TWO DIMENSIONAL MOS-TRANSISTOR MODELING

S.SELBERHERR, A.SCHÜTZ and H.PÖTZL

Institut für Allgemeine Elektrotechnik und Elektronik
Abteilung für Physikalische Elektronik
Technische Universität Wien
Gußhausstraße 27
A-1040 Wien, AUSTRIA
    and
Ludwig Boltzmann Institut für Festkörperphysik

ABSTRACT - The advent of Very Large Scale Integration has been an incentive to concentrate persistently on device modeling. The fundamental properties which represent the basis for all device modeling activities are summarized. The sensible use of physical and technological parameters is discussed and the most important physical phenomena which are required to be taken into account are scrutinized. The assumptions necessary for finding a reasonable trade-off between efficiency and effort for a model synthesis are recollected. Methods to bypass limitations induced by these assumptions are pin-pointed. Simple and easy to use formulae for the physical parameters of major importance are presented. The necessity of a careful parameter-selection, based on physical information, is shown. Some glimpses on the numerical solution of the semiconductor equations are given. The discretisation of the partial differential equations with finite differences is outlined. Linearisation methods and algorithms for the solution of large sparse linear systems are sketched. Results of our two dimensional MOSFET model - MINIMOS - are discussed with typical applications. Much emphasis is laid on the didactic potential of such a complex high order model. In addition to its academic importance, the role of modeling as a tool to optimize transistor performance is stressed.

CONTENTS

1. INTRODUCTION

The first integrated circuits which just contained a few devices became commercially available in the early 1960's. Since that time an evolution has taken place so that today the manufacture of integrated circuits with over 400.000 transistors per single chip is possible. This advent of the so-called Very-Large-Scale-Integration (VLSI) certainly revealed the need of a better understanding of the basic device physics. The miniaturization of the single transistor, which is one of the inseparable preconditions of VLSI, brought about a collapse of the classical device models, because totally new phenomena emerged and even dominated the device behaviour. One consequence of this evidence led to an unimaginable number of suggestions of how to modify the classical models to incorporate various of the new phenomena. Additionally new activities have been initiated to explore the physical principles which make a device operationable. The number of scientific publications which utilize the terms "device analysis", "device simulation" and "device modeling" (c.f./4/, /53/, /83/) grew in an incredible manner.

At first it seems necessary to clarify these frequently used terms to facilitate the intelligibility of the subsequent chapters. Consulting a dictionary one will find among many more the following interpretations:

Analysis
- separation of a whole into its component parts, possibly with comment and judgement
- examination of a complex, its elements, and their relations in order to learn about

Simulation
- imitative representation of the functioning of one system or process by means of the functioning of another
- examination of a problem not subject to experimentation

Modeling
- to produce a representation or simulation of a problem or process
- to make a description or analogy used to help visualize something that cannot be directly observed

Therefore, analysis is at least intended to mean "Exact Analysis" and simulation must inferentially mean "Approximate Simulation" using only to some extent physically motivated models. Modeling is necessary for analysis and simulation, but with a different objective. However, any model should at least reflect the underlying physics.

The characteristic feature of early modeling was the separation of the interiour of the device into different regions, the treatment of which could be simplified by various assumptions like special doping profiles, complete depletion and quasineutrality. These separately treated regions were simply connected to produce the overall solution. If analytic results are intended, any other approach is prohibitive. Fully numerical modeling based on partial differential equations /172/ which describe all different regions of semiconductor devices in one unified manner was first suggested by Gummel /69/ for the one dimensional bipolar transistor. This approach was further developed and applied to pn-junction theory by De Mari /39/, /40/ and to IMPATT diodes by Scharfetter and Gummel /132/.

A two dimensional numerical analysis of a semiconductor device was carried out the first time by Kennedy and O'Brien /85/ who investigated the junction field effect transistor. Since then two dimensional modeling has been applied to fairly all important semiconductor devices. There are so many papers of excellent repute that it would be unfair to cite only a few. Recently also the first results on three dimensional device modeling have been published. The time dependence has been investigated by e.g. /96/, /107/ and models for three space dimensions have been announced by e.g. /25/, /185/, /186/.

In spite of all these important and successful activities, the need for economic and highly user oriented computer programs became more and more apparent in the field of device modeling. Especially for MOS devices which have evolved since their invention by Kahng and Atalla /82/ to an incredible standard, modeling has become inherently important because current flow controlled by a perpendicular field is an intrinsically two dimensional problem. One such program which has been applied successfully in many laboratories is called CADDET /165/. We have also tried to bridge that gap and developed MINIMOS /140/ for the two dimensional static analysis of planar MOS transistors.

In the next chapter the fundamental properties which are the basis for all device models are summarized. Much effort is laid on the documentation of various physical effects which possibly have to be taken into account when synthezising a device model for some special application. The assumptions which are usually made to ease modeling are presented and their validity is, at least qualitatively, discussed. Simple and easy to use formulae are presented which allow phenomenological simulation of the most important physical parameters with which the modelist has to deal.

Introduction

In the third chapter the numerical solution of the basic semiconductor equations is discussed. The two main methods for the solution of differential equations (i.e. finite differences and finite elements) are briefly compared. The discretisation of the general quasiharmonic equation is explained, because Poisson's equation as well as the continuity equations can be classified as quasiharmonic equations. A few linearisation schemes are presented and judged for adaequacy in terms of effort and efficiency. Classical algorithms for the solution of the sparse algebraic systems which are obtained by linearisation of the discrete semiconductor equations are explained.

The fourth chapter entirely deals with applications of MINIMOS. A didactic example which should make transparent the applicability of a high order model like MINIMOS for the analysis of device behaviour is given first. This example has been chosen to demonstrate the influence of ion implantation in the channel region of a very small scale MOSFET on threshold voltage and punch through. Secondly, an example of a process sensitivity analysis is presented which is an application ideally suited for numerical investigations. A more sophisticated application is shown next to demonstrate that it is possible to comprehend some complex interaction of different physical phenomena with device modeling. This is accomplished by explaining the reasons for the snap-back effect in the characteristics of a miniaturized MOSFET. The final sections deal with the analysis of coupled devices. For that purpose an n-Mos Inverter with a depletion load transistor and a CMOS inverter are investigated. These few examples are intended to stress the power and versatility of device modeling and its informative potential for the semiconductor industry.

This paper cannot intend to be any more than a summary with only moderate impact on details. The influence of process modeling, for instance, on the accuracy of the final result, important as it is, could not be dealt with. The reader with specific interest in any one subject will find the references useful for further information.

Throughout this paper all constants and quantities are given in the following units, if not specified otherwise: lengths in cm, times in s, temperature in K, voltages in V, currents in A. The units are often omitted to gain transparency in the formulae.

Introduction

2. SOME FUNDAMENTAL PROPERTIES

To accurately analyze an arbitrary semiconductor structure which is intended as a self-contained device under various operating conditions, a mathematical model has to be given. The equations which form this mathematical model are often called the fundamental semiconductor equations; these will be discussed in the first section of this chapter.

The second section will deal with assumptions which have to be made for special applications additionally to those which have already been used in the derivation of the equations and which are beyond the scope of this presentation. Furthermore, all quantities which are involved in the basic equations will be outlined more or less qualitatively.

It will become apparent that the fundamental equations employ a set of physical and technological parameters. An in-depth analysis of all these parameters is far from being finished at the moment – or the results of such an analysis are of overwhelming complexity – because of inherent methodical difficulties.

The third section will deal with additional assumptions which can be made to ease and speed up models for MOS-devices.

The topic of the fourth section of this chapter is the description of some suggestions for a heuristic simulation of the most important parameters based, as it were, on physical principles.

2.1 The Fundamental Semiconductor Equations

The most familiar model of carrier transport in a semiconductor device has been proposed by Van Roosbroeck /172/. It consists of Poisson's equation (2.1-1), the current continuity equations for electrons (2.1-2) and holes (2.1-3) and the current relations for electrons (2.1-4) and holes (2.1-5)

$$\text{div } \mathcal{E} \text{ grad } \Psi = -q \, (\, p - n + C \,) \qquad\qquad (2.1\text{-}1)$$

$$\text{div } \vec{J}_n = -q \, (\, G - R \,) \qquad\qquad (2.1\text{-}2)$$

$$\text{div } \vec{J}_p = q \, (\, G - R \,) \qquad\qquad (2.1\text{-}3)$$

$$\vec{J}_n = -q \, (\, \mu_n \, n \text{ grad } \Psi - D_n \text{ grad } n \,) \qquad\qquad (2.1\text{-}4)$$

$$\vec{J}_p = -q \, (\, \mu_p \, p \text{ grad } \Psi + D_p \text{ grad } p \,) \qquad\qquad (2.1\text{-}5)$$

The Fundamental Semiconductor Equations

These relations form a system of coupled partial differential equations. Poisson's equation, which is one of Maxwell's laws, describes the charge distribution in the interior of a semiconductor device. The balance of sinks and sources for electron- and hole currents is characterized by the continuity equations. The current relations describe the absolute value, direction and orientation of electron- and hole currents. The continuity equations and the current relations can be derived from Boltzmann's equation by not at all trivial means. The ideas behind these considerations cannot be presented here due to limited space. The interested reader should refer to /172/ and its secondary literature or text books on semiconductor physics e.g. /18/, /78/, /136/, /148/.

