j+\frac{1}{2}} + x_{j-\frac{1}{2}}), \quad j = 1, 2, \cdots, N.$$

Here, x_0 and x_{N+1} correspond to the end points $x_0 = -1$ and $x_{N+1} = 1$. The point $\bar{x}_j$ is the midpoint of the two midpoints $x_{j+\frac{1}{2}}$ and $x_{j-\frac{1}{2}}$. The maximum mesh size h is defined by

$$h = \max_{0 \le j \le N} h_j.$$

For the evaluation of functions at mesh points , we use the following notations:

$$f(x_j) = f_j, \quad f(x_{j+\frac{1}{2}}) = f_{j+\frac{1}{2}}, \quad f(x_{j-\frac{1}{2}}) = f_{j-\frac{1}{2}}, \quad f(\bar{x}_j) = f_{\bar{j}}.$$

3.1.1 Discretization of Poisson equation

Now we discretize the Poisson equation of the semiconductor problem. The scaled form is

$$\lambda^2 \frac{d^2\psi}{dx^2} = n - p - C(x). \tag{3.1}$$

Integrating the Poisson equation (3.1) over the interval $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$, we have

$$\lambda^2 \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \frac{d^2\psi(x)}{dx^2} \, dx = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} n - p - C(x) \, dx. \tag{3.2}$$

By applying the one-dimensonal box method, equation (3.2) can be approximated as:

$$\lambda^2 \left((\frac{d\psi}{dx})_{j+\frac{1}{2}} - (\frac{d\psi}{dx})_{j-\frac{1}{2}} \right) \approx \frac{1}{2}(h_{j-1} + h_j)(n_j^- - p_j^- - C_j^-). \tag{3.3}$$

We evaluate $(\frac{d\psi}{dx})_{j+\frac{1}{2}}$ and $(\frac{d\psi}{dx})_{j-\frac{1}{2}}$ by central differences, that is,

$$(\frac{d\psi}{dx})_{j+\frac{1}{2}} \approx \frac{\psi_{j+1} - \psi_j}{h_j}, \tag{3.4}$$

$$(\frac{d\psi}{dx})_{j-\frac{1}{2}} \approx \frac{\psi_j - \psi_{j-1}}{h_{j-1}}. \tag{3.5}$$

Substituting these approximations (3.4), (3.5) into (3.3), we obtain the standard three point formula for discretization of the Poisson equation:

$$\lambda^2 \frac{2}{h_{j-1} + h_j} \left(\frac{\psi_{j+1} - \psi_j}{h_j} - \frac{\psi_j - \psi_{j-1}}{h_{j-1}} \right) - (n_j^- - p_j^- - C_j^-) = 0. \tag{3.6}$$

The space charge $n(x) - p(x) - C(x)$ is approximated by the midpoint of the midpoints scheme [25].

3.1.2 Discretization of continuity equations

Now we construct the discretization of the continuity equations. We deal only with the continuity equation for electrons:

$$\frac{dJ_n}{dx} = R, \tag{3.7}$$

with electron current density:

$$J_n = \mu_n(\frac{dn}{dx} - n\frac{d\psi}{dx}). \tag{3.8}$$

The treatment of the continuity equation for holes is similar.

By applying the box method, as in the Poisson equation, to the electron continuity equation, we obtain

$$\left(Jn_{j+\frac{1}{2}} - Jn_{j-\frac{1}{2}}\right) \approx \frac{1}{2}(h_{j-1} + h_j)R_j. \tag{3.9}$$

The evaluation of $Jn_{j+\frac{1}{2}}$ and $Jn_{j-\frac{1}{2}}$ is interesting and has to be treated very carefully. As a simple illustration, we consider the straightforward central difference scheme as applied in the Poisson equation. The discrete form for electron current equation (3.8) by the central difference scheme is

$$Jn_{j+\frac{1}{2}} = \mu n_{j+\frac{1}{2}} \left(\frac{n_{j+1} - n_j}{h_j} - \frac{n_{j+1} - n_j}{2}\frac{\psi_{j+1} - \psi_j}{h_j}\right). \tag{3.10}$$

This scheme is not regarded as a good choice for evaluating $Jn_{j+\frac{1}{2}}$, since it requires a very fine mesh for computing accurately the electrostatic potential ψ.

To simply explain the reason, consider the thermal equilibrium condition, i.e. $Jn = 0$. The discretization (3.10) of the electron current equation for $Jn = 0$ yields

$$\psi_{j+1} - \psi_j = \frac{n_{j+1} - n_j}{n_{j+1} + n_j} = \frac{e^{\psi_{j+1} - \psi_j} - 1}{e^{\psi_{j+1} - \psi_j} + 1},$$

where $n = \delta^2 e^{\psi}$ at thermal equilibrium. This shows that the variation of the electrostatic potential ψ must be small on each mesh interval, that is,

$$\max_j | \psi_{j+1} - \psi_j | \ll 1.$$

This condition requires a very small mesh-size inside the junction layer since the electrostatic potential varies greatly in the thin layer as we know from the singular perturbation analysis. Therefore, in practical numerical computation, this scheme is not feasible for discretizing the current equations.

In order to have suitable evaluations of $Jn_{j+\frac{1}{2}}$ and $Jn_{j-\frac{1}{2}}$, we may simplify the electron current equation by using the state variables (ψ, u, v):

$$J_n = \delta^2 \mu_n e^\psi \frac{du}{dx}. \tag{3.11}$$

The electron mobility μ_n at the midpoint $x_{j+\frac{1}{2}}$ can be evaluated by the average, i.e.

$$\mu_{n_{j+\frac{1}{2}}} = \frac{1}{2}(\mu_{n_j} + \mu_{n_{j+1}}).$$

We approximate the derivative of the unknown function u by the difference quotient:

$$(\frac{du}{dx})_{j+\frac{1}{2}} = \frac{(u_{j+1} - u_j)}{h_j}.$$

One way to approximate e^ψ at $x_{j+\frac{1}{2}}$ is

$$(e^\psi)_{j+\frac{1}{2}} \approx e^{\frac{\psi_j + \psi_{j+1}}{2}}, \tag{3.12}$$

but this approximation will yield the same problem as the straightforward difference scheme. Mock [28] claims that this approximation (3.12) will lead to inaccuracies.

An alternative estimation for e^ψ at $x_{j+\frac{1}{2}}$ can be given by calculating the average of the weighted function $e^\psi(\frac{de^{-\psi}}{dx})$. Let us consider the integration of this weighted function over the interval $[x_j, x_{j+1}]$:

$$\int_{x_j}^{x_{j+1}} e^\psi \frac{de^{-\psi}}{dx}\, dx = -\int_{x_j}^{x_{j+1}} \frac{d\psi}{dx}\, dx = -(\psi_{j+1} - \psi_j). \tag{3.13}$$

By employing the mean value theorem in the integral on the left hand side of the above equation (3.13), we have

$$\int_{x_j}^{x_{j+1}} e^\psi \frac{de^{-\psi}}{dx}\, dx = e^{\psi(\xi)} \int_{x_j}^{x_{j+1}} \frac{de^{-\psi}}{dx}\, dx = e^{\psi(\xi)}(e^{-\psi_{j+1}} - e^{-\psi_j}), \tag{3.14}$$

with $\xi \in [x_j, x_{j+1}]$.

If we take ξ to be the midpoint $x_{j+\frac{1}{2}}$, we obtain the approximation for $(e^\psi)_{j+\frac{1}{2}}$:

$$(e^\psi)_{j+\frac{1}{2}} \approx \frac{\psi_j - \psi_{j+1}}{e^{-\psi_{j+1}} - e^{-\psi_j}} = e^{\psi_{j+1}} B(\psi_{j+1} - \psi_j), \qquad (3.15)$$

where $B(x)$ is the Bernoulli function defined by

$$B(x) = \frac{x}{e^x - 1}.$$

The equation (3.15) is proposed by Bank et al. [26] for approximating e^ψ at the midpoint. Substituting the expression (3.15) into (3.11), we obtain the electron current density J_n at $x_{j+\frac{1}{2}}$:

$$J_{n_{j+\frac{1}{2}}} = \delta^2 \mu_{n_{j+\frac{1}{2}}} e^{\psi_{j+1}} B(\psi_{j+1} - \psi_j) \frac{u_{j+1} - u_j}{h_j}. \qquad (3.16)$$

The analogous evaluation for J_n at $x_{j-\frac{1}{2}}$ is

$$J_{n_{j-\frac{1}{2}}} = \delta^2 \mu_{n_{j-\frac{1}{2}}} e^{\psi_j} B(\psi_j - \psi_{j-1}) \frac{u_j - u_{j-1}}{h_{j-1}}. \qquad (3.17)$$

Thus, for the variables (ψ, u, v), the discretization of the electron continuity equation can be obtained by inserting the expressions (3.16), (3.17) into (3.9). This discretization of the electron continuity equation is

$$\frac{2\delta^2}{h_{j-1} + h_j} \left\{ \mu_{n_{j+\frac{1}{2}}} e^{\psi_{j+1}} B(\psi_{j+1} - \psi_j) \frac{u_{j+1} - u_j}{h_j} - \right.$$

$$\left. \mu_{n_{j-\frac{1}{2}}} e^{\psi_j} B(\psi_j - \psi_{j-1}) \frac{u_j - u_{j-1}}{h_{j-1}} \right\} - R_j = 0. \qquad (3.18)$$

The discretization using variables (ψ, n, p) is easily obtained by setting $n = \delta^2 e^{\psi} u$, that is,

$$\frac{2}{h_{j-1} + h_j} \left\{ \frac{\mu n_{j+\frac{1}{2}}}{h_j} B(\psi_{j+1} - \psi_j) n_{j+1} \right.$$

$$- \left[\frac{\mu n_{j+\frac{1}{2}}}{h_j} B(\psi_j - \psi_{j+1}) + \frac{\mu n_{j-\frac{1}{2}}}{h_{j-1}} B(\psi_j - \psi_{j-1}) \right] n_j$$

$$\left. + \frac{\mu n_{j-\frac{1}{2}}}{h_{j-1}} B(\psi_{j-1} - \psi_j) n_{j-1} \right\} - R_j = 0. \tag{3.19}$$

The above discrete equation (3.19) was suggested first by Scharfetter and Gummel [8] motivated by physical consideration. It is therefore called the Scharfetter-Gummel scheme.

An alternative derivation for the Scharfetter-Gummel discretization is that the electron current equation may be regard as determining the distribution of electron density n over a mesh interval $[x_j, x_{j+1}]$. We assume that the electric field $E = -\frac{d\psi}{dx}$, electron mobility μn and current density $J n$ are constant within the mesh interval $[x_j, x_{j+1}]$. Now we solve the electron current equation

$$J n = \mu n \left(\frac{dn}{dx} - n \frac{d\psi}{dx} \right) \tag{3.20}$$

over $[x_j, x_{j+1}]$ with the boundary condition

$$n(x_j) = n_j, \quad n(x_{j+1}) = n_{j+1}. \tag{3.21}$$

This first order differential equation can be exactly solved in the explicit form:

$$n(x) = \left\{ e^{\psi(x)-\psi_j} - \frac{e^{\psi_{j+1}-\psi_j}(1 - e^{\psi(x)-\psi_j})}{1 - e^{\psi_{j+1}-\psi_j}} \right\} n_j$$

$$+ \left\{ \frac{1 - e^{\psi(x) - \psi_j}}{1 - e^{\psi_{j+1} - \psi_j}} \right\} n_{j+1} \tag{3.22}$$

for $x \in [x_j, x_{j+1}]$.

The electrostatic potential ψ is assumed to be a piecewise linear function between the mesh points. Thus

$$\psi(x) - \psi_j = (x - x_j) \frac{\psi_{j+1} - \psi_j}{h_j} \tag{3.23}$$

for $x \in [x_j, x_{j+1}]$. Substituting (3.23) into (3.22), we obtain the so-called exponential fitting for electron density n:

$$n(x) = (1 - g_j(x; \psi)) n_j + g_j(x; \psi) n_{j+1} \tag{3.24}$$

for $x \in [x_j, x_{j+1}]$, where the growth function $g_j(x; \psi)$ is defined by

$$g_j(x; \psi) = \frac{1 - e^{\frac{\psi_{j+1} - \psi_j}{h_j}(x - x_j)}}{1 - e^{\psi_{j+1} - \psi_j}} \tag{3.25}$$

for $x \in [x_j, x_{j+1}]$.

Therefore, we can evaluate the electron density n and the derivative $\frac{dn}{dx}$ at $x_{j+\frac{1}{2}}$ and $x_{j+\frac{1}{2}}$ from the expression (3.24) of the electron density n. Substituting these evaluations into the electron current equation (3.8) with central difference for the derivative of electrostatic potential ψ, we have the evaluation for electron current density J_n at $x_{j+\frac{1}{2}}$ and $x_{j-\frac{1}{2}}$:

$$J_{n_{j+\frac{1}{2}}} = \frac{\mu n_{j+1\frac{1}{2}}}{h_j} \{ B(\psi_{j+1} - \psi_j) n_{j+1} - B(\psi_j - \psi_{j+1}) n_j \} \tag{3.26}$$

$$J_{n_{j-\frac{1}{2}}} = \frac{\mu n_{j-1\frac{1}{2}}}{h_{j-1}} \{ B(\psi_j - \psi_{j-1}) n_j - B(\psi_{j-1} - \psi_j) n_{j-1} \} \tag{3.27}$$

We again obtain the Scharfetter-Gummel discretization of the electron continuity equation by inserting the expressions (3.26), (3.27) into (3.9). A completely analogous discussion holds for the hole continuity equation.

We now summarize the discrete equations as follows.