However, it is of prime importance to note that the equations (2.1-4) and (2.1-5) do not characterize effects which are caused by degenerate semiconductors (e.g. heavy doping). /97/, /171/, /174/ discuss some modifications of the current relations, which partially take into account the consequences introduced by degenerate semiconductors (e.g. invalidity of Boltzmann's statistics, bandgap narrowing). These modifications are not at all simple and lead to problems especially for the formulation of boundary conditions /116/, /173/. In case of modeling MOS devices, degeneracy is, owing to the relatively low doping in the channel region, practically irrelevant. For modern bipolar devices, though, bearing in mind shallow and extraordinarily heavily doped emitters, it is an absolute necessity to account for local degeneracy of the semiconductor.

Furthermore, (2.1-4) and (2.1-5) do not describe velocity overshoot phenomena which become apparent at feature lengths of $0.1 \mu m$ for silicon and $1 \mu m$ for gallium-arsenide /60/; and certainly no effects which are due to ballistic transport, the existence of which is still questionable /77/, are included. The latter start to become important for feature sizes below $0.01 \mu m$ for silicon and $0.1 \mu m$ for gallium-arsenide /61/. Considering the state of the art of device miniaturization, neither effect has to bother the modelists of silicon devices. For gallium-arsenide devices new ideas are mandatory for the near future /60/, /110/, /111/.

2.2 Assumptions and Discussion of Parameters

It is imperative to discuss the parameters of the semiconductor equations to get some insight into the complexity of that mathematical model and the difficulty of a more or less rigorous solution.

The permittivity ε in Poisson's equation in the most general

case is a rank two tensor. Because all common semiconductor
materials grow in cubic crystal structure and because silicon-
dioxide is amorphous no anisotropy exists and the permittivity
can be treated as a scalar quantity. Furthermore, one can savely
assume that the permittivity is homogenous with sufficient
accuracy for even degenerate semiconductors.

The electrically active net doping concentration C in
Poisson's equation is the most important technological parameter.
To obtain this quantity by mathematical analysis /51/ is at least
as cumbersome as to accurately analyze some semiconductor device,
because the physics of the technological processes which
determine the doping concentration still lacks basic
understanding. The need of modeling in this area is drastically
increasing in view of VLSI devices. One-dimensional process
modeling is fairly well established nowadays, but two-dimensional
simulation is just appearing /51/, /164/. Some glimpses of
modeling doping profiles with handy analytical expressions will
be given in section 2.4.1. One assumption which is usually made
with fairly satisfactory success is the total ionization of all
dopants (2.2-1).

$$C = N_D - N_A = N_D^+ - N_A^- \qquad (2.2-1)$$

As long as the Fermi level is separated several thermal
voltages from the impurity level, this assumption holds quite
nicely. For modern bipolar transistors, however, it certainly
becomes questionable for the emitter region (degenerate
material).

The electron density n and the hole density p are commonly
assumed to obey Boltzmann's statistics (2.2-2).

$$n = n_i \cdot e^{(\psi - \varphi_n)/U_T} \qquad p = n_i \cdot e^{(\varphi_p - \psi)/U_T} \qquad (2.2-2)$$

This assumption principally neglects degeneracy; but
moderate degeneracy can be included /55/ by introducing an
effective, doping dependent intrinsic number (2.2-3).

$$n_i = n_i(T,N) \qquad (2.2-3)$$

$$n_i(T,N) = n_i(T)\, e^{52.7(\ln(N/10^{17})+\sqrt{(\ln(N/10^{17}))^2 +0.5})/T}$$

$$n_i(T) = 3.88 \cdot 10^{16} \cdot T^{1.5} \cdot e^{-7000/T}$$

$$N = N_D + N_A$$

Assumptions and Discussion of Parameters

The temperature dependence of the intrinsic number is based on the influence of the effective carrier masses and the bandgap. More elaborate formulae for these effects which might be imperative for low temperature applications can be found in /62/. The formula for bandgap narrowing in (2.2-3) was first suggested by Slotboom /146/. For a doping concentration of $1.3 \cdot 10^{17}$ cm^{-3} the intrinsic number has already increased by twenty percent.

The mobility of electrons μ_n and holes μ_p is in principle a rank two tensor function of many arguments. One ends up with a "so called" mobility after averaging and combining various physical mechanisms which are still not analyzed thouroghly enough to be modeled satisfactorily /79/. Some formulae for a mobility model for silicon will be summarized in section 2.4.2.

Another assumption which is unfortunately not at all free of doubts is the validity of the Einstein-Nernst relations (2.2-4).

$$D_n = \mu_n \cdot U_T \qquad D_p = \mu_p \cdot U_T \qquad (2.2-4)$$

Some guidelines on how to extend these relations for degenerate material are given in e.g. /8/. It is important to remember that the current relations (2.1-4) and (2.1-5) do not differentiate between lattice temperature and electron temperature. Therefore, if one has to deal with hot electrons in a precise manner, the current relations have to be updated; in particular the mathematical structure of the diffusion current term has to be refined.

The last parameter which remains to be dealt with for a qualitative characterization is the net generation/recombination rate (G-R) in (2.1-2) and (2.1-3). This quantity has to describe a number of physical processes which are responsible for generation/recombination of electron-hole pairs. These processes and their interactions are also not analyzed to a satisfactory level so that one has to use heuristic expressions for a model which is at least plausible in the underlying physics. Some suggestions for these formulae will be given in section 2.4.3.

2.3 Additional Assumptions for MOS-Models

The fundamental semiconductor equations describe the internal behavior of any semiconductor device. However, for certain devices these equations may be simplified without significant loss of accuracy. As the MOSFET is a minority carrier device, the current is given mainly by the continuity equation of one carrier type. If avalanche is neglected, only little carrier generation occurs in the MOSFET.

Therefore, the eqs. (2.1-2)-(2.1-3) may be rewritten as

$$\text{div } \vec{J}_n = 0 \tag{2.3-1}$$

$$\vec{J}_p = 0 \tag{2.3-2}$$

for the n-channel device and

$$\text{div } \vec{J}_p = 0 \tag{2.3-3}$$

$$\vec{J}_n = 0 \tag{2.3-4}$$

for the p-channel device. However, it should be kept in mind that these assumptions are valid only if the avalanche effect is neglected.

The channel width of an MOSFET is usually much larger than the depletion widths. As a consequence the partial derivatives in that direction can be neglected and the semiconductor equations reduce to two dimensions. The neglection of the derivative of the potential in source-drain direction is a proper assumption only for long-channel devices. The so called "gradual-channel approximation" was the basis of a lot of one-dimensional models. These models fail to predict accurately the behavior of modern miniaturized devices.

If the avalanche effect should be included, the assumptions (2.3-1)-(2.3-4) are no longer valid and both continuity equations have to be solved with inhomogeneity terms. As a consequence, the ionization-generated majority carriers (holes for an n-channel MOSFET) flow to the substrate as they are repelled from the source and drain junctions. There are several options to account for the voltage drop which is induced by the substrate current: (a) a truely three-dimensional analysis; (b) extension of the simulation over the entire bulk area; (c) extension of the two-dimensional simulation over the depletion region and using an (effective) bulk resistor (Fig. 2.3-1). If one wants to avoid excessive computing time associated with (a), option (c) is to be preferred because it allows inclusion of current spread into the third dimension and, also, consumes less computing time than (b). In that way the voltage drop across the parasitic bulk resistor simulates a more positive bulk bias and, if large enough, is able to forward-bias the parasitic bipolar npn transistor (according to source, bulk, and drain). This causes a larger drain current and facilitates the breakdown which then will occur at smaller drain voltages /133/.

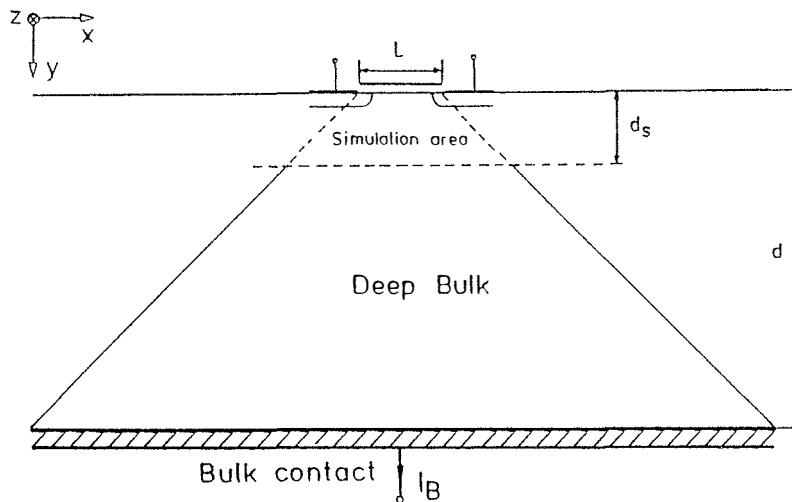Additional Assumptions for MOS-Models

Fig. 2.3-1: Current flow in deep bulk

In the following we should like to suggest an easy method to estimate the value of the bulk resistor. It is assumed that the current spreads at an angle of 45 degrees /15/ into both directions perpendicular to its flow (x- and z- direction in Fig. 2.3-1). This assumption is arbitrary but not implausible, and, furthermore, if we neglect any diffusion current, we obtain the following expression for the electric field in the deep substrate.

$$\frac{d\Psi}{dy} = \frac{I_B}{KA} = \frac{I_B}{K(L+2y)(W+2y)}$$ (2.3-5)

with K standing for the conductivity of the substrate and A the area of the current flow. L and W are channel length and channel width, respectively. Integrating this equation along y from the end of the simulation area d_s to the bulk contact we obtain

$$R_{Bulk} = \frac{\int_{d_s}^{d} \frac{d\Psi}{dy}\, dy}{I_B} = \frac{1}{2K(W-L)}\left(\ln(\frac{L+2d}{L+2d_s}) - \ln(\frac{W+2d}{W+2d_s}) \right) .$$ (2.3-6)

For L=W this equation simplifies to

$$R_{Bulk} = \frac{d-d_s}{K(L+2d)(L+2d_s)} .$$ (2.3-7)

This calculation is fairly crude compared to the elaborate solution of the basic equations. However, any more precise calculation would be very complicated and the present method is sufficient to investigate the influence of the parasitic bulk resistance at least qualitatively.