The discrete Poisson equation f_{ψ_j} is:

$$f_{\psi_j} = \lambda^2 \frac{2}{h_{j-1} + h_j} \left(\frac{\psi_{j-1}}{h_{j-1}} - (\frac{1}{h_{j-1}} + \frac{1}{h_j})\psi_j + \frac{\psi_j}{h_j} \right)$$

$$- (n_j - p_j - C_j) = 0. \tag{3.28}$$

The discrete electron continuity equation f_{n_j} is:

$$f_{n_j} = \frac{2}{h_{j-1} + h_j} \left\{ \frac{\mu n_{j-\frac{1}{2}}}{h_{j-1}} B(\psi_{j-1} - \psi_j)n_{j-1} \right.$$

$$- \left[\frac{\mu n_{j-\frac{1}{2}}}{h_{j-1}} B(\psi_j - \psi_{j-1}) + \frac{\mu n_{j+\frac{1}{2}}}{h_j} B(\psi_j - \psi_{j+1}) \right] n_j$$

$$\left. + \frac{\mu n_{j+\frac{1}{2}}}{h_j} B(\psi_{j+1} - \psi_j)n_{j+1} \right\} - R_j = 0. \tag{3.29}$$

The discrete hole continuity equation f_{p_j} is:

$$f_{p_j} = \frac{2}{h_{j-1} + h_j} \left\{ \frac{\mu p_{j-\frac{1}{2}}}{h_{j-1}} B(\psi_j - \psi_{j-1})p_{j-1} \right.$$

$$- \left[\frac{\mu p_{j-\frac{1}{2}}}{h_{j-1}} B(\psi_{j-1} - \psi_j) + \frac{\mu p_{j+\frac{1}{2}}}{h_j} B(\psi_{j+1} - \psi_j) \right] p_j$$

$$\left. + \frac{\mu p_{j+\frac{1}{2}}}{h_j} B(\psi_j - \psi_{j+1})p_{j+1} \right\} - R_j = 0. \tag{3.30}$$

The discrete electron current equation J_{n_j} is:

$$J_{n_j} = \mu n_j \left\{ B(\psi_{j+1} - \psi_j) \frac{n_{j+1} - n_j}{h_j} - n_j \frac{\psi_{j+1} - \psi_j}{h_j} \right\}. \qquad (3.31)$$

The discrete hole current equation J_{p_j} is:

$$J_{p_j} = -\mu p_j \left\{ B(\psi_j - \psi_{j+1}) \frac{p_{j+1} - p_j}{h_j} + p_j \frac{\psi_{j+1} - \psi_j}{h_j} \right\}. \qquad (3.32)$$

3.2 Analysis for the discretizations

The Scharfetter-Gummel discretization method is an exponentially fitted scheme for the current equations. In this subsection we estimate the local truncation error and study the convergence of the discretization scheme.

We first consider the Poisson equation in terms of the differential operator:

$$L \,\psi(x) = \lambda^2 \, \frac{d^2\psi(x)}{dx^2} = f(x), \quad -1 \le x \le 1, \qquad (3.33)$$

with $f(x) = n(x) - p(x) - C(x)$ and $\psi(x)$ satisfying the Dirichlet boundary conditions. The corresponding difference operator of the three point discretization scheme is written as:

$$L_h \, \Psi_j = \lambda^2 \frac{2}{h_{j-1} + h_j} \left(\frac{\Psi_{j+1} - \Psi_j}{h_j} - \frac{\Psi_j - \Psi_{j-1}}{h_{j-1}} \right) = f_j, \quad j = 1, 2, \cdots, N. \qquad (3.34)$$

with $\Psi_0 = \Psi(-1)$, $\Psi_{N+1} = \Psi(1)$ and $f_j = n_j - p_j - C_j$.

The local truncation error $T_j[\psi]$ is defined by

$$T_j[\psi] = L_h \psi(x_j) - L\psi(x_j), \quad j = 1, 2, \cdots, N. \qquad (3.35)$$

Thus, the local truncation error estimates the amount by which the solution of the differential equation fail to satisfy the discretization of the differential equation. By simple calculation and Taylor's expansion up to the fourth term, we obtain

$$T_j[\psi] = \lambda^2 \left\{ \frac{1}{3}(h_j - h_{j+1})\frac{d^3\psi(x_j)}{dx^3} + \frac{1}{12}(h_j^2 - h_j h_{j+1} + h_{j-1}^2)\frac{d^4\psi(x_j)}{dx^4} \right.$$

$$\left. +O\left(\frac{h_j^4 + h_{j-1}^4}{h_j + h_{j-1}}\right) \right\}. \tag{3.36}$$

This shows that the standard three point discretization scheme is consistent with order 1. This means that

$$\max_{1 \le j \le N} |T_j[\psi]| = O(h)$$

holds for a non-uniform mesh. For a uniform or almost uniform mesh, that is

$$h_j = h_{j-1}(1 + O(h_{j-1})),$$

the local truncation error can be improved to second order accuracy. That is

$$\max_{1 \le j \le N} |T_j[\psi]| = O(h^2).$$

We now consider the global error estimation. Let the global error

$$e_j = \psi(x_j) - \Psi_j, \quad j = 1, 2, \cdots, N.$$

Then we have

$$L_h e_j = L_h \psi(x_j) - L\psi(x_j) = T_j[\psi], \quad j = 1, 2, \cdots, N. \tag{3.37}$$

This system of differential equations (3.37) can be shown to be stable in the maximum norm. This means that there exist a constant k [27] such that

$$\| L_h^{-1} \| \le k. \tag{3.38}$$

The three point difference scheme for the non-uniform mesh is then convergent because of the stability and consistency of this scheme. From (3.37), (3.38), the global error satisfies

$$\max_{1\leq j\leq N} |\psi(x_j) - \Psi_j| \leq k \max_{1\leq j\leq N} |T_j[\psi]| \tag{3.39}$$

The Scharfetter-Gummel discretization method is an exponentialy fitted scheme for the continuity equations. The exponential factor

$$B(\psi_{j+1} - \psi_j) \rightarrow 1 \quad as \quad \psi_{j+1} \rightarrow \psi_j.$$

The electrostatic potential varies slowly in the regions far away from the junction by singular perturbation analysis. Therefore, this scheme is close to the three point scheme in those regions. Some exponentially fitted methods have been proven to be uniformly convergent for boundary layer problems by Doolan, Miller and Schilders [28]. The Scharfetter-Gummel scheme is not completely understood from mathematical analysis; although it has been used well for around twenty years in numerical semiconductor device computation.

We use the same notations $T_j[J_n]$, $T_j[n]$ as in the Poisson equation to denote the local truncation errors of the Scharfetter-Gummel discretization for the electron continuity equation and the current equation, respectively. Taylor's expansion gives

$$T_j[J_n] = \frac{1}{4}(h_j - h_{j+1})\frac{d^2 J_n(x_j)}{dx^2} + \frac{1}{24}(h_j^2 - h_j h_{j+1} + h_{j-1}^2)\frac{d^3 J_n(x_j)}{dx^3}$$

$$+ O\left(\frac{h_j^4 + h_{j-1}^4}{h_j + h_{j-1}}\right) \tag{3.10}$$

for the electron continuity equation and

$$T_j[n] = \mu_n \frac{h_j}{2}\left\{\frac{d^2 n(x_j)}{dx^2} - \frac{d\psi(x_j)}{dx}\frac{dn(x_j)}{dx} - \frac{d^2\psi(x_j)}{dx^2}n(x_j)\right\} + O(h_j^2) \tag{3.11}$$

for the electron current equation, provided ψ_{j+1} is close to ψ_j. Thus, this scheme yields first order accuracy and we expect the convergence rate $O(h)$ if it is stable. The global error estimates have been investigated [13] by solving directly the combined current and continuity equations and their discretization assuming the recombination-generation rate $R = 0$. The combined current continuity equations with $R = 0$ are:

$$\frac{dJ_n(x)}{dx} = 0,$$

$$J_n = \mu n \left(\frac{dn}{dx} - n \frac{d\psi}{dx} \right) \tag{3.42}$$

for $-1 \le x \le 1$, with boundary conditions $n(-1) = n_0$ and $n(1) = n_{N+1}$.

Integrating (3.42) and substituting the boundary conditions, we obtain the exact solution:

$$n(x) = e^{\psi(x) - \psi(-1)} n_0 + J_n e^{\psi(x)} \int_{-1}^{x} \frac{e^{-\psi(t)}}{\mu n(t)} dt \tag{3.43}$$

$$J_n = \frac{n_{N+1} - e^{\psi(1) - \psi(-1)} n_0}{e^{\psi(1)} \int_{-1}^{1} \frac{e^{-\psi(t)}}{\mu n(t)} dt} \tag{3.44}$$

The solution of the corresponding difference equations from the Scharfetter-Gummel method is

$$n_j = e^{\psi_j - \psi_0} n_0 + J_n e^{\psi_j} \sum_{0}^{j-1} \frac{e^{-\psi_k}}{\mu_{n_k + \frac{1}{2}}} \frac{h_k}{\psi_{k+1} - \psi_k} \tag{3.45}$$

$$J_n = \frac{n_{N+1} - e^{\psi_{N+1} - \psi_0} n_0}{e^{\psi_{N+1}} \sum_{0}^{N} \frac{e^{-\psi_k}}{\mu_{n_k + \frac{1}{2}}} \frac{h_k}{\psi_{k+1} - \psi_k}} \tag{3.46}$$

Note that if the integrals are approximated by the sums, then the discrete solution (3.45), (3.46) has the same structure as the continuous solution (3.43), (3.44).

For the given continuous functions ψ, μ_n, the bound

$$\left| \int_{x_j}^{x_{j+1}} \frac{e^{-\psi(t)}}{\mu_n(t)} \, dt - \frac{e^{-\psi_j}}{\mu_{n_{j+\frac{1}{2}}}} \frac{h_j}{\psi_{j+1} - \psi_j} \right| \leq k_1 h_j \qquad (3.17)$$

holds for a constant k_1. Thus, we obtains the global error estimates:

$$\max_{1 \leq j \leq N} |n_j - n(x_j)| \leq Kh \qquad (3.18)$$

for an arbitary mesh. The constant K only depends on ψ. This result shows that the Scharfetter-Gummel discretization is uniformly convergent for zero recombination-generation rate.

3.3 Mesh designing

Choosing a suitable mesh is essential in numerical computation for problems whose solutions change rapidly within some narrow regions. The solutions of the basic semiconductor device equations exhibit this junction layer character, as the singular perturbation analysis shows. This suggests using a coarse mesh outside the layer and a fine one near the junctions. Our mesh generation is then based on this solution character and the idea of equidistributing the local truncation error.

In our simulator, we first compute the zero bias (i.e., thermal equilibrium) solution before calculating the solution for nonzero-bias voltage. The current densities J_n, J_p vanish identically at thermal equilibrium. Therefore, we need only solve the Poisson equation instead of the full set of the equations. As an initial approximation for the iterative solution of the discrete Poisson equation, we can use the solution derived from the assumption of zero space charge, i.e., $\psi = 2 \sinh^{-1} \frac{C}{\delta^2}$.

As an illustration for constructing the initial mesh, we consider a symmetric p-n junction diode with the junction position at $x_J = 0$ and the boundary at $x_0 = -1$, $x_{N+1} = 1$. The initial mesh is designed as follows:

$$x_{j+1} = x_j + h_j, \quad j = 0, 1, \cdots, N,$$
$$h_j = \tfrac{1}{2} h_{j-1}, \qquad j = 1, 2, \cdots, J - 2,$$
$$h_j = 2 h_{j-1}, \qquad j = J + 2, J + 3, \cdots, N.$$

We set the minimun mesh size $\hat{h}$ near the junction with the order of magnitude of the junction layer. That is

$$\hat{h} = h_{J-2} = h_{J-1} = h_J = h_{j+1} = \lambda.$$

The above initial mesh is consistent with the solution character and the initial guess for the discrete Poisson equation, since it is a coarse mesh refined near the junction. This initial mesh has many fewer mesh points than the uniform mesh. To achieve the minimum mesh size $\hat{h} = \lambda = 10^{-3}$, for example, the uniform mesh requires about 2×10^3 mesh points, but the above initial mesh only needs $2 \times (\frac{3}{\log 2}) \approx 23$.

Having an initial mesh and initial guess for the solution, we solve the Poisson equation by Newton's method to get the approximate solution on the initial mesh. Then new mesh points are inserted wherever the estimates of the local truncation error exceed a prescribed accuracy. This is the idea of equidistributing the local truncation error. We solve the Poisson equation to get a new solution and equidistribute the local truncation error again if necessary. The final mesh is constructed by repeating the above mesh-refinment step until achieving a specified accuracy or the total number of mesh points exceeds a prescribed maximum number.

4. NUMERICAL SCHEMES IN SEMICONDUCTOR DEVICE SIMULATION

In this chapter we present the central ideas in our numerical methods for solving the system of discrete semiconductor device equations. The discrete formulation has been discussed in the preceding chapter. The discretization procedure leads to a large nonlinear algebraic system since the semiconductor equations are a nonlinear system of differential equations.

In general, iterative methods must be used for solving the nonlinear algebraic system. Two basic iterative methods are widely used in semiconductor device simulation. The first approach is the nonlinear block Gauss-Seidel method which is a decoupled iteration originally used by Gummel. The second basic approach is the coupled Newton method which has the advantage of locally quadratic convergence.

When we use Newton's method we must find an initial approximation that is close enough to the solution. The natural continuation method is introduced to overcome this problem. This method is an elegant approach for finding a good initial approximation based on the concept of a homotopy.

The multi-solution characteristic of a thyristor creates numerical computation difficulties for the natural continuation method. Such a solution curve has turning points. To pass the turning points, we use the pseudo arc-length continuation method.

At each iterative step, linearizing the algebraic system leads to a large linear system to be solved. The coefficient matrix of this linear system is sparse, i.e. contains mostly zero entries, because of the discretization. For solving the large sparse linear system, it is worth while to use some special techniques which avoid storing and calculating with the zero elements.

4.1 Iterative methods

In this section we discuss the two basic iterative schemes, Gummel's method and Newton's method, used in our device simulator. For notational simplicity, we will write the nonlinear system of discrete semiconductor equations as

$$F(z) = \begin{pmatrix} f_\psi(z) \\ f_n(z) \\ f_p(z) \end{pmatrix} = 0 \qquad (4.1)$$

where

$$z = (\psi, n, p)^T = (\psi_1, \psi_2, \cdots, \psi_N, n_1, n_2, \cdots, n_N, p_1, p_2, \cdots, p_N)^T \qquad (4.2)$$

is the vector of variables of the discrete equations at mesh points. The difference operator $f_\psi = (f_{\psi_1}, f_{\psi_2}, \cdots, f_{\psi_N})^T$ is the discretization (3.28) of the Poisson equation. The difference operators $f_n = (f_{n_1}, f_{n_2}, \cdots, f_{n_N})^T$ and $f_p = (f_{p_1}, f_{p_2}, \cdots, f_{p_N})^T$ are the discretizations (3.29), (3.30) of the combined current relation and continuity equation for electrons and holes.