Additional Assumptions for MOS-Models

2.4 Models of Physical Parameters

2.4.1 Formulae for Modeling Doping Profiles

A one dimensional doping profile which can be calculated fairly accurately with a process simulation program (e.g. /6/) may be heuristically converted to two dimensions for a structure with an ideal oxide mask as shown in Fig. 2.4-1 using (2.4-1).
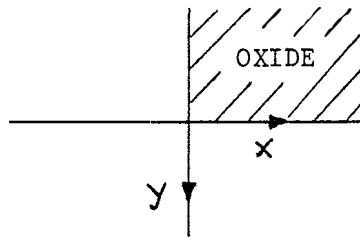

Fig. 2.4-1 Coordinate nomenclature for an ideal oxide mask

$$C(x,y) = C(\sqrt{y^2 + max(x/f,0)^2}) \qquad (2.4-1)$$

This formula is extraordinarily simple to use and needs only one fitting parameter: f which controls the amount of lateral diffusion. For most applications f lies in the range of 0.5 to 0.9. An elliptic rotation at x=0 (c.f. Fig. 2.4-1) of the one-dimensional profile is performed to obtain the doping concentration below the oxide mask. Out-diffusion effects which occur near the mask edge are not at all taken into account.

Lee /89/, /90/ recently published expressions which are still fairly simple to use, but which are based on more physical reasoning. (2.4-2) can be used for the simulation of a predeposition step. Ld denotes the diffusion length; D: diffusion constant, t: diffusion time, N_s: desired surface concentration.

$$Ld = 2 \cdot \sqrt{D \cdot t} \qquad (2.4-2)$$

$$C_p(x,y) = 0.5 \cdot N_s \cdot e^{-(y/Ld)^2} \cdot erfc(x/Ld)$$

The distribution of implanted ions under mask edges has also been investigated extensively e.g. /128/. The formulae (2.4-3) which have been taken from /89/, /90/ allow simulation of diffusion with an initial ion-implantation. Rp denotes the projected range, ΔRp: projected standard deviation, Dose: implantation dose.

$$a = (2 + (Ld/\Delta Rp)^2)^{-1/2} \qquad (2.4-3)$$

$$K(y) = e^{-(a \cdot (Rp-y)/\Delta Rp)^2} \cdot \text{erfc}(-a \cdot ((Rp/\Delta Rp)+\sqrt{2 \cdot y/Ld}))$$

$$C_i(x,y) = (a/(4 \cdot \Delta Rp \cdot \sqrt{\pi})) \cdot \text{Dose} \cdot (K(y)+K(-y)) \cdot \text{erfc}(x/Ld)$$

In the derivation of (2.4-2) and (2.4-3) it is assumed that the diffusion "constant" is really constant. This limits the application to relatively low peak values of the implanted profile. For high peak values one might fit the diffusion lengths Ld to obtain a desired junction depth.

The diffusion constant D can be estimated, again for fairly low concentrations, with the classical exponential law (2.4-4).

$$D = D_0 \cdot e^{T_a/T} \qquad (2.4-4)$$

Element	$D_0/(cm^2 s^{-1})$	$T_a/(K)$
B	0.5554	$-3.975 \cdot 10^4$
P	3.85	$-4.247 \cdot 10^4$
Sb	12.9	$-4.619 \cdot 10^4$
As	24.	$-4.735 \cdot 10^4$

The projected range parameters Rp and ΔRp which are nonlinear functions of the implantation energy can be looked up in standard tables /64/. These tables are principally tedious to implement in computer programs, so that one might prefer some polynomial fit (2.4-5); x denotes here the implantation energy.

$$Rp = \sum_{i=1}^{n} a_i \cdot x^i \quad (\mu m) \qquad (2.4-5)$$

$$\Delta Rp = \sum_{i=1}^{n} b_i \cdot x^i \quad (\mu m)$$

Models of Physical Parameters

The coefficients for such polynomials are given in Fig. 2.4-2 for Rp in silicon, in Fig. 2.4-3 for ΔRp in silicon and in Fig. 2.4-4 for Rp in silicon-dioxide.

Element	B	P	Sb	As
a_1	$3.338 \cdot 10^{-3}$	$1.259 \cdot 10^{-3}$	$8.887 \cdot 10^{-4}$	$9.818 \cdot 10^{-4}$
a_2	$-3.308 \cdot 10^{-6}$	$-2.743 \cdot 10^{-7}$	$-1.013 \cdot 10^{-5}$	$-1.022 \cdot 10^{-5}$
a_3		$1.290 \cdot 10^{-9}$	$8.372 \cdot 10^{-8}$	$9.067 \cdot 10^{-8}$
a_4			$-3.056 \cdot 10^{-10}$	$-3.442 \cdot 10^{-10}$
a_5			$4.028 \cdot 10^{-13}$	$4.608 \cdot 10^{-13}$

Fig. 2.4-2 Coefficients for Rp in silicon

Element	B	P	Sb	As
b_1	$1.781 \cdot 10^{-3}$	$6.542 \cdot 10^{-4}$	$2.674 \cdot 10^{-4}$	$3.652 \cdot 10^{-4}$
b_2	$-2.086 \cdot 10^{-5}$	$-3.161 \cdot 10^{-6}$	$-2.885 \cdot 10^{-6}$	$-3.820 \cdot 10^{-6}$
b_3	$1.403 \cdot 10^{-7}$	$1.371 \cdot 10^{-8}$	$2.311 \cdot 10^{-8}$	$3.235 \cdot 10^{-8}$
b_4	$-4.545 \cdot 10^{-10}$	$-2.252 \cdot 10^{-11}$	$-8.310 \cdot 11^{-10}$	$-1.202 \cdot 10^{-10}$
b_5	$5.525 \cdot 10^{-13}$		$1.084 \cdot 10^{-13}$	$1.601 \cdot 10^{-13}$

Fig. 2.4-3 Coefficients for ΔRp in silicon:

Element	B	P	Sb	As
a_1	$3.258 \cdot 10^{-3}$	$9.842 \cdot 10^{-4}$	$7.200 \cdot 10^{-4}$	$7.806 \cdot 10^{-4}$
a_2	$-2.113 \cdot 10^{-6}$	$-2.240 \cdot 10^{-7}$	$-8.054 \cdot 10^{-6}$	$-7.899 \cdot 10^{-6}$
a_3			$6.641 \cdot 10^{-8}$	$7.029 \cdot 10^{-8}$
a_4			$-2.422 \cdot 10^{-10}$	$-2.653 \cdot 10^{-10}$
a_5			$3.191 \cdot 10^{-13}$	$3.573 \cdot 10^{-13}$

Fig. 2.4-4 Coefficients for Rp in silicon-dioxide

The maximum error of the projected range parameters calculated with these coefficients and (2.2-5) is in the energy range of 5keV to 300keV only a few percent compared to /64/. More data are given in /141/.

Models of Physical Parameters

If an implantation is performed through an oxide, the projected range in the semiconductor has to be reduced /129/ e.g. with (2.4-6).

$$Rp = Rp_{se} \cdot (1 - T_{iox}/Rp_{ox})$$

(2.4-6)

T_{iox} denotes the thickness of the oxide, Rp_{se}/Rp_{ox}: projected range in semiconductor/oxide.

2.4.2 Formulae for Mobility Modeling

The mobility of carriers is, as already mentioned, an eminently complex quantity. Additionally it is an important parameter, because all errors in the mobility a proportional error of the current through the multiplicative dependence. This is certainly one of the primary results any model should yield reliably. The formulae which will be given below describe phenomenologically the mobility in silicon; the subscripts n and p denote electrons and holes, respectively.

To model mobility at least plausibly, several scattering mechanisms have to be taken into account, the basis of which is lattice scattering. This effect can be described by a simple power law /79/, /136/ in dependence of temperature (2.4-7).

$$\mu_L(T) = A \cdot T^{-g} \quad (cm^2/Vs)$$

(2.4-7)

$$A_n = 7.12 \cdot 10^8 \qquad\qquad A_p = 1.35 \cdot 10^8$$
$$g_n = 2.3 \qquad\qquad\qquad g_p = 2.2$$

The pure lattice mobility is reduced through the scattering processes at ionized impurities. (2.4-8) is a well established formula which models temperature dependent ionized impurity scattering /24/ and electron-hole scattering /55/. The latter is extremely important in low doped regions where high injection takes place.

$$\mu_{LI}(N,T) = \mu_L(T) \cdot a + \mu_{min} \cdot (1 - a) \quad (cm^2/Vs)$$

(2.4-8)

$$a = \frac{1}{1 + (T/300)^b \cdot (N/N_0)^c}$$

$$N = 0.67 \cdot (N_D^+ + N_A^-) + 0.33 \cdot (n + p)$$

$$\mu_{minn} = 55.24 \qquad\qquad \mu_{minp} = 49.7$$
$$b_n = -3.8 \qquad\qquad\qquad b_p = -3.7$$
$$c_n = 0.73 \qquad\qquad\qquad c_p = 0.7$$
$$N_{0n} = 1.072 \cdot 10^{17} \qquad\qquad N_{0p} = 1.606 \cdot 10^{17}$$

Models of Physical Parameters

Similar expressions which have been partly deduced from measurement and/or theory have been presented in /7/, /41/, /47/, /92/, /132/.