The nonlinear algebraic system (4.1) can be solved by the nonlinear Gauss-Seidel process. Let $z^k = (\psi^k, n^k, p^k)^T$ be the solution of the kth iterative step and $z^0 = (\psi^0, n^0, p^0)^T$ be given. The nonlinear Gauss-Seidel iteration for solving the

system (4.1) is the following:

$$Solve \quad f_\psi(\ (\psi, n^k, p^k)^T\ ) = 0 \qquad for \quad \psi = \psi^{k+1};$$

$$Solve \quad f_n(\ (\psi^{k+1}, n, p^k)^T\ ) = 0 \qquad for \quad n = n^{k+1};$$

$$Solve \quad f_p(\ (\psi^{k+1}, n^{k+1}, p)^T\ ) = 0 \quad for \quad p = p^{k+1};$$

$$k = 0, 1, \cdots.$$

This iteration has the advantage that only one equation is solved on each step. Its disavantage is that the convergence can be rather slow for high currents.

In the original version of Gummel's method, the mobilities and recombination-generation rate were evaluated using the previous iterative values and the quasi-Fermi level variables $(\psi, \varphi_n, \varphi_p)$ were used instead of variables (ψ, n, p). The electron and hole continuity equations are then linear and can be directly solved. The discrete Poisson equation $f_\psi = 0$ for the variables $(\psi, \varphi_n, \varphi_p)$ in the kth step nonlinear Gauss-Seidel procedure is

$$\lambda^2 \frac{2}{h_{j-1} + h_j} \left(\frac{\psi_{j-1}^{k+1}}{h_{j-1}} - (\frac{1}{h_{j-1}} + \frac{1}{h_j})\psi_j^{k+1} + \frac{\psi_j^{k+1}}{h_j} \right)$$

$$- \delta^2 \left(e^{\psi_j^{k+1} - \varphi_n^k} - e^{\varphi_{p_j}^k - \psi_j^{k+1}} \right) - C_j = 0 \qquad (1.3)$$

since $n = \delta^2 e^{\psi - \varphi_n}$ and $p = \delta^2 e^{\varphi_p - \psi}$. In Gummel's original algorithm, just one step of Newton's method is performed on this equation (4.3). For variables (ψ, n, p), the equation (4.3) is easily tranformed into

$$\lambda^2 \frac{2}{h_{j-1} + h_j} \left(\frac{\psi_{j-1}^{k+1}}{h_{j-1}} - (\frac{1}{h_{j-1}} + \frac{1}{h_j})\psi_j^{k+1} + \frac{\psi_j^{k+1}}{h_j} \right)$$

$$- \left(n_j^k e^{\psi_j^{k+1} - \psi_j^k} - p_j^k e^{\psi_j^k - \psi_j^{k+1}} \right) - C_j = 0 \qquad (4.1)$$

The other way to solve the Poisson equation in this Gauss-Seidel procedure is to solve the linear equation

$$\lambda^2 \frac{2}{h_{j-1}+h_j} \left(\frac{\psi_{j-1}^{k+1}}{h_{j-1}} - (\frac{1}{h_{j-1}} + \frac{1}{h_j})\psi_j^{k+1} + \frac{\psi_j^{k+1}}{h_j} \right)$$

$$- \left(n_j^k - p_j^k - C_j \right) = 0 \tag{4.5}$$

It is preferable to solve (4.4) rather than (4.5) fairly accurately even though (4.5) is more easily solved.

Gummel's method was widely used in early device simulation because it decouples the full system. Kerkhoven [31] showed that the Gummel method unconditionally generates iterates that are successively closer to the solution while the distance to the solution is sufficiently large. However, it is now well known that this method may fail to converge for high currents, or when the recombination-generation rate plays an important role. Our results show that it converges rapidly for the first few iterative steps, even from a poor initial guess, but slows down close to the solution.

Since the Gummel method(decoupled algorithm) does not always converge rapidly, we use Newton's method. That is, let z^0 be given, and for each $k = 0, 1, 2, \cdots$ solve

$$\frac{dF}{dz}(z^k)s^k = -F(z^k)$$

for the Newton correction s^k, where $\frac{dF}{dz}$ is the Frechet derivative(Jacobian matrix) of F; set

$$z^{k+1} = z^k + s^k.$$

It is well known that Newton's method is quadratically convergent from a good initial guess if the Frechet derivative is nonsingular. The most important result about

Newton's method is due to Kantorovich. Using the contraction mapping theorem, he proved the convergence of Newton's method without assuming that a solution exists.

We state the Newton-Kantorovich theorem here without proofs. A proof can be found in [30].

Theorem 4.1 *Newton-Kantorovich: Let* $F : R^N \to R^N$ *be continuously differentiable. For some* $z^0 \in R^N$, *and a neighborhood* $N(z^0, r)$ *of* z^0, *let* F *satisfy:*

(a) $\frac{dF}{dz}(z^0)$ *is nonsingular with*

$$\left\| \left[\frac{dF}{dz}(z^0) \right]^{-1} \right\| \leq \beta;$$

(b)

$$\left\| \left[\frac{dF}{dz}(z^0) \right]^{-1} F(z^0) \right\| \leq \eta;$$

(c)

$$\left\| \frac{dF}{dz}(u) - \frac{dF}{dz}(v) \right\| \leq \gamma \|u - v\|, \quad for \quad u, v \in N(z^0, r)$$

If $\alpha = \beta\gamma\eta \leq \frac{1}{2}$ *and* $r \geq (1 - \sqrt{1 - 2\alpha})/\beta\gamma$ *then the sequence* z^k *produced by Newton's method:*

$$z^{k+1} = z^k - \left[\frac{dF}{dz}(z^k) \right]^{-1} F(z^k), \quad k = 0, 1, 2, \cdots.$$

is well defined and converges to a root z^* *of* $F(z) = 0$. *If* $\alpha < \frac{1}{2}$ *then* z^* *is the unique root of* $F(z) = 0$ *in the closure of a neighborhood* $N(z^0, r_1)$ *of* z^0, *where* $r_1 = \min\{r, (1 - \sqrt{1 - 2\alpha})/\beta\gamma\}$ *and*

$$\|z^k - z^*\| \leq (2\alpha)^{2^k} \frac{\eta}{\alpha}, \quad k = 0, 1, 2, \cdots.$$

Two problems need be considered in the application of Newton's method for practical numerical computation: overshooting, and the choice of initial iterate. The Newton correction s^k may overestimate the length of the step which should have been taken. This is called overshoot. To illustrate this problem, we use the Newton method to solve the realistic problem:

$$F(u) = e^{40u} - e^{-40} = 0.$$

The solution of this example is $u = -1$. Using $u^0 = -1.2$ as the initial guess, in the first step Newton method, we obtain $u^1 = 73.3$. It is farther away from the solution than the initial guess. We can not continue to compute the next Newton correction because of exponential overflow.

The damped Newton method is introduced to overcome the overshooting problem. That is, for given z^0, at each iteration k step, solve

$$\frac{dF}{dz}(z^k)s^k = -F(z^k),$$

for the Newton correction s^k. The solution sequence $\{x^k\}$ is generated by setting

$$z^{k+1} = z^k + d^k s^k, \quad k = 0, 1, \cdots.$$

The damping parameter d^k is chosen to make the Newton correction decrease the residual. In device simulation, the damped Newton method can be used to solve Poisson's equation with quasi-Fermi level variables. We have not encountered the overshooting problem in our device simulation experience while using the variables (ψ, n, p).

The Newton method can fail to converge if started from a bad initial guess. In the device problem, we can use Gummel's method to provide an initial guess

which is in the region of convergence of Newton's method. Then only a few Newton iterations must be performed to achieve an accurate solution. One needs a good criterion for switching from the Gummel method to the Newton method. We employ the continuation method to provide an initial guess. The applied external voltage is taken as the continuation parameter. We start with the solution at zero bias and then calculate the tangent vector to predict an initial approximation for a non-zero bias, say, the applied voltage V_1. Having the solution at the voltage V_1, we can repeat the continuation method to provide the initial approximation for the voltage $V_2 = V_1 + \Delta V$ for sufficiently small ΔV. We use successfully this continuation method in calculating the current-voltage curve of the p-n diode. Finding a reasonable initial guess directly for a specified set of applied external voltages is not achieved so far as we know.

4.2 Continuation method

In this we discuss the continuation method for calculating the current-voltage characteristic. The solution $z = (\psi, n, p)^T$ of the discrete steady state semiconductor equations (4.1), can be regarded as a function $z = z(V)$ of V for an applied external voltage $V \in (V^-, V^+)$. It may be a multivalued function of V if there are multiple steady solutions. The current density $J = J_n + J_p$ depends on the solution vector z, that is, $J = J(z, V)$. Therefore, the nonlinear system (4.1) of discrete equations can be regarded as parameterized system $F : R^{3N} \times R \longrightarrow R^{3N}$

$$F(z, V) = \begin{pmatrix} f_\psi(z, V) \\ f_n(z, V) \\ f_p(z, V) \end{pmatrix} = 0 \qquad (4.6)$$

with the natural parameter $V \in R$.

In order to get the current-voltage $(J - V)$ curve, we must compute the solution path $\Gamma = \{(z, V) | F(z, V) = 0\}$. At a desired bias $V = V_1$, we can apply Newton's method to solve the system (4.6). As discussed in the preceding section, it is difficult to find an initial approximation close to the solution.

Let $(z_0(V_0), V_0)$ be a root of $F(z(V), V) = 0$ and $\frac{dF}{dz}(z(V_0))$ be nonsingular with bounded inverse, i.e., $\left\| \left[\frac{dF}{dz}(z(V_0)) \right]^{-1} \right\| \leq \beta$. For notational simplicity, we write $z_i = z(V_i)$ for the solution of $F(z(V), V) = 0$ at $V = V_i$. If $F(z, V)$ and $\frac{dF}{dz}(z, V)$ are continuous on $N(z_0, r_1) \times N(V_0, r_2)$, where $N(z_0, r_1)$ is a neighborhood of z_0 and $N(V_0, r_2)$ is a neighborhood of V_0, then the implicit function theorem implies the existence of a unique solution branch $z = z(V)$ in the neighborhood $N(V_0, r_2)$ of V_0.

We will construct an approximation z_1^0 to the solution $z(V_1)$ of $F(z, V_1) = 0$ using the solution z_0. From the implicit function theorem we have

$$F(z(V), V) = 0, \qquad V \in N(V_0, r_2). \tag{4.7}$$

By differentiating (4.7) with respect to V and the chain rule, we get

$$\frac{dF}{dz}(z, V)\dot{z}(V) + \frac{dF}{dV}(z, V) = 0, \tag{4.8}$$

here, $\dot{z}(V)$ denotes the derivative of $z(v)$ with respect to V. Evaluating (4.8) at (z_0, V_0), we have the tangent vector

$$\dot{z}_0 = \dot{z}(V_0) = - \left[\frac{dF}{dz}(z_0, V_0) \right]^{-1} \frac{dF}{dV}(z_0, V_0) \tag{4.9}$$

Then, we take the initial approximation z_1^0 for $z_1 = z(V_1)$ by Euler predictor as

$$z_1^0 = z_0 + \Delta V_0 \dot{z}_0 \tag{4.10}$$

here, $V_1 = V_0 + \Delta V_0 \in N(V_0, r_2)$ for small enough ΔV_0. The error is given by

$$z_1 - z_1^0 = \frac{1}{2}\ddot{z}_0 \Delta V_0 + O(\Delta V_0) \tag{4.11}$$

For calculating the solution $z(V_1)$ of $F(z, V_1) = 0$, we use z_1^0 as the initial gusee for Newton iteration:

$$z_1^{k+1} = z_1^k - \left[\frac{dF}{dz}(z_1^k, V_1)\right]^{-1} F(z_1^k, V_1), \quad k = 0, 1, \cdots. \tag{4.12}$$

Suppose that $\left\| \left[\frac{dF}{dz}(z, V)\right]^{-1} \right\| \le \beta$ and $\left\| \frac{d^2 F}{dz^2}(z, V) \right\| \le \gamma$, then the sequence $\{z_1^k\}^{k=0,1,\cdots}$ generated by the Newton method (4.12) converges to the solution (z_1, V_1) for sufficiently small ΔV_0. In order to apply the Newton-Kantorovich theorem, we need the condition

$$\left\| \left[\frac{dF}{dz}(z_1^0)\right]^{-1} F(z_1^0) \right\| \le \eta.$$

Since $\left\| \frac{d^2 F}{dz^2}(z, V) \right\| \le \gamma$, we have

$$\left\| F(z_1^0) - F(z_0) - \frac{df}{dz}(z_0)\Delta V_0 \dot{z}_0 \right\| \le \frac{\gamma}{2}\|\Delta V_0 \dot{z}_0\|^2. \tag{4.13}$$

From (4.9), (4.13) and $F(z_0) = 0$, we obtain the estimate

$$\|f(z_1^0)\| \le |\Delta V_0| \left\| \frac{dF}{dV}(z_0) \right\| + \frac{\gamma \beta^2}{2}|\Delta V_0|^2 \left\| \frac{dF}{dV}(z_0) \right\|^2. \tag{4.14}$$

Therefore,

$$\left\| \left[\frac{dF}{dz}(z_1^0)\right]^{-1} F(z_1^0) \right\| \le \beta \left(|\Delta V_0| \left\| \frac{dF}{dV}(z_0) \right\| + \frac{\gamma \beta^2}{2}|\Delta V_0|^2 \left\| \frac{dF}{dV}(z_0) \right\|^2 \right) = \eta. \tag{4.15}$$

For sufficiently small ΔV_0, we can require $\alpha = \beta \gamma \eta < \frac{1}{2}$, and $r \ge r_0 = (1 - \sqrt{1 - 2\alpha})/\beta\gamma$ such that the neighborhood $N(z1^0, r) \subset N(z_0, r_1)$. Then the convergence follows from the Newton-Kantorovich theorem.