To properly simulate the mobility in MOS transistors, one has to deal with surface roughness and field dependent surface scattering. /30/, /130/, /153/ presented interesting measured results on inversion layer mobility; /162/, /163/ gave some excellent ideas on how to treat theoretically these and other scattering mechanisms; /182/ suggested a heuristic formula for field dependent surface scattering which is applicable for two-dimensional simulations, but whose adequacy is questioned in /162/. However, we have developed (2.4-9) which models phenomenologically with best fit to measurement surface roughness as well as field dependent surface scattering /143/.

$$\mu_{LIS}(y,E_p,E_t,N,T) = \mu_{LI}(N,T) \cdot \frac{y+y_r}{y+b \cdot y_r} \quad (cm^2/Vs) \qquad (2.4-9)$$

$$y_r = y_0/(1+E_p/E_{p0})$$

$$b = 2+E_t/E_{t0}$$

$$E_p = \max(0,(E_x \cdot J_x+E_y \cdot J_y)/(J_x^2+J_y^2)^{1/2})$$

$$E_t = \max(0,(E_x \cdot J_y-E_y \cdot J_x) \cdot J_x/(J_x^2+J_y^2))$$

$$y_{0n} = 5 \cdot 10^{-7} \qquad\qquad y_{0p} = 4 \cdot 10^{-7}$$

$$E_{p0n} = 10^4 \qquad\qquad E_{p0p} = 8 \cdot 10^3$$

$$E_{t0n} = 1.8 \cdot 10^5 \qquad\qquad E_{t0p} = 3.8 \cdot 10^5$$

In regions with a high electric field component parallel to current flow, drift velocity saturation has to be taken into account. (2.4-10) combines, also phenomenologically, this physical effect and the lattice-impurity-surface mobility using a Mathiessen-type rule with a weakly temperature dependent saturation velocity /23/, /79/, /80/.

$$\mu_{tot}(y,E_p,E_t,N,T) = (\mu_{LIS}(...)^{\beta}+(v_s/E_p)^{\beta})^{1/\beta} \qquad (2.4-10)$$

$$v_{sn} = 1.53 \cdot 10^9 \cdot T^{-0.87} \qquad\qquad v_{sp} = 1.62 \cdot 10^8 \cdot T^{-0.52}$$

$$\beta_n = -2 \qquad\qquad\qquad \beta_p = -1$$

Models of Physical Parameters

2.4.3 Formulae for Modeling Generation/Recombination

To simulate satisfactorily transfer phenomena of majority carrier current and minority carrier current in just a simple diode, it is an absolute necessity to model carrier recombination and generation as carefully as possible. (2.4-11) represents the well known Shockley-Read-Hall term for modeling thermal generation/recombination. The carrier lifetimes can be simulated as being doping dependent /35/, /103/.

$$(\text{ G } - \text{ R })_{th} = \frac{n_i^2 - p \cdot n}{\tau_n(p+p_1)+\tau_p(n+n_1)} \qquad (1/cm^3 s) \qquad (2.4-11)$$

$$\tau_n = 3.95 \cdot 10^{-5}/(1+N/7.1 \cdot 10^{15}) \qquad \tau_p = 3.52 \cdot 10^{-5}/(1+N/7.1 \cdot 10^{15})$$

Surface generation/recombination /74/ can be treated in a fairly similar manner by (2.4-12).

$$(\text{ G } - \text{ R })_s = \frac{n_i^2 - p \cdot n}{(p+p_1)/s_n + (n+n_1)/s_p} \cdot \delta(y) \qquad (1/cm^3 s) \qquad (2.4-12)$$

$\delta(y)$: Dirac-Delta function, $y=0$ denotes an interface

$$s_n = 100 \qquad\qquad s_p = 100$$

Impact ionization can be modeled by an exponentially field dependent generation term /27/, /28/. The constants in (2.4-13) are essentially taken from /170/.

$$G_a = \frac{|\vec{J}_n|}{q} A_n \exp \left(- \frac{B_n|\vec{J}_n|}{\vec{E} \cdot \vec{J}_n} \right) +$$

$$+ \frac{|\vec{J}_p|}{q} A_p \exp \left(- \frac{B_p|\vec{J}_p|}{\vec{E} \cdot \vec{J}_p} \right) \qquad (1/cm^3 s) \qquad (2.4-13)$$

$$A_n = 7 \cdot 10^5 \qquad\qquad A_p = 1.588 \cdot 10^6$$
$$B_n = 1.23 \cdot 10^6 \qquad\qquad B_p = 2.036 \cdot 10^6$$

Models of Physical Parameters

It should be noted that this form of simulating avalanche is relatively crude compared to more exact considerations, but the underlying physical principles are so complex that a trade-off in accuracy and complexity leads to that type of formula. The ionization probabilities $\alpha_{n,p}$ for silicon as a function of the electric field have been measured by various authors: Mc Kay /101/, /102/, Miller /105/, /106/, Chynoweth /27/, /28/, Lee /88/, Moll /112/, /113/, Ogawa /118/, Van Overstraeten /170/, Grant /65/, Dalal /36/. Their results are summarized in Fig. 2.4-5 for electrons and in Fig. 2.4-6 for holes. Additionally, the measured results are compared to theoretical results of Baraff /10/ (material constants from Sze /157/, /158/). Also drawn in Fig. 2.4-5 and Fig. 2.4-6 are theoretical limits published by Okuto /122/, /123/, which imply that all the energy the carriers can obtain from the electric field is used to generate additional carriers. Furthermore, the energy loss per single ionization has been taken to be 1.6eV for electrons and 1.8eV for holes (see also /75/). A more concise treatment of the ionization probabilities has been undertaken theoretically by /5/, /26/, /91/, /145/, /160/, /161/, /162/, /169/, /181/ and experimentally by /95/, /131/, /149/.

To analyze high injection conditions, Auger recombination has to be included as an antagonism to avalanche generation. Already the use of a simple formula like (2.4-14) in general gives satisfactory results /31/, /35/, /52/, /55/.

$$(G - R)_{Aug} = (n_i^2 - p{\cdot}n) (C_n{\cdot}n + C_p{\cdot}p) \qquad (1/cm^3 s) \quad (2.4-14)$$

$$C_n = 2.8{\cdot}10^{-31} \qquad\qquad C_p = 9.9{\cdot}10^{-32}$$

Finally, all generation/recombination phenomena have to be combined to one total quantity. The usual way to do so is to simply sum up all terms (2.4-15). However, that means that no interaction of the different phenomena does exist.

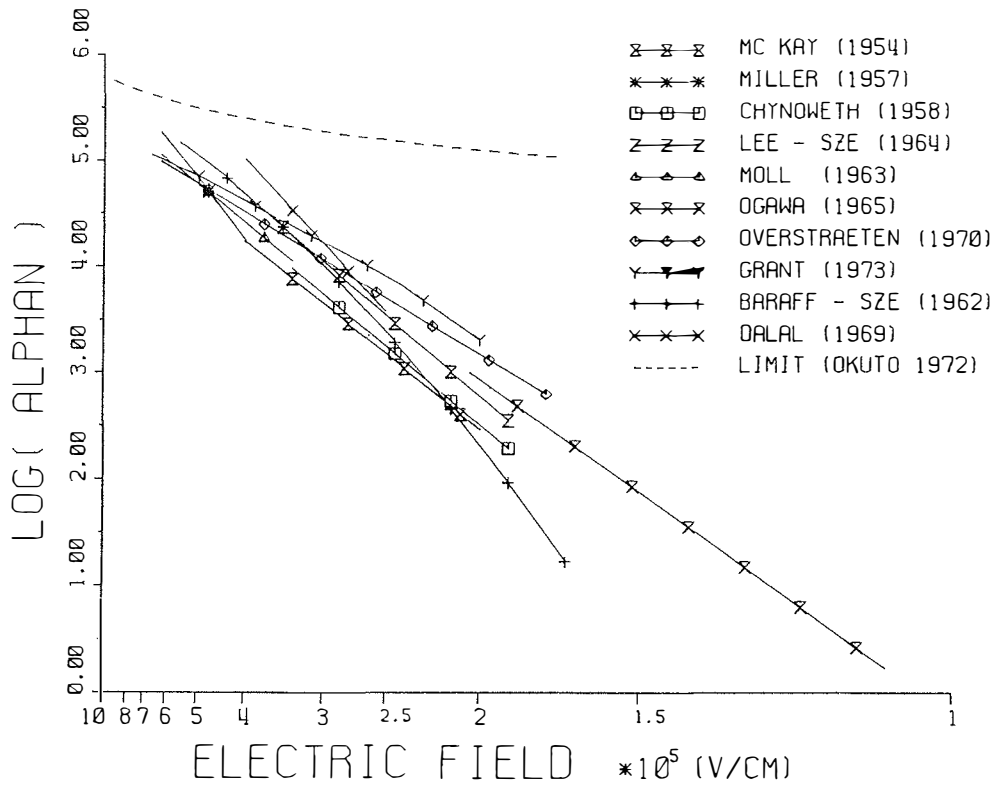$$(G-R)_{tot} = (G-R)_{th} + (G-R)_{s} + (G-R)_{Aug} + G_a \qquad (2.4-15)$$

Models of Physical Parameters

Fig. 2.4-5: Ionization probabilities for electrons
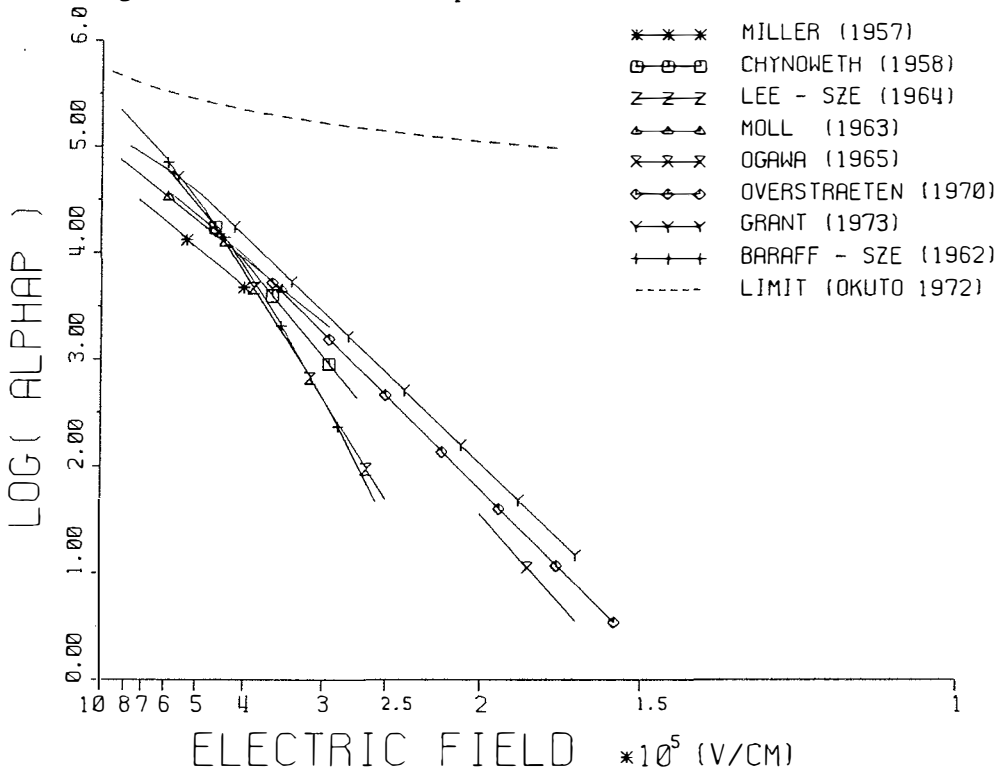
Fig. 2.4-6: Ionization probabilities for holes

Models of Physical Parameters

3. NUMERICAL SOLUTION OF SEMICONDUCTOR EQUATIONS

The major difficulty in designing a high order numerical model of a semiconductor device is the adaption of adequate numerical methods for the solution of the basic semiconductor equations and their associated, often very complex, physical parameters, as outlined in the previous chapter.

In section 3.1 we should like to discuss the discretisation of the basic equations. The classical method of replacing derivatives with finite differences will be explained. The last part of that section will deal with automatic and adaptiv mesh generation which is a task of primary importance for user oriented models, but which has as yet not been scrutinized thoroughly.

The linearization of discrete equations will be treated in section 3.2 with some emphasis on the severly strong nonlinearity of the semiconductor equations. For that purpose some modified Newton schemes are presented which yield an incredible gain in computer efficiency.

Algorithms for the solution of the linearized discrete equations are discussed in section 3.3. A review of the most attractive methods for linear systems with special sparsity structure is given and also some cautious judgement is ventured.

3.1 Discretisation of Semiconductor Equations with Finite Differences

Unfortunately, the basic semiconductor equations cannot be solved in closed form by analytical methods. To utilize a numerical method, first of all the domain in which a solution is wanted has to be split into a finite number of small parts. These parts have to be sufficiently small so that all dependent variables of the basic equations behave like some arbitrarily chosen, but nevertheless simple functions; the equations have to be discretized. However, one should always bear in mind that one can, following the above sketched outline, obtain only an exact solution of the discretized problem, which is just an approximate solution of the analytically formulated equations. The difference between the discrete solution and the solution of the real problem depends obviously on the partitioning of the domain and the selection of the approximating functions.

There exist basically two classical methods for obtaining algebraic equations, which approximate the differential equation and which can be solved numerically, namely: the Finite Difference method and the Finite Element method. The fundamental

Discretisation of Semiconductor Equations

difference between these two methods can be summarized, at least
qualitatively, as follows. By applying the finite difference
method, all derivatives in the differential equation are replaced
by finite differences between discrete points in the interiour of
the domain and the residual of the resulting difference equation
is set to zero on every discrete point. The finite element
method in its residual formulation demands that the weighted
residuum integrated over the whole domain be zero. This can be
achieved algebraically by setting all residual integrals for
every finite element for which the solution is assumed to obey
some simple functional relation to zero. From our point of view
it is impossible to favourize one method distinctly; both methods
have their advantages and bottlenecks. Following the literature
many renowned authors have concentrated their work on finite
elements e.g. /1/, /13/, /20/, /21/, /22/, /25/, /32/, /70/,
/71/, /119/, or finite differences e.g. /56/, /66/, /67/, /76/,
/84/, /85/, /93/, /94/, /107/, /108/, /109/, /121/, /154/. We
have also concentrated our activities on finite differences,
because it looks as if the mathematical background required to
produce a running program were somewhat smaller for the finite
difference method than for the finite element method. Some
interesting extensions of the finite difference method have been
recently proposed by Adler /2/, /3/. When fully utilizing these
ideas, one advantage of the finite element method, high
flexibility at the partitioning task, should also be reached with
the finite difference method.

We should like to explain the discretisation with
five-point-star differences, which is probably the best known
approach of the finite difference method for two dimensional
partial differential equations (PDEs). The domain in which the
solution of a PDE is desired is first partitioned into small
areas by grid lines parallel to some arbitrary coordinate system.
For the sake of simplicity a rectangular domain and a cartesian
coordinate system will be assumed. By laying NX vertical grid
lines (parallel to y-axis) and NY horizontal grid lines (parallel
to x-axis) one gets NX·NY intersections. On these intersections
one wants to obtain an approximate solution of the PDE of
sufficient accuracy. For that purpose the PDE is replaced on
every inner point (i,j) (see Fig. 3.1-1) by a difference equation
which uses the inner point (i,j) and its four nearest neighbours
(i+1,j), (i-1,j), (i,j+1) and (i,j-1). The major assumption for
the derivation of the difference equation is that the solution
can be approximated with a piecewise linear function along the
verteces between the inner point (i,j) and its neighbours. Thus
one gets (NX-2)·(NY-2) difference equations because that is
exactly the number of inner points. At the boundary of the
domain the solution of the PDE has to fulfill some boundary
conditions from which one can obtain equations for the boundary

Discretisation of Semiconductor Equations

points in a similiar manner; there exist $2 \cdot (NX+NY-2)$ boundary points. The total number of equations equals, therefore, the total number of points and a unique solution can be found.
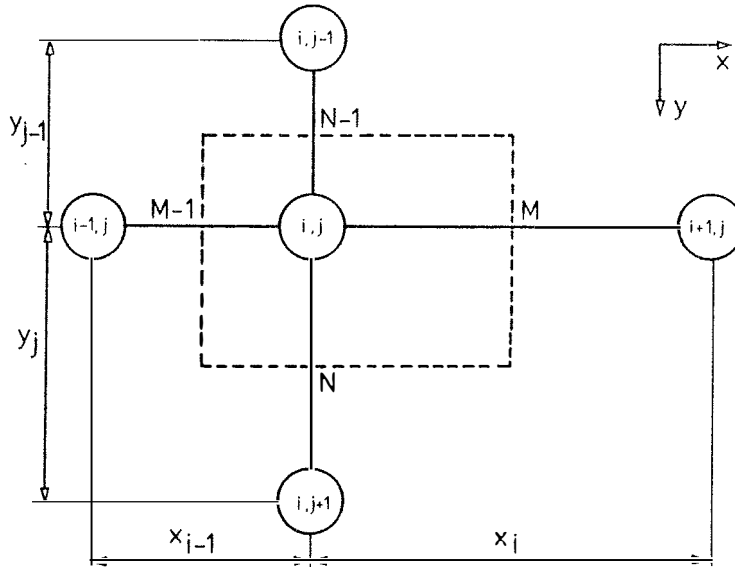

Fig. 3.1-1: The index convention used

In the next section the discretisation of the quasiharmonic equation in a precise manner will be dealt with, because the linearized forms of Poisson's equation as well as continuity equations belong to this important category of PDEs and many publications have been written on that subject e.g. /58/, /59/, /98/, /147/.