We now compute the solution path $\Gamma = \{(z, V) | F(z, V) = 0, \quad V \in [V_0, V_+]\}$

starting with giving solution (z_0, V_0) by the natural continuation procedure as follows:

begin starting with an solution (z_i, V_i).

predictor calculating the initial approximation z_i^0 by Euler prediction (4.10).

corrector computing the solution (z_{i+1}, V_{i+1}) by Newton method

$$z_i^{k+1} = z_i^k - \left[\frac{dF}{dz}(z_i^k, V_i)\right]^{-1} F(z_i^k, V_i), \quad k = 0, 1, \cdots. \qquad (4.16)$$

end If $V_{i+1} = V_+$, then stop; else replace (z_i, V_i) by (z_{i+1}, V_{i+1}) and go to *begin* step.

If each subsequence $\{z_i^k\}^{k=1,2,\cdots}$ produced by the Newton correction (4.26) converges to the solution (z_i, V_i), then we have the discrete solution path.

Some devices have multiple solutions for certain bias. For example, the four-layer structure device p-n-p-n thyristor has three steady state solutions. A curve containing multiple solutions will have singular points where the Jacobian $\frac{dF}{dz}$ is not invertible. The natural continuation procedure will break down at singular points. For example, at a turning point(simple limit point) the Euler predictor step may have ΔV_i for which no nearby solution exists, and the Newton correction will generally fail to converge. This leads us to modify the natural continuation method to overcome the turning point problem.

We define the regular points and turning points mathematically.

1. *Regular Point:*

   The solution point (z_0, V_0) of $F(z, V) = 0$ is called a *regular point* if the Jacobian $\frac{dF}{dz}(z_0, V_0)$ is nonsingular.

2. *Turning point:*

The solution point (z_0, V_0) of $F(z, V) = 0$ is called a *turning point* if

(a) $\frac{df}{dz}(z_0, V_0)$ is singular,

(b) $\frac{df}{dV}(z_0, V_0) \notin R(\frac{df}{dz}(z_0, V_0))$.

(c) $\dim N\left(\frac{df}{dz}(z_0, V_0)\right) = \operatorname{codim} R\left(\frac{df}{dz}(z_0, V_0)\right) = 1$.

Here, N denotes the null space and R denotes the range.

The pseudo arc-length continuation method is introduced for passing the turning points. In the pseudo arc-length continuation method, we reparametrize the solution (z, V) by a pseudo arc-length parameter s instead of the natural parameter V. The discrete nonlinear system (4.6) can be parametrized by pseudo arc-length s as follows:

$$F(z(s), V(s)) = \begin{pmatrix} f_\psi(z(s), V(s)) \\ f_n(z(s), V(s)) \\ f_p(z(s), V(s)) \end{pmatrix} = 0 \qquad (4.17)$$

with an unit arc-length equation

$$\|\dot{z}(s)\|^2 + |\dot{V}(s)|^2 - 1 = 0, \qquad (4.18)$$

where $\dot{z}$ and $\dot{V}$ are derivatives of z and V with respect to the pseudo arc-length parameter s. To compute the solution path $\Gamma = \{(z, V)|F(z, V) = 0 \ s \in [s_0, s_+]\}$ and pass the turning point , the pseudo arc-length continuation method solves the augmented system

$$G(z, V, s) = \begin{pmatrix} F(z(s), V(s)) \\ \|\dot{z}(s)\|^2 + |\dot{v}(s)|^2 - 1 \end{pmatrix} = 0 \qquad (4.19)$$

instead of just solving $F(z, V) = 0$.

The pseudo arc-length continuation procedure starts with a given solution point $(z_0, V_0) \equiv (z(s_0), V(s_0))$ as follows:

begin starting with an solution $(z_i, V_i) \equiv (z(s_i), V(s_i))$.

tangent vector compute the tangent vector $(\dot{z}_i, \dot{V}_i) \equiv (\dot{z}(s_i), \dot{V}(s_i))$. Differentiating (4.27) with respect to s as in the natural continuation method and using the chain rule, we have

$$\frac{dF}{dz}(z_i, V_i)\dot{z}_i + \frac{dF}{dV}(z_i, V_i)\dot{V}_i = 0, \tag{4.20}$$

Solving equations (4.20), (4.18), we get the tangent vector:

$$\dot{z}_i = g_i \dot{V}_i,$$
$$\dot{V}_i = \pm\sqrt{\frac{1}{\|g_i\|^2 + 1}} \tag{4.21}$$

here g_i satisfies

$$\frac{dF}{dz}(z_i, V_i)g_i = -\frac{dF}{dV}(z_i, V_i). \tag{4.22}$$

predictor calculating the initial approximation z_i^0, V_i^0 by Euler prediction along the tangent for a step size $\Delta s_i = s_{i+1} - s_i$:

$$z_{i+1}^0 = z_i + \Delta s_i \dot{z}_i,$$
$$V_{i+1}^0 = V_i + \Delta s_i \dot{V}_i. \tag{4.23}$$

corrector computing the solution (z_{i+1}, V_{i+1}) by Newton method. Since the arc-length equation (4.18) is not suited for computation purposes, we approximate this equation by a linearization:

$$N(z(s), V(s), s) \equiv < \dot{z}_i, z(s) - z_i > + \dot{V}(V(s) - V_i) - (s - s_i) = 0 \tag{4.24}$$

A geometrical interpretation for equation (4.24) is that this is a hyperplane perpendicular to the tangent vector $(\dot{z}_i, \dot{V}_i)$ at a distance $|s - s_i|$ from (z_i, V_i). This hyperplane intersects the solution curve Γ if $|s - s_i|$ and the curvature of Γ are small enough. We now solve the augmented system

$$G(z, V, s) = \begin{pmatrix} F(z(s), V(s), s) \\ N(z(s), V(s), s) \end{pmatrix} = 0 \tag{4.25}$$

for (z_{i+1}, V_{i+1}) by Newton's method with initial guess (4.23)

$$\begin{bmatrix} z_{i+1}^{k+1} \\ V_{i+1}^{k+1} \end{bmatrix} = \begin{bmatrix} z_{i+1}^{k} \\ V_{i+1}^{k} \end{bmatrix} - \left[\frac{dG}{d(z,v)}(z_{i+1}^k, V_{i+1}^k) \right]^{-1} G(z_{i+1}^k, V_{i+1}^k), \quad k = 0, 1, \cdots. \tag{4.26}$$

Here, $\frac{dG}{d(z,v)}$ is the Jacobian matrix:

$$\frac{dG}{d(z,v)} = \begin{bmatrix} \frac{\partial F}{\partial z} & \frac{\partial F}{\partial V} \\ \frac{\partial N}{\partial z} & \frac{\partial N}{\partial V} \end{bmatrix} \tag{4.27}$$

Keller has shown that $\frac{dG}{d(z,v)}$ is nonsingular at the regular points and simple limit points. Therefore, there is no computation problem for solving (4.26). That is, this method can pass turning points.

end If $s_{i+1} = s_+$, then stop; else replace (z_i, V_i) by (z_{i+1}, V_{i+1}) and go to *begin* step.

4.3 Implementation and conditioning

In this section we discuss the implementation details in our device simulator. We first solve the Poisson equation by Newton's method at thermal equilibrium, i.e.

zero bias. For the p-n diode, we use the natural continuation method for computing the solutions with nonzero biases starting with the thermal equilibrium solution. For the p-n-p-n thyristor, we use the pseudo arc-length continuation method for computing the solution path. The two iterative schemes, Gummel's method and Newton's method are the basic tools in the continuation algorithm.

The most time-consuming part of the iterative methods for solving the discrete semiconductor equations is calculating the Jacobian and computing the solutions of the linear system resulting from the linearization. To maintain quadratic convergence, the Jacobian is evaluated exactly by computing the derivatives with respect to ψ, n, p in our computer codes. The coefficient matrix of the linear system, i.e. the Jacobian matrix, is sparse because of the discretization. Some sparse direct methods are used since they give us an exact solution of the sparse linear system if it is well conditioned.

We now turn to compute the solution of the Poisson equation at thermal equilibrium by using Newton's method. The linearization of the discrete Poisson equation at each Newton's iterative step in matrix-vector notation is

$$As = b, \tag{4.28}$$

where A is the Jacobian matrix, s is the correction vector and b is the residual vector of the Poisson equation. The Jacobian matrix $A = (a_{ij})$ is a tridiagonal matrix with elements

$$a_{i,i} = \frac{2\lambda^2}{h_{i-1} + h_i}\{\frac{1}{h_{i-1}} + \frac{1}{h_i} + \delta^2 e^{\psi_i} + \delta^2 e^{-\psi_i}\},$$

$$a_{i,j+1} = a_{i-1,j} = \frac{2\lambda^2}{h_{i-1} + h_i}h_i.$$

Matrix A is a diagonally dominant and positive definite tridiagonal matrix. Therefore, we can use Linpack dgtsl or dptsl directly solving the system (4.28).

For Gummel's method, the coefficient matrices of the discrete electron and hole continuity equations and the Jacobian of the Poisson equation at non-zero biases are also tridiagonal matrices. We still use Linpack dgtsl to solve these equations.

Using Newton's method for solving the coupled system, we must rearrange the order of the equations and the variables to get band matrix form. The unknowns are arranged as

$$z = (z_i)^T \equiv (\psi_1, n_1, p_1, \psi_2, n_2, p_2, \cdots, \psi_N, n_N, p_N)^T$$

with

$$z_i = (\psi_i, n_i, p_i),$$

and the discrete equations are numbered as

$$F(z) = (F_i(z))^T \equiv (f_{\psi_1}, fn_1, fp_1, f_{\psi_2}, fn_2, fp_2, \cdots, f_{\psi_N}, fn_N, fp_N)^T$$

with

$$F_i = (f_{\psi_i}, fn_i, fp_i).$$

Under this ordering, the Jacobian matrix $A = \frac{dF}{dz}$ of the coupled system takes the following form:

$$A = \begin{bmatrix} D_1 & U_1 & & & & & \\ L_2 & D_2 & U_2 & & & 0 & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & & & \\ & 0 & & & L_{N-1} & D_{N-1} & U_{N-1} \\ & & & & & L_N & D_N \end{bmatrix}$$

(4.29)

$$L_i = \begin{bmatrix} \dfrac{\partial f_{\psi_i}}{\partial \psi_{i-1}} & \dfrac{\partial f_{\psi_i}}{\partial n_{i-1}} & \dfrac{\partial f_{\psi_i}}{\partial p_{i-1}} \\[2em] \dfrac{\partial f_{n_i}}{\partial \psi_{i-1}} & \dfrac{\partial f_{n_i}}{\partial n_{i-1}} & \dfrac{\partial f_{n_i}}{\partial p_{i-1}} \\[2em] \dfrac{\partial f_{p_i}}{\partial \psi_{i-1}} & \dfrac{\partial f_{p_i}}{\partial n_{i-1}} & \dfrac{\partial f_{p_i}}{\partial p_{i-1}} \end{bmatrix} \tag{4.30}$$

$$D_i = \begin{bmatrix} \dfrac{\partial f_{\psi_i}}{\partial \psi_i} & \dfrac{\partial f_{\psi_i}}{\partial n_i} & \dfrac{\partial f_{\psi_i}}{\partial p_i} \\[2em] \dfrac{\partial f_{n_i}}{\partial \psi_i} & \dfrac{\partial f_{n_i}}{\partial n_i} & \dfrac{\partial f_{n_i}}{\partial p_i} \\[2em] \dfrac{\partial f_{p_i}}{\partial \psi_i} & \dfrac{\partial f_{p_i}}{\partial n_i} & \dfrac{\partial f_{p_i}}{\partial p_i} \end{bmatrix} \tag{4.31}$$

$$U_i = \begin{bmatrix} \dfrac{\partial f_{\psi_i}}{\partial \psi_{i+1}} & \dfrac{\partial f_{\psi_i}}{\partial n_{i+1}} & \dfrac{\partial f_{\psi_i}}{\partial p_{i+1}} \\[2em] \dfrac{\partial f_{n_i}}{\partial \psi_{i+1}} & \dfrac{\partial f_{n_i}}{\partial n_{i+1}} & \dfrac{\partial f_{n_i}}{\partial p_{i+1}} \\[2em] \dfrac{\partial f_{p_i}}{\partial \psi_{i+1}} & \dfrac{\partial f_{p_i}}{\partial n_{i+1}} & \dfrac{\partial f_{p_i}}{\partial p_{i+1}} \end{bmatrix} \tag{4.32}$$

The Jacobian matrix A (4.29) is in a block tridiagonal form or can be regarded as a band matrix. Therefore, at each Newton's iterative step the linear system $As = b$ can be solved by a block tridiagonal solver or a band system solver.

The linearized equations for the thyristor are ill-conditioned along the low-current branch but for the diode are well-conditioned under Ohmic contacts. We decompose the domain to three overlapping subdomains each containing one junc-

tion. Then each subdomain behaves like a diode. Under this domain decomposition, the Jacobian matrix A is separated into three overlapping submatrices A_1, A_2, A_3 as follows:

$$A_1 = \begin{bmatrix} D_1 & U_1 & & & & & \\ L_2 & D_2 & U_2 & & & 0 & \\ & \cdot & \cdot & \cdot & & & \\ & & L_{j_1} & D_{j_1} & U_{j_1} & & \\ & & & \cdot & \cdot & \cdot & \\ & 0 & & & L_{j_2-1} & D_{j_2-1} & U_{j_2-1} \\ & & & & & L_{j_2} & D_{J_2} \end{bmatrix} \tag{4.33}$$

$$A_2 = \begin{bmatrix} D_{j_1} & U_{j_1} & & & & \\ L_{j_1+1} & D_{j_1+1} & U_{j_1+1} & & 0 & \\ & \cdot & \cdot & \cdot & & \\ & & L_{j_2} & D_{j_2} & U_{j_2} & \\ & & & \cdot & \cdot & \cdot \\ & & & L_k & D_k & U_k \\ & & & & \cdot & \cdot & \cdot \\ & 0 & & & & L_{J_3-1} & D_{J_3-1} & U_{J_3-1} \\ & & & & & & L_{j_3} & D_{J_3} \end{bmatrix} \tag{4.34}$$

$$
A_3 = \begin{bmatrix}
D_k & U_k & & & & & \\
L_{k+1} & D_{k+1} & U_{k+1} & & & 0 & \\
& \cdot & \cdot & \cdot & & & \\
& & L_{j3} & D_{j3} & U_{j3} & & \\
& & & \cdot & \cdot & \cdot & \\
& 0 & & & L_{N-1} & D_{N-1} & U_{N-1} \\
& & & & & L_N & D_N
\end{bmatrix} . \tag{4.35}
$$

Each submatrix is block tridiagonal. Therefore, we can use the block Gauss-Seidel method to solve the linear system $As = b$.

For the pseudo arc-length continuation method, we will illustrate the determination of the direction of the tangent vector, the selection of step size and the solution of the linear system (4.26) in the Newton correction step.

In the tangent vector step (4.20), (4.21), (4.22), the choice of the sign is determined by the direction of the solution path. If we have computed two previous solutions (z_{i-1}, V_{i-1}), (z_i, V_i), then the choice depends on the inner product

$$
\theta_i = < \dot{z}_i, z_i - z_{i-1} > + \dot{V}_i(V_i - V_{i-1}).
$$

The positive inner product $\theta_i > 0$ implies that one travels in the direction from (z_{i-1}, V_{i-1}) to (z_i, V_i).

In the predictor step (4.23), we must keep the step size Δs_i small enough that the initial guess (4.23) is in the convergence range of Newton's method. It requires

to estimate the curvature of the solution path and the convergence radius for sophisticated step size control. Bank and Mittelmann [32] propose a strategy for step size selection in continuation procedures with damped Newton's method.

We choose the step size based on the residuals at Newton correction steps. At first we use a trial step size $\Delta s_1 = s_1 - s_0$ to perform the first few Newton iterations. If the residuals decrese quadratically then we multiply the trial step size by two, else by one-half. The above procedure is repeated until a suitable first step size is achieved. For all other step sizes $\Delta s_i = s_{i+1} - s_i$, the previous step size is taken to be the trial step size, i.e., $\Delta s_i = \Delta s_{i-1}$ for the Newton iteration. If the residuals do not decrease on the first few steps then we multiply the step size by one-half; else we accept the step size.

In the correction step (4.26), we must solve the linear system

$$\begin{bmatrix} \frac{\partial F}{\partial z} & \frac{\partial F}{\partial V} \\ \frac{\partial N}{\partial z} & \frac{\partial N}{\partial V} \end{bmatrix} \begin{bmatrix} s_z \\ s_V \end{bmatrix} = - \begin{bmatrix} F \\ N \end{bmatrix}. \tag{4.36}$$

At regular points this linear system (4.36) can be solved by block Gaussian elimination:

(1) Solve

$$\frac{\partial F}{\partial z} w = \frac{\partial F}{\partial V}$$

and

$$\frac{\partial F}{\partial z} y = \frac{\partial F}{\partial V}$$

(2) Set

$$s_V = \left(-N - \frac{\partial N}{\partial z} \right) \Big/ \left(\frac{\partial N}{\partial V} - \frac{\partial N}{\partial z} w \right).$$

(3) Set

$$s_z = y + s_V w.$$

Only two linear systems with coefficient matrix $\frac{\partial F}{\partial z}$ must be solved. This matrix $\frac{\partial F}{\partial z}$ is a block tridiagonal matrix. Therefore, these two linear systems can be solved by a block tridiagonal solver or band solver.

5. RESULTS AND CONCLUSION

5.1 Computation results

In this section we present some computational results. In order to demonstrate the numerical methods discussed in the preceding chapters, we illustrate the condition number and convergence rates. A one dimensional p-n diode and a p-n-p-n thyristor are simulated to highlight the structure of solutions. The main result is the current-voltage characteristic of the thyristor obtained by the pseudo arc-length continuation method.

The length of the domain for both of the models is $l = 50\mu m$. Constant mobilities $\mu_n = 1500 \ cm^2/volt-sec$, $\mu_p = 480 \ cm^2/volt-sec$ and intrinsic carrier density $n_i = 1.5 \times 10^{10} \ cm^{-3}$ are adopted for silicon at room temperature [4]. Both models include SRH, Auger and impact ionization recombination-generation rates. The symmetrical abrupt junction doping profile $C(x)$ for the diode is used as follows:

$$C(x) = \begin{cases} -10^{17} \ cm^{-3} & -1 \leq x < 0. \\ 10^{17} \ cm^{-3} & 0 \leq x < 1. \end{cases}$$

This yields $\lambda^2 \approx 10^{-7}$ and $\delta^2 \approx 10^{-7}$. The condition number of the Jacobian for this diode model is 10^{10} measured by using LINPACK routine dgbco. It has six significant digits for double precision with sixteen significant digits.

The convergence rates of Gummel's method and Newton's method at applied

contact voltage $V = 1.5$ *volts* are illustrated in Figure 5.1. Observe that Gummel's method converges rapidly for the first iterative steps but slows down close the solution. Newton's method shows quadratic convergence.

The forward bias scaled current-voltage characteristic of the diode is demonstrated in Figure 5.2, where the voltage V is expressed in units of the thermal voltage $V_T = 0.026$ *volts*. The current is very small at voltage less than the cut-in voltage ≈ 28. Figure 5.3 shows the characteristic for reverse bias. The small saturation current is about 10^{-6} and the breakdown voltage is about 800. The junction layer widths in the diode for various applied voltages are shown in Figure 5.4. This shows that the widths decrease as applied contact voltage increases.

The doping profile $C'(x)$ for the thyristor model is as follows:

$$C'(x) = \begin{cases} -10^{17}, & -1 \le x \le -0.4, \\ 10^{14}, & -0.4 < x \le 0.6, \\ -10^{15}, & 0.6 < x \le 0.9, \\ 10^{18}, & 0.9 < x \le 1. \end{cases}$$

This doping profile is close to reality in a practical physical model and yields the parameters $\lambda^2 \approx 10^{-8}$ and $\delta^2 \approx 10^{-8}$.

Figure 5.5 illustrates the three steady state current-voltage characteristic of thyristor. This curve is obtained by solving the full set of semiconductor equations with realistic physical parameters by using the pseudo arc-length continuation method. We start with the thermal equilibrium solution and then take a larger step size to jump directly to the on-state high current solutions. From an on-state solution, the pseudo arc-length continuation procedure walks backward and uses small enough step size $\Delta s \approx 10^{-2}$ to pass the turning point(holding voltage ≈ 0.3 *volts*).

Kurata [12] uses the p-i-n diode doping profile to reach a high current on-state. Then he switches to the p-n-p-n thyristor doping profile at a certain high current. Thereafter, he decreases the voltage gradually and gets empirically the holding current through a number of computations.

The condition number of the Jacobian for the thyristor in the high current on-state has the same order of magnitude as that for the diode. Therefore, the high current solutions can be gotten directly by using a band solver or block tridiagonal solver at each Newton iterative step. However, the condition number of the Jacobian is 10^{21} at the low current blocking-state, making no significant digits. The condition number of the Jacobian for each subdomain containing only one junction reduces to 10^{10}. That is, the submatrix containing only one junction of the Jacobian is well-conditioned. This motivates use of the block Gauss-Seidel method as discussed in Chapter 4. In Figure 5.5, the blocking-state low currents are obtained by this method at each Newton iterative step.

The electric fields of the thyristor for various applied voltages in the blocking state is shown in Figure 5.6. The middle junction layer spreads as the applied voltage increses. This *Early effect* [4] describing the effect on the current of reverse biased spreading junction layer prevents current saturation. Rubinstein [20] and Steinruck [21], using the zero space charge approximation, found current saturation. In another paper [29], Steinruck rescaled the potential with λ^2 to construct the S-shape characteristics. Ward et al. [22] incorporated the *Early effect* into the reduced problem and showed how this layer spreading prevents current saturation.

The electrostatic potential distribution for the blocking state and the unstable state at positive applied voltages comparing with thermal equilibrium are illustrated

in Figures 5.7 and 5.8. Figure 5.8 shows that the middle junction is reverse biased and the others are forward biased. At on-state, Figure 5.9 shows that the three junction are all forward biased. These phenomena have been explained in device physics [4]. The unstable state shows the same biasing as the blocking state.

The electron current density J_n and hole current density J_p distributions are demonstrated in Figures 5.10, 5.11 and 5.12 for the three steady states. Figure 5.10 shows the current densities are smooth in the on-state. In the unstable state, Figure 5.11 shows the current densities have no jump at the middle junction. Figure 5.12 shows the current densities have junction layers for each junction in the blocking state. Rubinstein [20] and Steinruck [21] contructed steady state solutions for the reduced equations using the condition $[J_n]x_j = 0$ and $[J_p]x_j = 0$, here x_j denotes the junction. The Figures show that continuity conditions only hold in the on-state. Recent work of Ward et al. [22] indicates that these conditions are not valid for all regions of the current-voltage curve and suggest solving the full set of equations.

5.2 Conclusion

A one-dimensional thyristor is simulated using the pseudo arc-length continuation method. It is interesting and difficult work because of the multiple solutions and ill-conditioned Jacobian of the model. Traditionally, the device development is based on experiment. However, the numerical simulations offer great benefits in shortened development time and lower development cost.

By singular perturbation analysis the solution exhibits junction layers. We use a special non-uniform mesh to save computer memory and calculation. It is efficient for solutions with junction layers. The Scharffetter-Gummel discretization scheme

is used to get the discrete equation. Three linear solvers are applied for solving the sparse linear system and overcoming the ill-conditioning at each iterative step. The current-voltage characteristics are obtained by the pseudo arc-length continuation method.

We get a non-monotone current-voltage characteristic for the thyristor with long carrier lifetimes in the blocking-state in a very small voltage interval illustrated in Figure 5.13. For short carrier lifetimes, it is monotone increasing. This phenomenon has not been reported in any paper so far as we know. Possibly, it is the effect of rounding error. However, Ward et al. [22] indicate that the solution is complicated for small current density $J \ll O(1)$. This requires further investigation.