3.1.1 The Quasiharmonic Equation

Let G be a finite domain in the (x,y) plane bounded by R which is piecewise continuously differentiable. Furthermore, let the functions $P(x,y)$, $S(x,y)$ and $F(x,y)$ be piecewise continuous in G. $P(x,y)$ be positive and not vanishing anywhere; $S(x,y)$ be positive or zero. Then (3.1-1) represents the quasiharmonic equation with solution $u(x,y)$.

$$\text{div}(P(x,y) \cdot \text{grad}(u(x,y))) - S(x,y) \cdot u(x,y) = F(x,y) \qquad (3.1-1)$$

subject to the boundary conditions:

$$A(x,y) \cdot u(x,y) + B(x,y) \cdot u(x,y)_n = C(x,y) \qquad (3.1-2)$$

Discretisation of Semiconductor Equations

where $A(x,y)$, $B(x,y)$ and $C(x,y)$ are defined in R, piecewise continuous and positive or zero, and $A(x,y)+B(x,y)$ is not vanishing anywhere. $u(x,y)_n$ denotes the derivate of $u(x,y)$ perpendicular to the boundary.

For a solution of this problem the differential equation has to be integrated in every area g_{ij}, obtained by partitioning as outlined above, around the inner point (i,j). The area g_{ij} is drawn with dashed lines in Fig. 3.1-1; it is represented by the rectangle around point (i,j).

$$\iint_{g_{ij}} \text{div}(P \cdot \text{grad}(u)) \cdot dx \cdot dy - \iint_{g_{ij}} S \cdot u \cdot dx \cdot dy = \iint_{g_{ij}} F \cdot dx \cdot dy \qquad (3.1\text{-}3)$$

Using Green's theorem, the area integral can be transformed into a closed boundary integral around g_{ij}.

$$\iint_{g_{ij}} \text{div}(P \cdot \text{grad}(u)) \cdot dx \cdot dy = \int_{r_{ij}} (P \cdot (\partial u/\partial x) \cdot dy - P \cdot (\partial u/\partial y) \cdot dx) \qquad (3.1\text{-}4)$$

Let x_i be the geometrical distance between the i.th and i+1.st vertical grid line and y_j the distance between the j.th and j+1.st horizontal grid line (cf. Fig. 3.1-1). Let P_M be the value of function $P(x,y)$ at point M which is placed exactly between points (i,j) and $(i+1,j)$; and assume the analogous relations for P_{M-1}, P_N and P_{N-1}, which can be easily made clear with Fig. 3.1-1. Then the following holds:

$$\int_{r_{ij}} (P \cdot (\partial u/\partial x) \cdot dy - P \cdot (\partial u/\partial y) \cdot dx) =$$

$$= 0.5 \cdot (y_j + y_{j-1}) \cdot (P_M \cdot (u_{i+1,j} - u_{i,j})/x_i +$$

$$+ \qquad P_{M-1} \cdot (u_{i-1,j} - u_{i,j})/x_{i-1}) +$$

$$+ 0.5 \cdot (x_i + x_{i-1}) \cdot (P_N \cdot (u_{i,j+1} - u_{i,j})/y_j +$$

$$+ \qquad P_{N-1} \cdot (u_{i,j-1} - u_{i,j})/y_{j-1}) +$$

$$+ o(x_{i-1} + x_i) + o(y_{j-1} + y_j) \qquad (3.1\text{-}5)$$

Discretisation of Semiconductor Equations

The second and third integral of Eq. 3.1-3 can be approximated straightforwardly under the assumptions that the functions $S(x,y)$ and $F(x,y)$ and the solution $u(x,y)$ are sufficiently smooth in the area g_{ij}.

$$\iint\limits_{g_{ij}} S \cdot u \cdot dx \cdot dy \doteq 0.25 \cdot S_{i,j} \cdot u_{ij} \cdot (x_i + x_{i-1}) \cdot (y_j + y_{j-1}) \quad (3.1-6)$$

$$\iint\limits_{g_{ij}} F \cdot dx \cdot dy \doteq 0.25 \cdot F_{ij} \cdot (x_i + x_{i-1}) \cdot (y_j + y_{j-1}) \quad (3.1-7)$$

After combining (3.1-5), (3.1-6) and (3.1-7) and separating the unknowns, one obtains for each inner point (i,j) a linear equation of the following form:

$$
\begin{aligned}
u_{i,j} \quad &\cdot ((y_j + y_{j-1}) \cdot (P_M/x_i + P_{M-1}/x_{i-1}) + \\
&+ (x_i + x_{i-1}) \cdot (P_N/y_j + P_{N-1}/y_{j-1}) + \\
&+ 0.5 \cdot S_{i,j} \cdot (x_i + x_{i-1}) \cdot (y_j + y_{j-1})) = \\
= u_{i+1,j} &\cdot ((y_j + y_{j-1}) \cdot P_M/x_i) + \\
+ u_{i-1,j} &\cdot ((y_j + y_{j-1}) \cdot P_{M-1}/x_{i-1}) + \\
+ u_{i,j+1} &\cdot ((x_i + x_{i-1}) \cdot P_N/y_j) + \\
+ u_{i,j-1} &\cdot ((x_i + x_{i-1}) \cdot P_{N-1}/y_{j-1}) - \\
- 0.5 \cdot F_{i,j} &\cdot (x_i + x_{i-1}) \cdot (y_j + y_{j-1})
\end{aligned}
\quad (3.1-8)
$$

In Eq. 3.1-8 no estimate of the discretisation error is given. For a non-aequidistant mesh $(x_i \neq x_{i-1}, \; y_j \neq y_{j-1})$ the discretisation error decreases approximately linearly with the mesh spacings. Some ideas on proper mesh selection to sufficiently bound that error will be given in section 3.1.4. However, more exact consideration should be looked up in the classical mathematical literature e.g. /58/, /147/.

The discretisation of the boundary conditions is basically no problem. It is treated quite carefully in many lecture books on numerical mathematics e.g. /47/ so that we can refrain from an explanation.

Discretisation of Semiconductor Equations

All equations obtained by the discretisation procedure can be combined to a sparse linear operator B (3.1-9) applied to a vector u of unknowns which represent the solution at the mesh points.

$$B(u) = 0 \qquad\qquad\qquad (3.1-9)$$

Therefore, the rank of the operator B equals the total number of meshpoints which is usually rather large. However, B is also very sparse; there exist at most five elements per row. The treatment of these specially structured equations will be outlined in the following sections.

3.1.2 Poisson's Equation

Poisson's equation (3.1-10) is an exponentially nonlinear elliptic equation.

$$\text{div } \boldsymbol{\varepsilon} \text{ grad } \boldsymbol{\Psi} = -q \cdot (n_i \cdot e^{(\boldsymbol{\varphi}_p - \boldsymbol{\Psi})/Ut} - n_i \cdot e^{(\boldsymbol{\Psi} - \boldsymbol{\varphi}_n)/Ut} + C) \qquad (3.1-10)$$

The geometry for the simulation of MOS transistors which we and many others use is shown in Fig. 3.1-2. Poisson's equation (also the continuity equations) has to be solved for the rectangular area A-F-G-H which represents the silicon region. In the area C-D-E-B which represents the gate oxide, only the Laplacian equation has to be solved because no space charge exists there. The boundary conditions are usually treated as follows: The contacts (A-B: source, E-F: drain, G-H: bulk) are assumed to be ideally ohmic. The potential is kept constant at the sum of the applied bias plus the built-in potential which is caused by the doping. At the vertical boundaries (A-H, F-G) the derivative of the potential perpendicular to the boundary (i.e. the lateral electrical field component) has to be zero. Certainly, this condition is only valid from the physical point of view if the source contact A-H and the drain contact E-F are sufficiently long. At the silicon to silicon dioxide interface the potential must obey Gauß's law (3.1-11). The existence of fixed surface states can be treated directly with Gauß's law, confer to /155/; however, we think it is more economic and sufficiently accurate to account for fixed surface states with the flatband voltage, because fixed surface states should be kept small anyway and, should, therefore, not effect the solution very much.

$$\boldsymbol{\varepsilon}_{ox} \cdot (\partial \boldsymbol{\Psi} / \partial y)_{ox} = \boldsymbol{\varepsilon}_{si} \cdot (\partial \boldsymbol{\Psi} / \partial y)_{si} \qquad , \qquad (3.1-11)$$
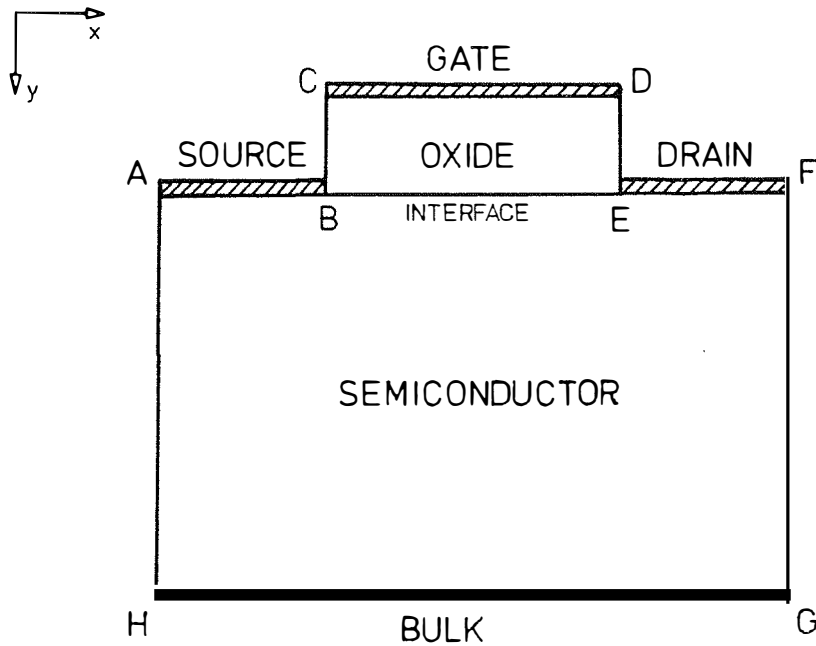
Discretisation of Semiconductor Equations

Fig. 3.1-2: The simulation geometry for planar MOSFETs

The Laplacian equation in the oxide is coupled with Poisson's equation via (3.1-11). At the gate contact (C-D) the potential is kept constant at the applied bias minus the flatband voltage; at the vertical boundaries of the oxide (C-B, D-E) the lateral electric field has to vanish.