The connection between the unstable-state and blocking state is not achieved. It is due to the overflow and extreme ill-conditioning. We need to study this problem further.
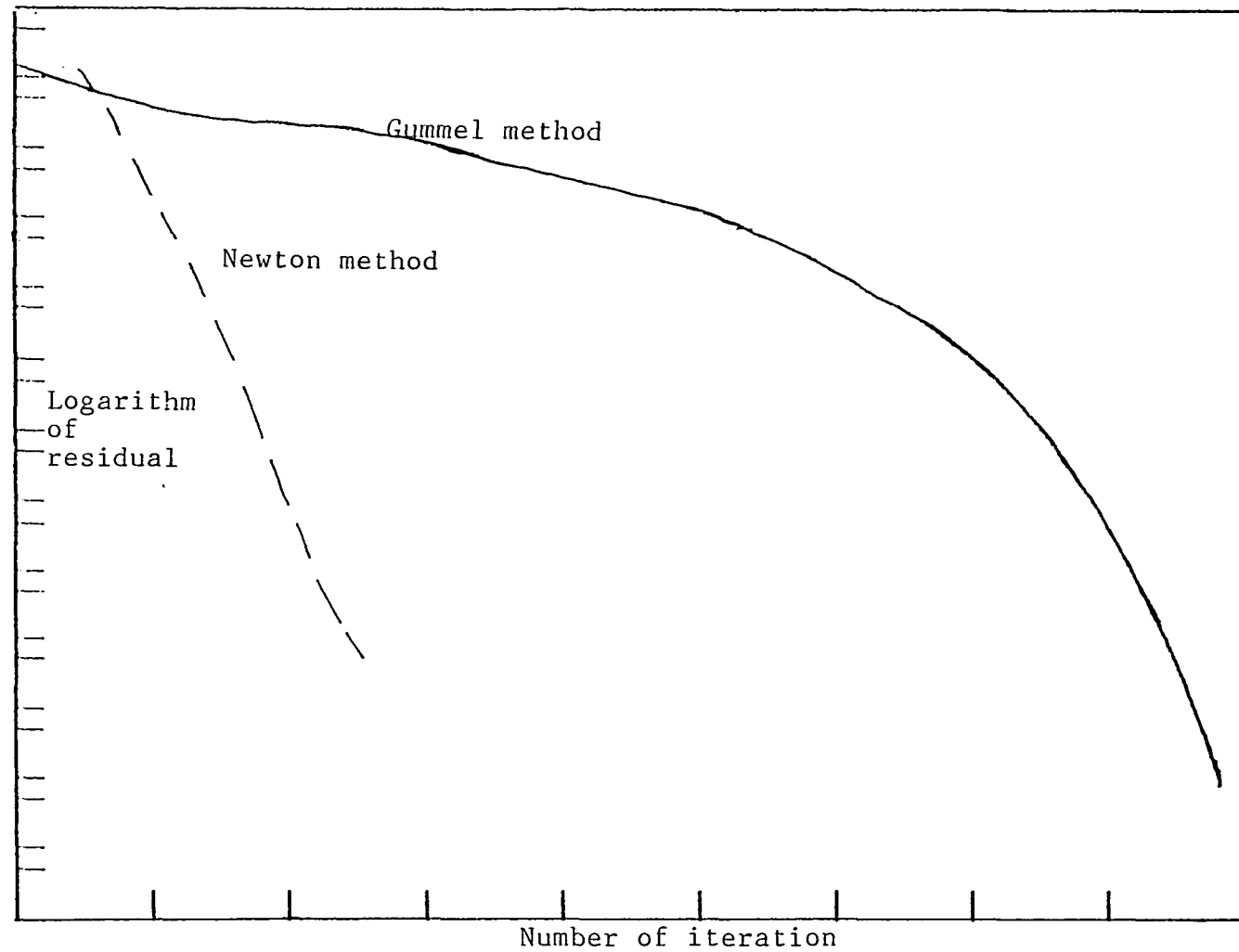
Figure 5.1: Speed of convergence
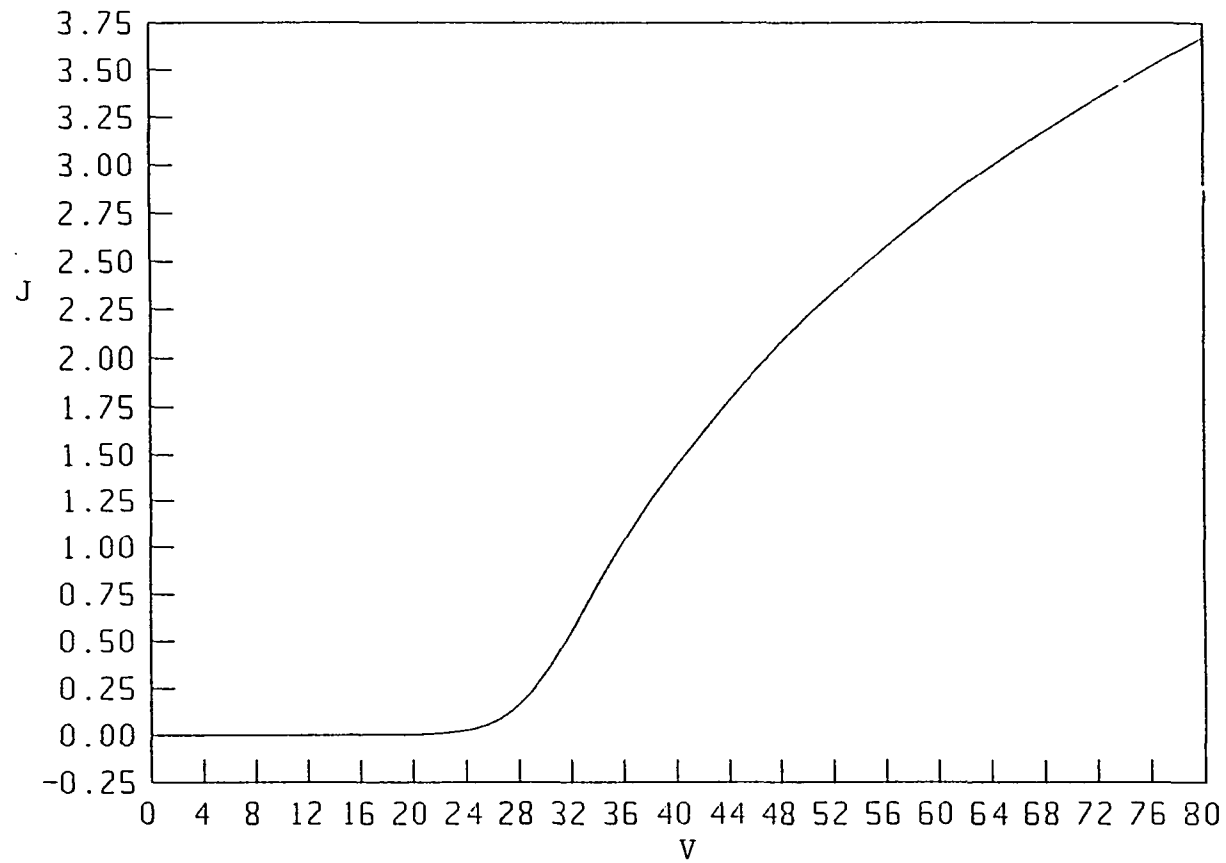
Figure 5.2: Current-voltage characteristic of a diode at forward bias
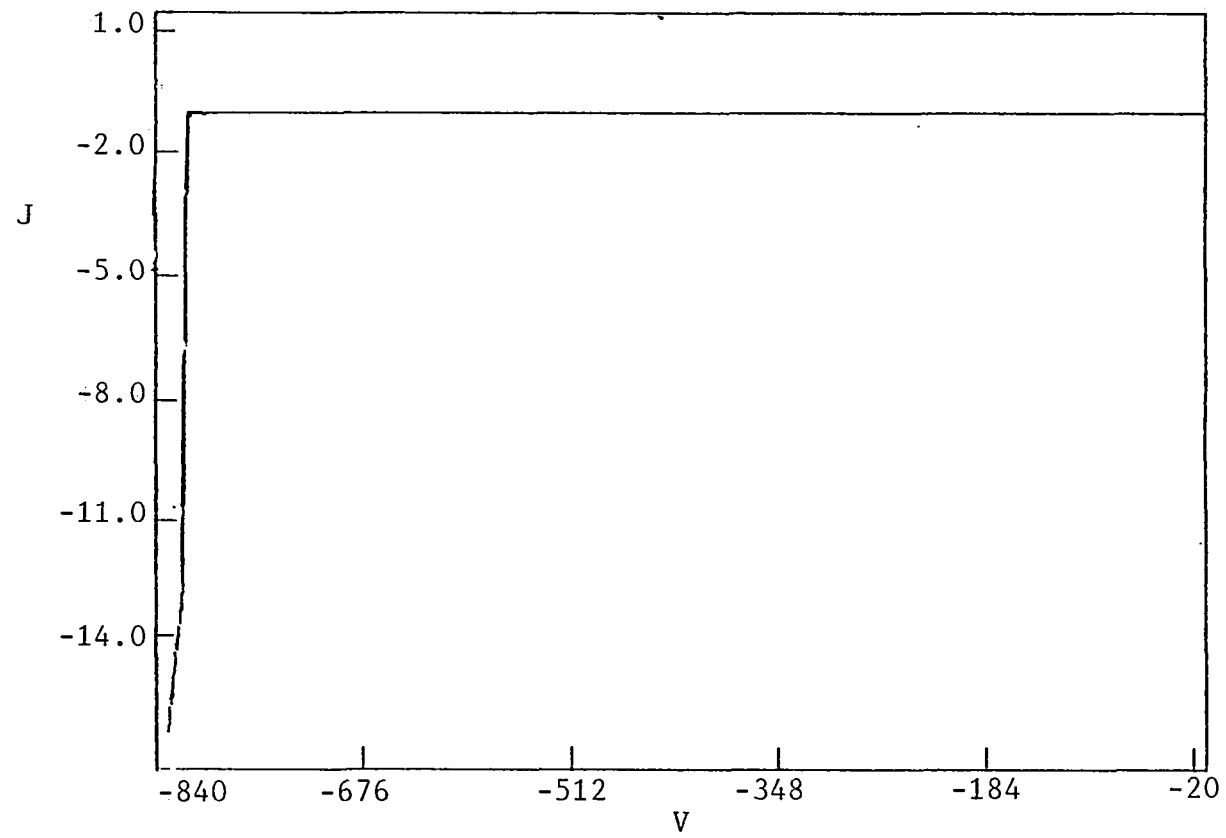
Figure 5.3: Current-voltage characteristic of a diode at reverse bias
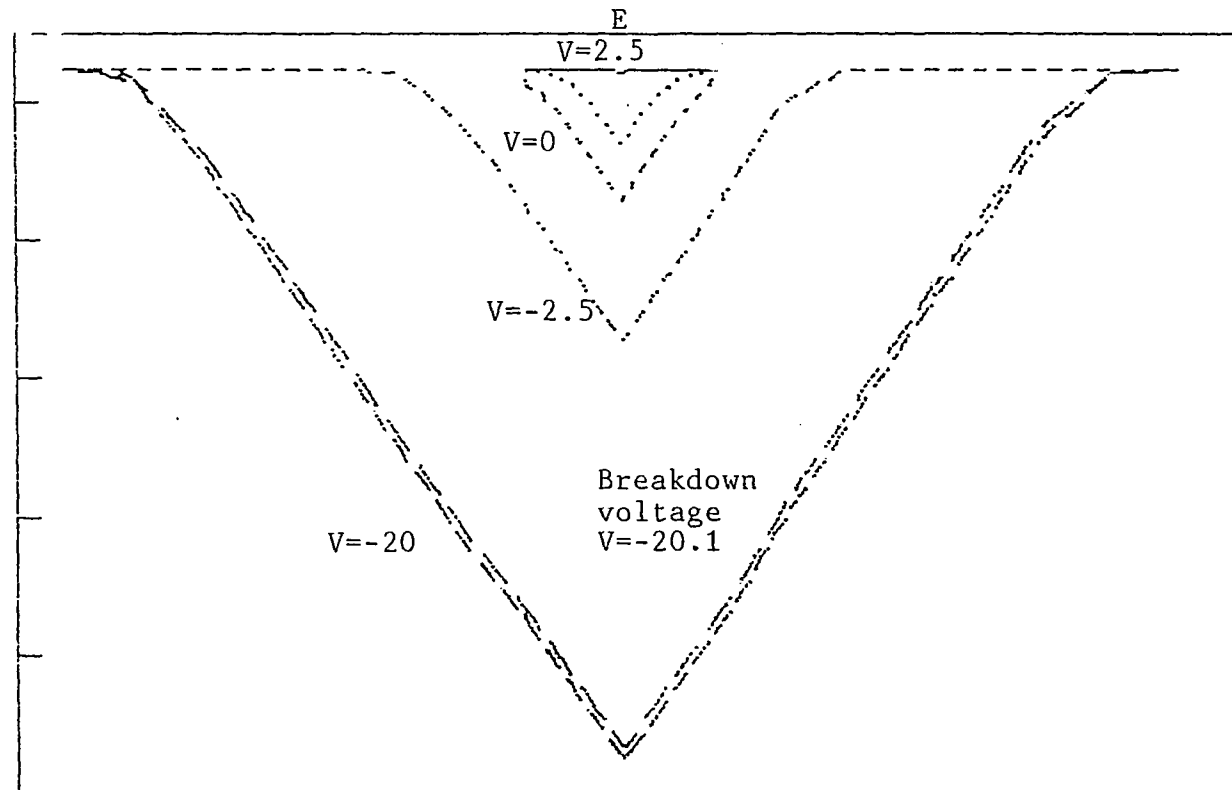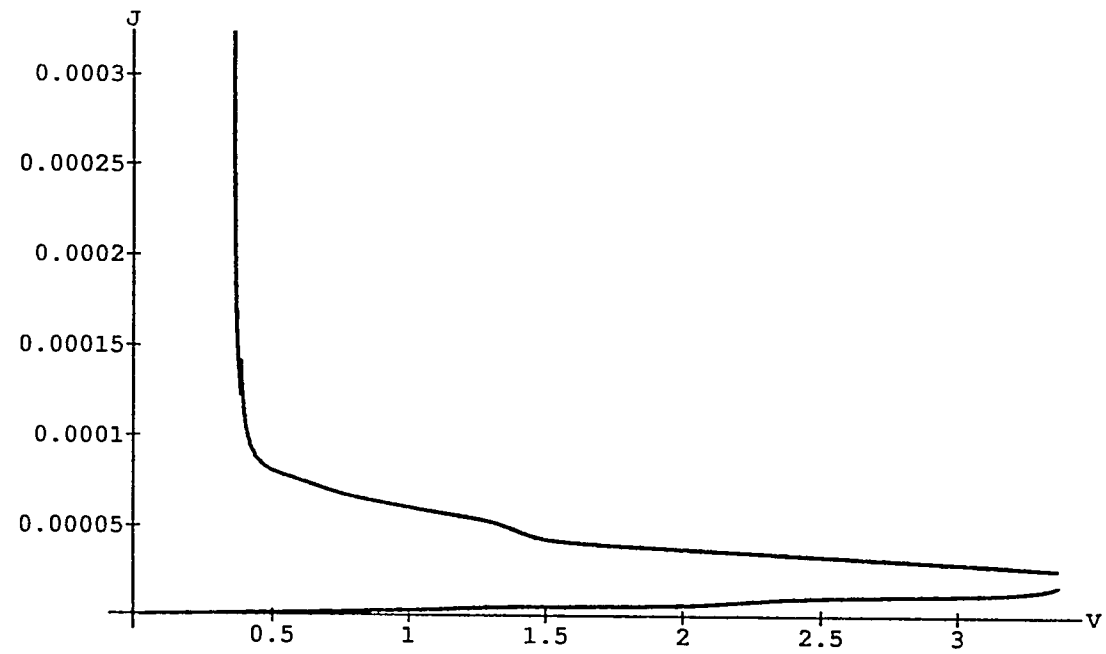
Figure 5.4:   Junction layer widths of a diode
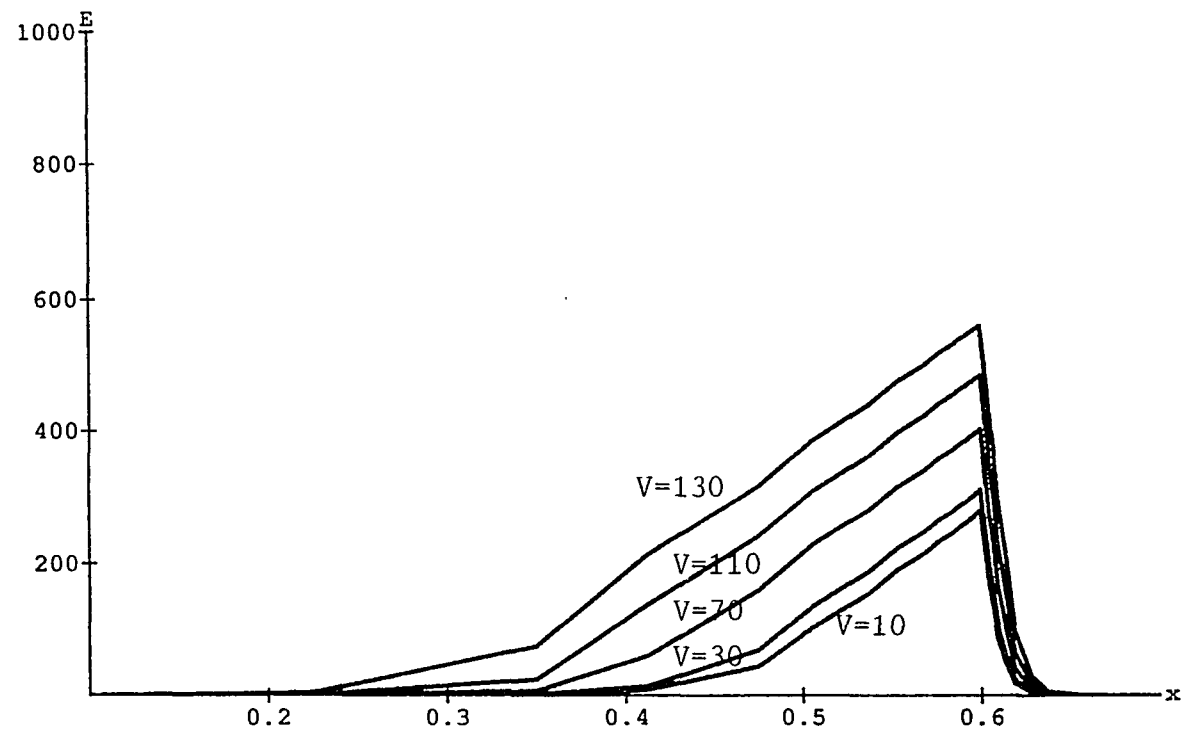
Figure 5.5: Current-voltage characteristic of a thyristor