It is interesting to note that many authors suggest a one dimensional voltage drop in the oxide e.g. /166/. In that manner one can obtain a mixed boundary condition (c.f. Eq. 3.1-2) for the potential at the interface. However, we feel that this is too crude an assumption for miniaturized MOS transistors.

3.1.3 Current Continuity Equations

Only the discretisation of the continuity equation for electrons will be treated in the following because the continuity equations for holes can be handled in an analogous manner. The major difficulty of the discretisation of the continuity equations is to find a proper, numerically stable formulation of the divergency of the current by just using information of physical quantities at meshpoints. A naive discretisation of the current relation (c.f. (3.1-12) for electrons) as only published in very early papers has been proved by various authors to be unstable.

Discretisation of Semiconductor Equations

$$\mathrm{div}\ \bar{J}_n = -q \cdot \mathrm{div}\ (\ \mu_n \cdot n \cdot \mathrm{grad}\ \psi - D_n \cdot \mathrm{grad}\ n\ ) \qquad (3.1\text{-}12)$$

However, Scharfetter and Gummel suggested already in 1969 a stable discretisation which has been physically motivated /132/. Their work can certainly be interpreted mathematically which helps in understanding various numerical phenomena associated with that discretisation. By assuming the validity of Boltzmann's statistics one gets with the substitution:

$$s = e^{-\phi_n/Ut} \qquad (3.1\text{-}13)$$

the following expression for the divergence of the electron current:

$$\mathrm{div}\ \bar{J}_n = q \cdot \mathrm{div}(D_n \cdot n_i \cdot e^{\psi/Ut} \cdot \mathrm{grad}\ s) \qquad (3.1\text{-}14)$$

This substitution is, as a matter of fact, essential because we have now a self-adjoint elliptic operator in s for the divergence of the current. For that type of operators the mathematical analysis is relatively easy and well investigated.

Recalling now the elliptic operator of the quasiharmonic equation (3.1-1), an analogy becomes evident. With:

$$P(x,y) = D_n \cdot n_i \cdot e^{\psi/Ut} \qquad (3.1\text{-}15)$$

we can use the results of section 3.1.1 for the discretisation. The fundamental problem, however, is to find a proper interpolation of (3.1-15) to obtain the mid-vertex values P_M etc. A naive linear interpolation of the exponential of the electric potential is definitely not appropriate. The very best one can do from the mathematical point of view is to use an exponential interpolation. (3.1-16) is an example for this type of interpolation between points (i,j) and (i+1,j). We should like to refrain from a proof of this relation as it is fairly lengthy. The interested reader should consult e.g. /124/.

$$
\begin{aligned}
e^{\psi_M} &= e^{\psi_{i,j}} \cdot (\psi_{i,j} - \psi_{i+1,j})/(e^{\psi_{i,j} - \psi_{i+1,j}} - 1) \\
&= e^{\psi_{i,j}} \cdot \mathrm{ber}(\psi_{i,j} - \psi_{i+1,j}) \qquad (3.1\text{-}16)
\end{aligned}
$$

with: $\mathrm{ber}(x) = x/(e^x - 1)$ (Bernoulli function) $\qquad (3.1\text{-}17)$

The actual programming of the Bernoulli functions (3.1-17) has to be undertaken with care to avoid underflow and overflow traps /73/. Furthermore, it should be noted that only differences of potential values occur in the Bernoulli functions.

Discretisation of Semiconductor Equations

Numerical stability is, therefore, greatly increased. The leading explicit exponential of ψ in (3.1-16) vanishes if it is combined with the s values (3.1-13) to electron densities which are numerically in an acceptable range compared to the exponentials of ψ and φ_n.

The mystery of the stability of the discretisation which has just been outlined lies in the fact that it represents a so-called "windward" difference approximation. For a large potential drop between neighbouring meshpoints the windward scheme degenerates in a forward- or backward difference scheme depending on the sign of the potential difference (i.e. electric field). Therefore the propagation of local errors is very small.

The boundary conditions are very simple for the continuity equations. At the contacts (A-B, E-F, G-H in Fig. 3.1-2) the carrier densities are set constant to their equilibrium value. At the remaining boundaries (B-E, F-G, A-H) no current component perpendicular to the boundary must exist.

A very interesting and successfully applied alternative to the outlined discretisation has been proposed by Mock /108/ for the MOS transistor. Through the introduction of so-called "stream-functions" one also obtains a self-adjoint operator for the divergence of the current with similar problems in the interpolation of exponentials of the electric potential. The treatment of inhomogeneities i.e. recombination/generation, however, is more complicated with stream-functions. Therefore we favour the other discretisation.

3.1.4 Grid Generation

To keep computer time as well as memory requirements reasonably small, it is necessary to limit the number of mesh points. A suitable tradeoff between accuracy and computing costs can be found once the discretisation errors are estimated. In critical regions with large discretisation errors grid spacing has to be kept small whereas it may be large in regions in which only small errors occur. Such considerations make it evident that an equidistant mesh is not suitable because in that case grid spacing has to be adapted to the critical regions and the number of mesh points would be very large.

As the discretisation errors depend on the distribution of the quantities ψ,n,p a suitable mesh cannot be estimated a priori, that is without knowledge of the solution. Therefore, grid generation is performed adaptively, i.e. a priliminary solution is calculated on the basis of an initial mesh, then the mesh is adapted to this solution and again the basic equations

Discretisation of Semiconductor Equations

are solved. Regeneration of the mesh can be done repeatedly if necessary.

Let f(x) be a four times continuously differentiable function; then one can savely write;

$$f_{i+1} \doteq f_i + f_i' h_i + f_i'' \frac{h_i^2}{2} + f_i''' \frac{h_i^3}{6} + f^{IV}(\xi)\frac{h_i^4}{24} \qquad (3.1-18)$$

$$f_{i-1} \doteq f_i - f_i' h_{i-1} + f_i'' \frac{h_{i-1}^2}{2} - f_i''' \frac{h_{i-1}^3}{6} + f^{IV}(\xi)\frac{h_{i-1}^4}{24} \qquad (3.1-19)$$

and we get for the second order differential quotient:

$$f''(x_i) \doteq 2 \cdot \frac{(f_{i+1}-f_i)/h_i + (f_{i-1}-f_i)/h_{i-1}}{h_i + h_{i-1}} +$$
$$+ f_i''' \frac{h_i - h_{i-1}}{3} + f^{IV}(\xi) \cdot \frac{h_i^2 - h_i h_{i-1} + h_{i-1}^2}{12} \qquad (3.1-20)$$

The first term on the right hand side of eq. (3.1-20) is the finite difference approximation the other two terms represent the discretisation error. For an aequidistant mesh ($h_i = h_{i-1}$) the second term on the right hand side vanishes and only differentials of at least forth order cause discretisation errors. Principially eq. (3.1-20) can be used to fix the mesh spacing h_i for given largest acceptable error, knowledge of the fourth differential provided, and also to bound the maximum mesh progression γ_i when knowing the third differential

$$\gamma_i = \max \left(\frac{h_i}{h_{i-1}}, \frac{h_{i-1}}{h_i}\right). \qquad (3.1-21)$$

The errors which are induced by the finite difference approximation of the inhomogeneity terms are going to be considered in the following. The inhomogeneities around each mesh point up to the next midpoint are approximated by

$$\int_{x_i - \frac{h_{i-1}}{2}}^{x_i + \frac{h_i}{2}} \int_{y_j - \frac{k_{j-1}}{2}}^{y_j + \frac{k_j}{2}} F(x,y) \cdot dx \cdot dy \doteq 0.25 \cdot F_{i,j}(h_i + h_{i-1})(k_j + k_{j-1}) \qquad (3.1-22)$$

with F(x,y) being the inhomogeneity term. For a one dimensional error estimation a series expansion of F(x) yields:

Discretisation of Semiconductor Equations

$$\int_{x_i-\frac{h_{i-1}}{2}}^{x_i+\frac{h_i}{2}} F(x)\cdot dx \doteq \frac{1}{2}F_i(h_i+h_{i-1})+F_i'\frac{h_i^2-h_{i-1}^2}{8}+F''(\xi)\frac{h_i^3+h_{i-1}^3}{24} \qquad (3.1-23)$$

The first term in (3.1-23) is the finite difference approximation as given in (3.1-22) and the two other terms describe the local error. The global error is extremely difficult to estimate. However, it is often of by one reduced order.