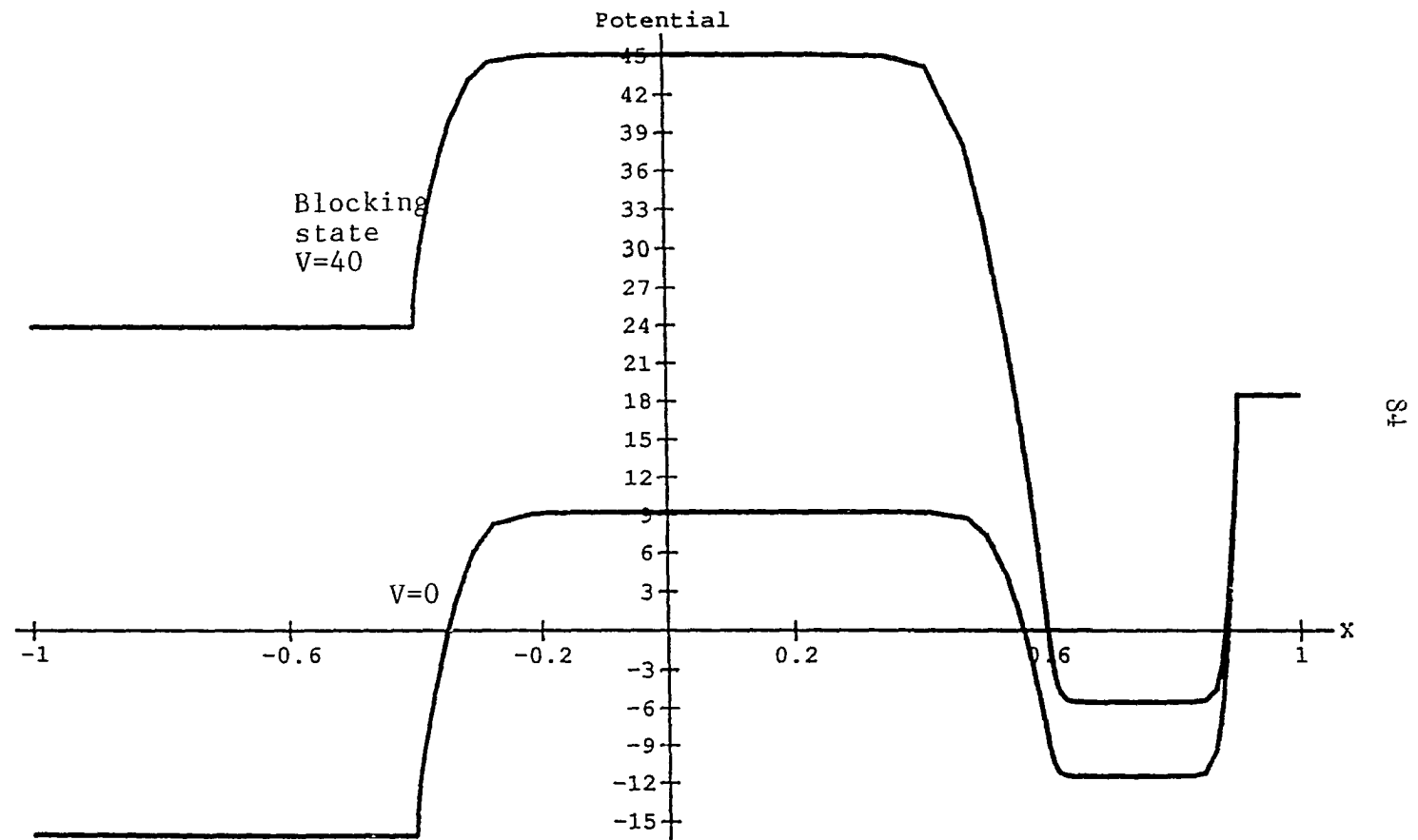Figure 5.6:   Electtric fields of a thyristor at blocking state
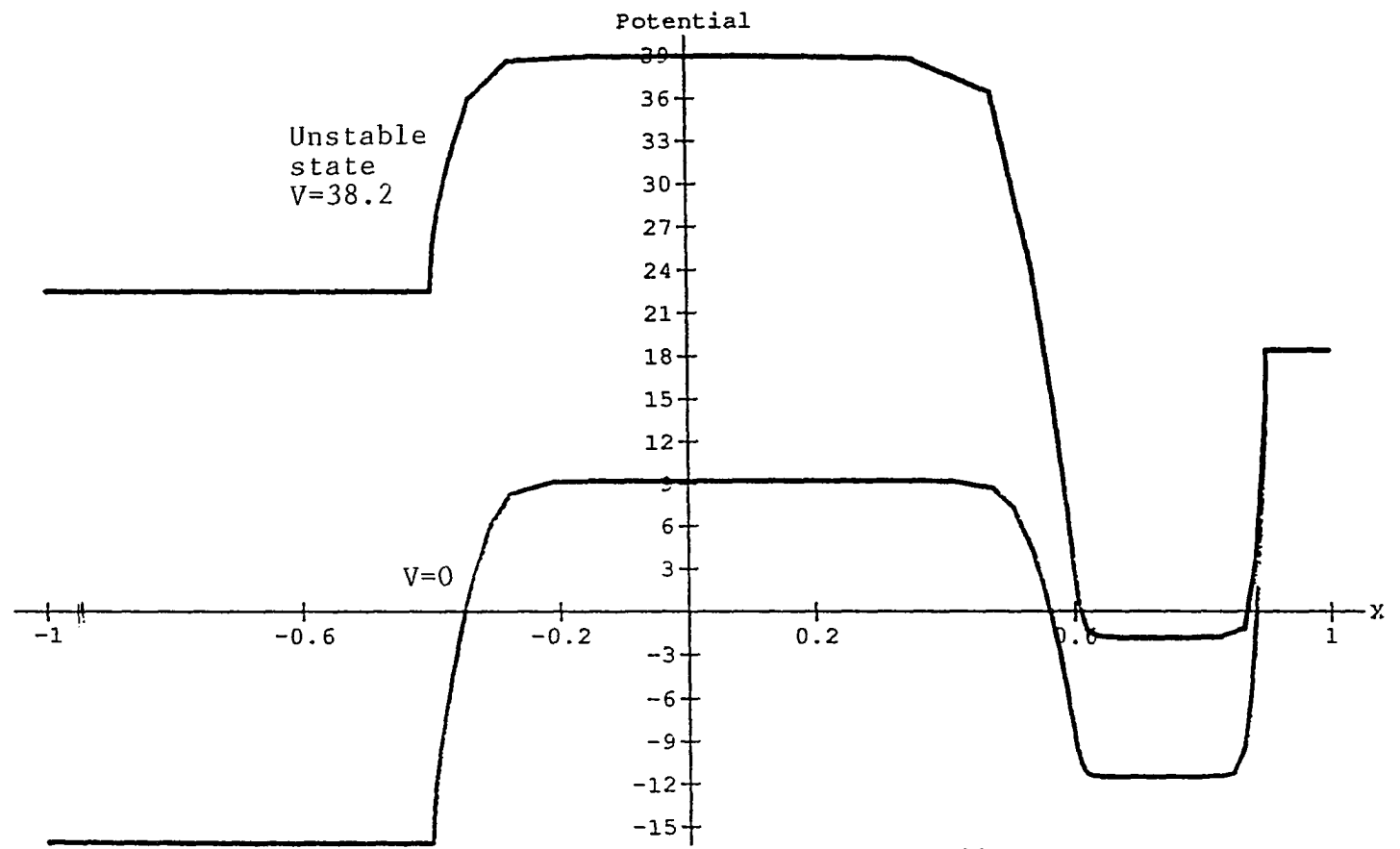
Figure 5.7: Electrostatic potential at blocking state
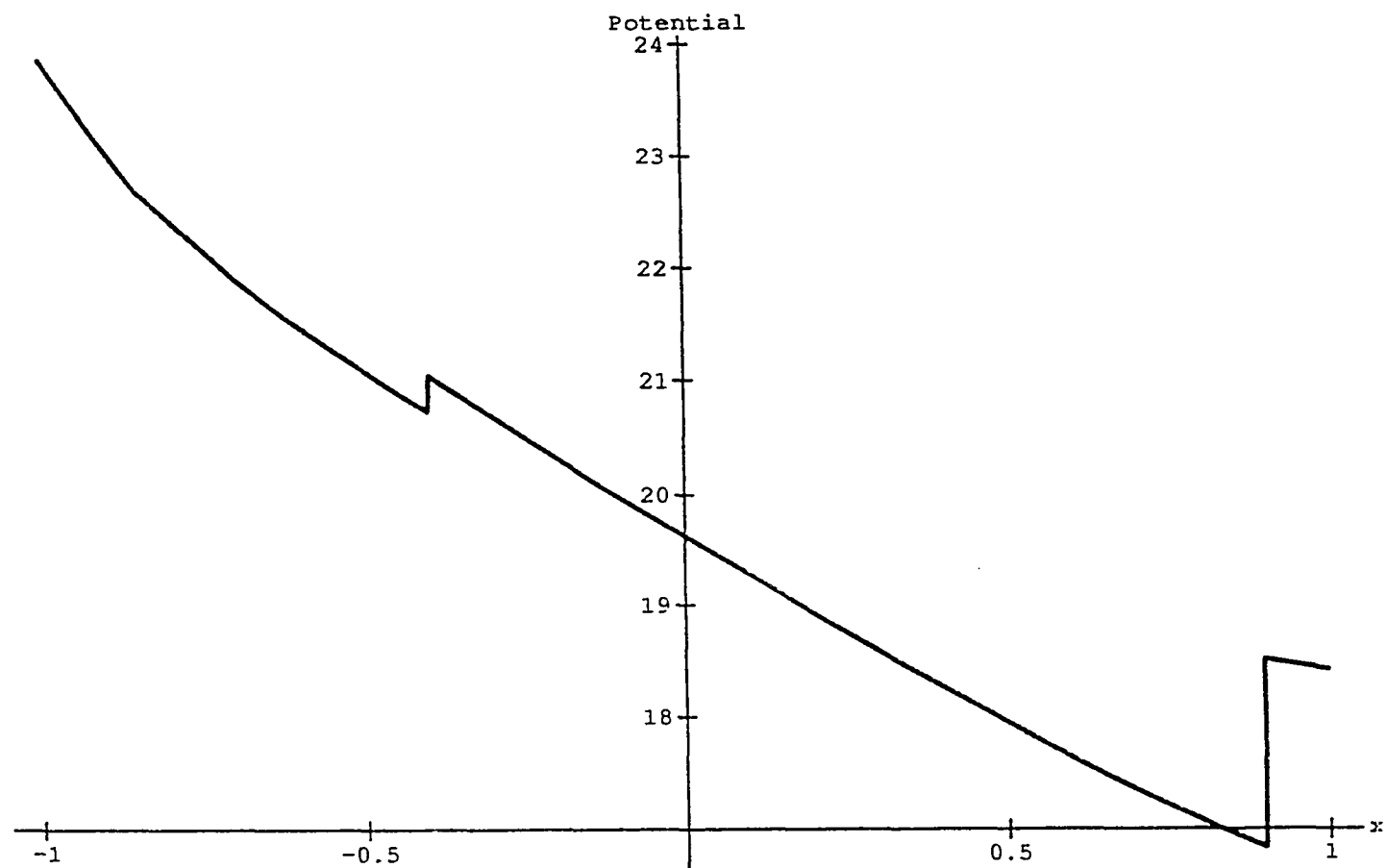
Figure 5.8: Electrostatic potential at unstable state
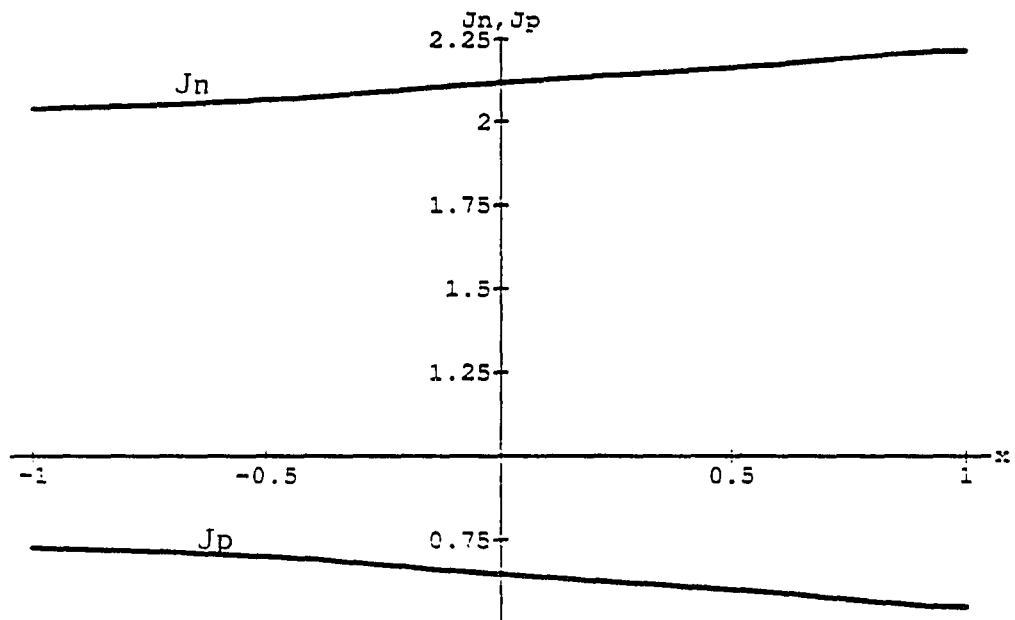
Figure 5.9:   Electrostatic potential at on-state

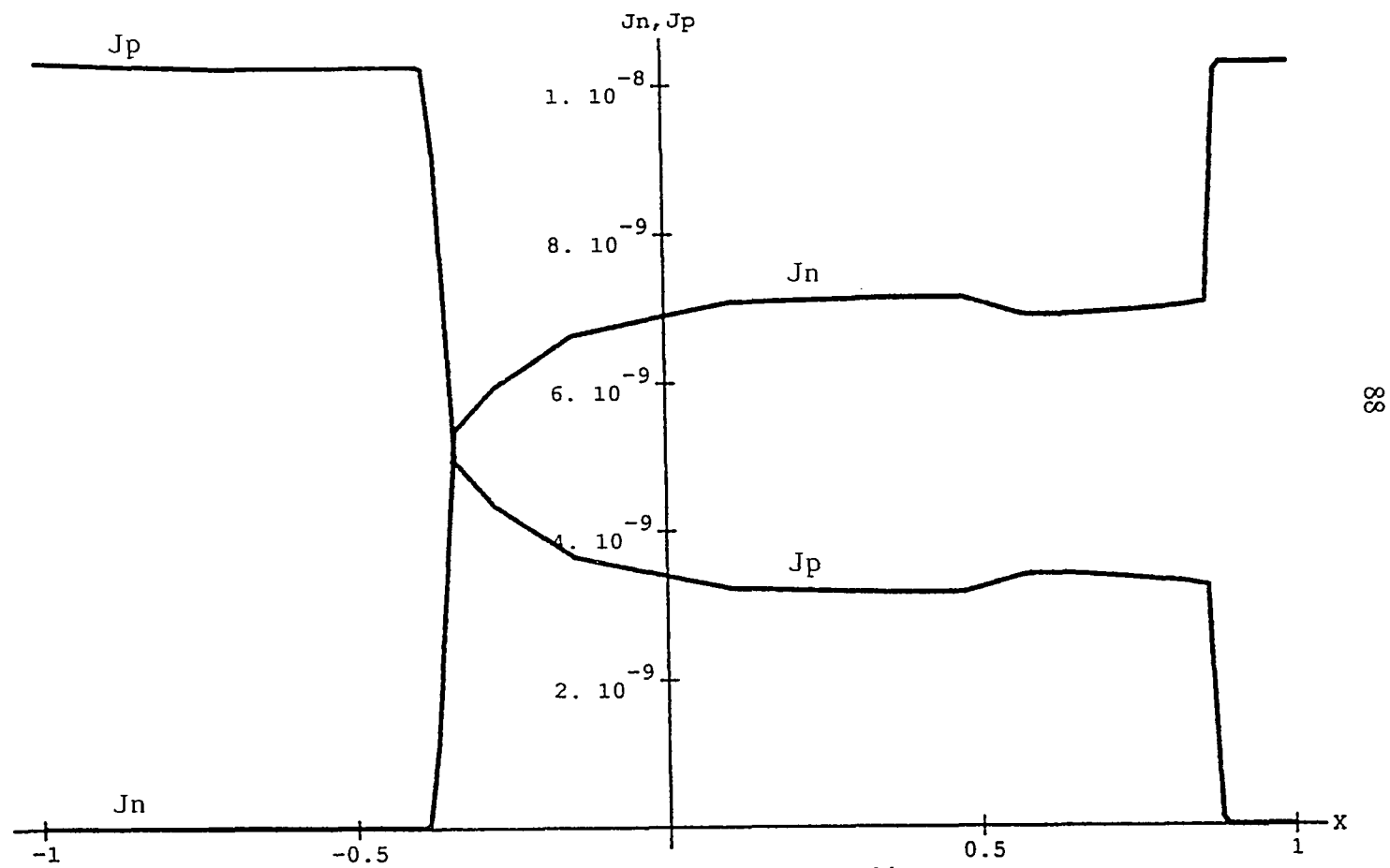Figure 5.10:   Current densities at on-state

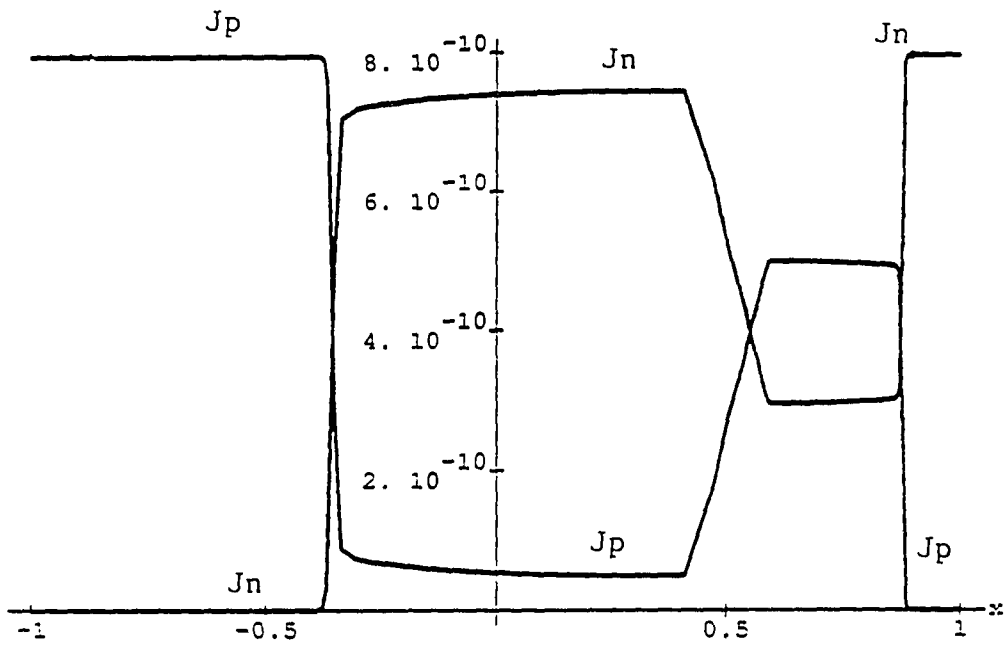Figure 5.11: Current densities at unstable state

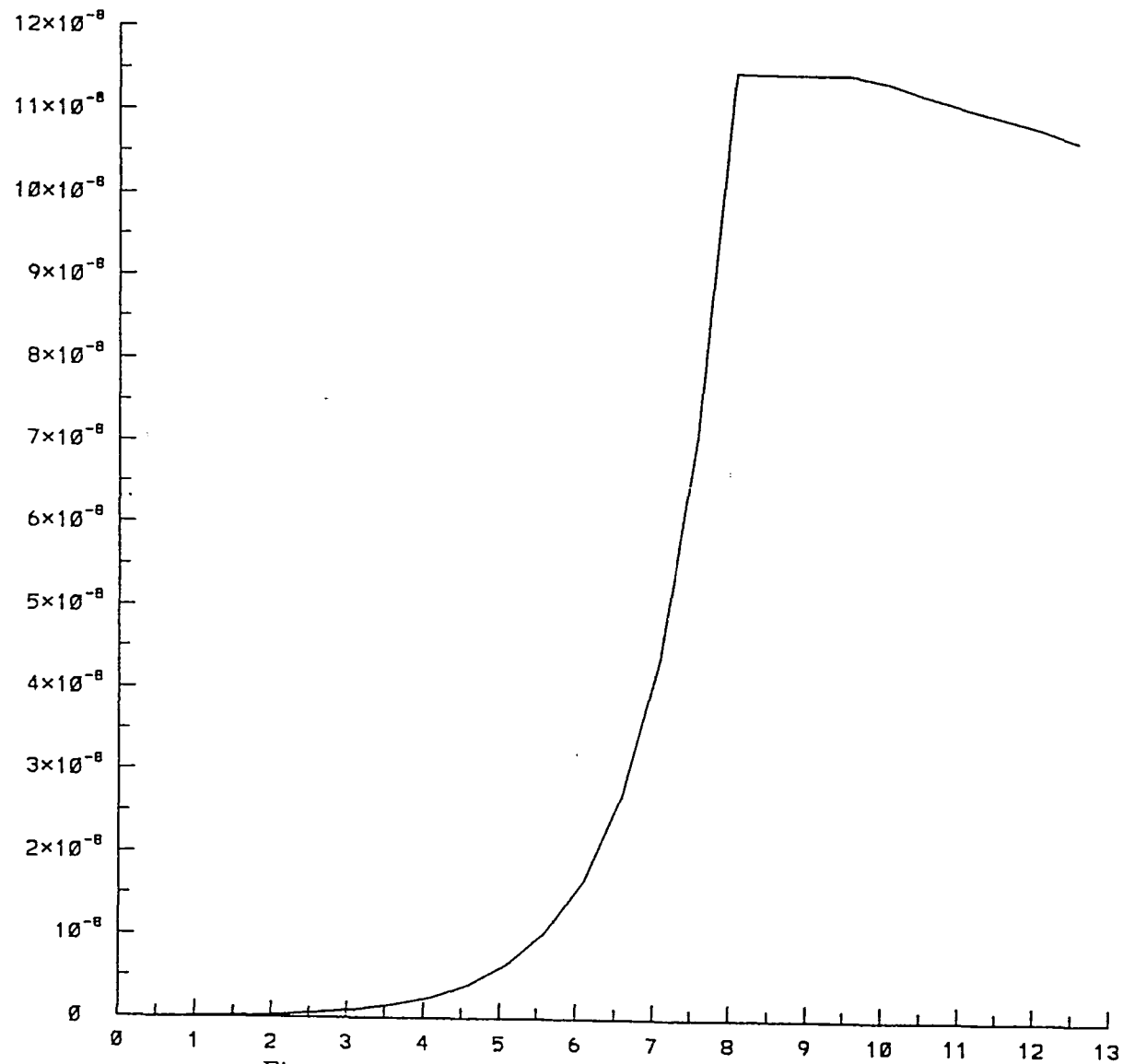Figure 5.12: Current densities at blocking state

Figure 5.13:   Non-monotone current-voltage characteristic

BIBLIOGRAPHY

[1] W. Shockley, *The Theory of p-n Junctions in semiconductors and p-n Junction Transistor*. Bell System Technical Journal 28 pp. 435-489 1949.

[2] W. V. Van Roosbroeck, *Theory of Flow of Electrons and Holes in Germanium and other Semiconductors*. Bell System Technical Journal 29 pp. 560-607 1950.

[3] W. Shockley, *A Unipolar Field Effect Transistor*. Proc. IRE pp. 1365-1377 November 1952.

[4] S. M. Sze, *Physics of Semiconductor Devices*. New York: J. Wiley 1981.

[5] H. K. Gummel, *A Self-consistent Iterative Scheme for One-dimensional Steady State Transtor Calculation*. IEEE Trans. Electron Devices, ED-16 pp. 455-465 1964.

[6] A. Demari, *An Accurate Numerical Steady-State One-dimensional Solution of the p-n Junction*. Solid-State Electron, 11, pp. 33-58 1968.

[7] A. Demari, *An Accurate Numerical Steady-State One-dimensional Solution of the p-n Junction under Arbirary Transient Conditions*. Solid-State Electron, 11, pp. 1021-2053 1968.

[8] D. L. Scharfetter, and H. K. Gummel, *Large-singnal Analysis of a Silicon Read Diode Oscillator*. IEEE Trans. Electron Devices, ED-20 no.1. pp. 64-67 1969.

[9] J. W. Slotboom, *Computor-aided Two-dimensional Analysis of Bipolar Transistors*. IEEE Trans. Electron Devices, ED-20 no.8. pp. 669-679, 1973.

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

[11] C. M. Snowden, *Introduction to Semiconductor Device Modeling*. World Scientific, Singpore, 1986.

[12] M. Kurata, *One-dimensional Calculation of Thyristor Forward Voltage and Holding Currents*. Solid-State Electron, 19, pp. 527-535 1976.

[13] P. A. Markowich, *The Stationary of Semiconductor Devices equations*. Springer-Verlag, 1986.

[14] R. E. Bank, and D. J. Rose, *Some Error Estimates for the Box Method*. SIAM J. Numer. Anal. 24, No. 4, pp. 777-787 1987.

[15] H. B. Keller, *Numerical Solution of Bifurcation and Nonlinear Eigenvalue Problem. Applications of Bifurcation Theory*, Rabinowitz, P. ed. Academic Press, pp. 359-384 1977.

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

[17] T. I. Seidman, *Steady State Solutions of Diffusion-Reaction Systems with Electrostatic Convection*. Nonlinear Analysis, Theory, Methods Appl. 4, No. 3, pp. 623-637 1980.

[18] J. W. Jerome, *Consistency of Semiconductor Modeling: An Existence/Stability Analysis for the Staionary Van Roosbroeck System*. SIAM J. Appl. Math. 45, No. 4, pp. 565-590 1985.

[19] M. S. Mock, *An Example of Nonuniqueness of Staionary Solutions in Semiconductor Devices Models*. COMPEL 1, No. 3, pp. 165-174 1982.

[20] I. Rubinstein, *Multiple Steady States in One-Dimensional Electrodiffusion th Local Electroneutrality*. SIAM J. Appl. Math. 47, No. 5, pp. 1076-1093 1987.

[21] H. Steinrück, *A biffurcation Analysis of the One-Dimensional Steady State Semiconductor Device Equations*. SIAM J. Appl. Math. 49, No. 4, pp. 1102-1121 1989.

[22] M. J. Ward, L. G. Reyna, and F. M. Odeh, *Multiple Steady-State Solutions in a Multijunction Semiconductor Device*. SIAM J. Appl. Math. 51, No. 1, pp. 99-123 1991.

[23] H. B. Keller, *Lectures on Numerical Methods in Bufurcation Problem*. Springer-Verlag, 1987.

[24] P. A. Markowich, *A Singular Perturbation Analysis of the Fundamental Semi-conductor Device Equations*. SIAM J. Appl. Math. 44, No. 5, pp. 896-928 1984.

[25] U. M. Ascher, R. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Prentice-Hall Inc. 1988.

[26] R. E. Bank, D. J. Rose, and W. Fichtner *Numerical Methods for Semiconductor Device Simulation*. SIAM J. Sci. Stat. Comput. 4, No. 3, pp. 416-435 1983.

[27] R. S. Varga, *On Diagonal Dominance Arguments for Bounding* $\|A^{-1}\|_\infty$. Linear Algebra Appl., 14, pp. 211-217 1976b.

[28] E. P. Doolan, J. J. H. Miller, and W. H. A. Schilders, *Uniform Numerical Methods for Problems with Initial and Boundary Layers*. Boole Press, Dublin, 1983.

[29] H. Steinrück, *Asymptotic Analysis of the Current-Voltage Curve of a pnpn Semiconductor Device*. IMA J. Appl. Math. 43, pp. 243-259 1989.

[30] J. M. Ortega, and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variable*. Academic Press, Inc 1970.

[31] T. Kerkhoven, *On the Effectiveness of Gummel's Method*. SIAM J. Sci. Stat. Comput. 9, No. 1, pp. 48-60 1988.

[32] R. Bank, and H. D. Mittelmann, *Stepsize selection in continuation procedures and damped newton's method*. Journal of computational and Applied Mathematics 26, pp. 67-77 1989.