If we consider Poisson's equation once again, the second differential of the space charge limits the mesh spacing and the first differential limits mesh progression.

$$\gamma_i - 1 = 4\frac{\varepsilon_{max}}{F_i' h_{i-1}} \qquad (3.1-24)$$

$$h_i^2 = 12\frac{\varepsilon_{max}}{F''(\xi)} \qquad (3.1-25)$$

As already discussed earlier ionisation rates are very sensitive to the electric field. Therefore, the generation rate exhibits an abrupt peak in the pinch-off region which can only be kept under control with a very fine discretisation. The integral of the generation rate over the total area gives the substrate current, and the discretisation error is, therefore, proportional to the error of the substrate current. If we consider only the first derivative of the electric field

$$E(x) \doteq E(x_i) + (x-x_i)\cdot E'(x_i)$$

we get for α'' on the basis of Chynoweth's law:

$$\alpha(x) \doteq A\cdot\exp\left[\frac{-B}{E_i}\cdot\left(1 - \frac{(x-x_i)E_i'}{E_i}\right)\right]$$

$$\alpha'(x_i) \doteq \frac{B}{E_i^2}\cdot E_i'\cdot\alpha(x_i) \qquad (3.1-26)$$

$$\alpha''(x_i) \doteq \left[\frac{B}{E_i^2}\cdot E_i'\right]^2\cdot\alpha(x_i) \qquad (3.1-27)$$

By substituting these expressions in (3.1-24), (3.1-25) we get some rules for the mesh generation. If one likes to limit the maximum relative error in the substrate current, it is useful to divide (3.1-26), (3.1-27) by the maximum ionization rate which occurs in the device.

Discretisation of Semiconductor Equations

3.2 Linearization of the Coupled System

In this section we should like to discuss some properties of one-step stationary iterative methods of the form (3.2-2) for the solution of systems of nonlinear equations (3.2-1).

$$F(x) = 0 \qquad\qquad (3.2-1)$$

$$x^{k+1} = G \cdot x^k \qquad\qquad (3.2-2)$$

The problem (3.2-1) be properly defined /124/. Suppose the operator G has a fixed point x^*. Then (3.2-2) will converge to x^* if G is F-differentiable /124/ at x^* and if the spectral radius of it's Jacobian $G'(x^*)$ satisfies Eq. (3.2-3).

$$\varrho(G'(x^*)) < 1 \qquad\qquad (3.2-3)$$

This very important theorem (Ostrowski theorem) is the basis of all investigations on the convergence of one-step iterations for the solution of nonlinear equations. In the following we should like to write the discretized semiconductor equations in a more abstract form to simplify the formulae. Let (3.2-4) be Poisson's equation and (3.2-5), (3.2-6) the continuity equations for electrons and holes in the unknowns ψ, φ_n and φ_p which represent vectors of the meshpoint values of the electrostatic potential, the quasifermilevel of electrons and holes, respectively.

$$F1 = F1(\psi, \varphi_n, \varphi_p) = 0 \qquad\qquad (3.2-4)$$

$$F2 = F2(\psi, \varphi_n, \varphi_p) = 0 \qquad\qquad (3.2-5)$$

$$F3 = F3(\psi, \varphi_n, \varphi_p) = 0 \qquad\qquad (3.2-6)$$

Then F1, F2 and F3 represent a nonlinear system of equations with the rank $3 \cdot (NX \cdot NY)$ which is three times the total number of meshpoints, usually a large number. We have now to find some operators G which are relatively easy to calculate and for which condition (3.2-3) holds. The best know method is certainly the classical Newton-Raphson method.

3.2.1 Newton's Method and Modified Newton Methods

For Newton's method the iteration is defined as follows:

$$
\begin{pmatrix} \psi \\ \varphi_n \\ \varphi_p \end{pmatrix}^{k+1} = \begin{pmatrix} \psi \\ \varphi_n \\ \varphi_p \end{pmatrix}^{k} - \begin{pmatrix} \partial F1/\partial\psi & \partial F1/\partial\varphi_n & \partial F1/\partial\varphi_p \\ \partial F2/\partial\psi & \partial F2/\partial\varphi_n & \partial F2/\partial\varphi_p \\ \partial F3/\partial\psi & \partial F3/\partial\varphi_n & \partial F3/\partial\varphi_p \end{pmatrix}_k^{-1} \cdot \begin{pmatrix} F1 \\ F2 \\ F3 \end{pmatrix} \qquad (3.2-7)
$$

The operator G is, therefore, defined as:

$$G = (I - F'^{-1} \cdot F)$$ (3.2-8)

It can be proved that the Jacobian G' of this operator has only zero eigenvalues at x^*, $F(x^*)=0$ and fulfills trivially condition (3.2-3). As all eigenvalues are zero, even quadratic convergence is anticipated as the solution is approached /124/, /126/. Although Newton's method is very attractive from the mathematical point of view, there are practical difficulties.

The main implementation problem is the evaluation of the derivative terms in equation (3.2-7), since the total equation (including modeled physical parameters) must be differentiated accurately with respect to the variables ψ, φ_n and φ_p. This can be done analytically, in prinicipal, but one loses much flexibility in changing the models of the physical parameters. Therefore, a numerical algorithm is necessary to automatically calculate the required derivatives. The best algorithm known at the moment has been published by Curtis and Reid /34/. Some very interesting comments on numerical differentiation have also been given in /81/.

Another fairly difficult problem when considering Newton's method is overshoot. The iteration process does not neccesarily converge monotonously to the solution. Especially, if one starts the iteration with a bad initial guess - which is the usual situation - monotonic convergence or convergence at all cannot be guaranteed. Therefore, one has to introduce a mechanism which dampens the increments resulting from the iteration process so that convergence is monotonic. The naive algorithms simply limit the increments to some maximum value e.g. /108/ or they use some function to continuously limit large increments e.g. /19/. Deuflhardt suggested a more elaborate method /44/, /45/. Roughly explained, he calculates a parameter ν in the range $]0,1]$ so that condition (3.2-9) holds.

$$F'^{-1}(x^k) \cdot F(x^k + \nu \cdot (x^{k+1} - x^k)) < F'^{-1}(x^k) \cdot F(x^k)$$ (3.2-9)

After having found ν, the solution x^{k+1} is calculated with Eq. (3.2-10).

$$x^{k+1} = x^k + \nu \cdot (x^{k+1} - x^k)$$ (3.2-10)

This procedure guarantees monotonic convergence; for $\nu=1$ it is the classical Newton method.

Another excellent method was proposed by Meyer /104/. He suggested to modify the iteration operator G by introducing a positive parameter λ as given in Eq. (3.2-11).

Linearization of the Coupled System

$$G = (I - (\lambda \cdot I + F')^{-1} \cdot F) \qquad\qquad (3.2\text{-}11)$$

λ has to be chosen as small as possible so that the euklidean norm of $F(x^{K+1})$ is smaller than the euklidean norm of $F(x^K)$. Some practical guidelines on how to find this parameter λ with reasonable effort have been recently presented by Bank and Rose /9/.

It is very time and also memory consuming to solve the large linear system which, nevertheless, has to be done for every Newton step. Therefore, many authors use another iterative algorithm, suggested by Gummel /69/, for the linearization of the semiconductor equations.

3.2.2 Block-Nonlinear Iteration

Gummel's idea, in essence, was to solve the semiconductor equations by independently linearizing each equation consecutively with respect to its dominant variable. The first step is to solve Poisson's equation with Newton's method assuming that the quasifermilevels are known functions of position; i.e. the best guess of the quasifermilevels is assumed to be correct. In the second step one of the continuity equations is solved assuming that the electric potential and the quasifermilevel of the other continuity equation are correct. In the third step the second continuity equation is solved under similiar assumptions. These three steps are performed repeatedly until a consistent solution is found. The effort for one cycle of this block-nonlinear iteration is obviously less than for one Newton step because the rank of one decoupled equation system is only a third of the rank of the total system.

A complete theoretical prof of the convergence of the block-nonlinear iteration algorithm has as yet not been published. However, strong theoretical indications have been given by Mock /109/. One should also not underestimate the "practical" indications; many authors have used Gummel's iterative schema with excellent success.

It should also be noted that a few authors have published modifications of the original Gummel method e.g. /14/, /133/, /156/. These modifications are basically motivated on intuition. They are, therefore, as arbitrary as the original approach itself. However, for special purpose applications some improvement in efficiency could be gained.

Linearization of the Coupled System

3.3 Solution of Large Sparse Linear Systems

For any of the linearization procedures which have been outlined in section 3.2 a large sparse linear equation system (3.3-1) has to be solved repeatedly.

$$A \cdot x = b \qquad (3.3-1)$$

We assume that A has been derived by linearizing five-point-star discretized PDEs. Hence matrix A has at most five nonzero entries per row; A is very sparse. For a full Newton scheme these entries are 3x3 matrices; for Gummel's scheme they are scalars. For the solution of these special types of linear equation systems two classes of methods, can, in principle, be used: direct methods which are based on elimination and iterative methods. An excellent survey on that subject has been published recently by Duff /48/.

3.3.1 Direct Methods

Classical Gaussian elimination is not feasible for our systems of equations because the rank of A in (3.3-1) is very large and A has many coefficients which are zero. Therefore, some modifications of the classical Gaussian elimination algorithm have to be introduced to account for the zero entries. There exist quite a few activities on that subject (c.f. /49/) and powerful algorithms which treat the nonzero coefficients only are available. Another serious drawback of direct methods lies in the fact that the upper triangular matrix which is created by the elimination process has to be stored for back substitution. This matrix has usually more nonzero entries than the matrix A. Therefore, memory requirement of direct methods is substantial.

In spite of all drawbacks of direct methods, their major advantage is high accuracy of the solution. However, we feel that for the semiconductor problems iterative algorithms are to be slightly favoured.

3.3.2 Relaxation Methods

The fundamental idea of relaxation methods is the splitting of the coefficient matrix A (3.3-1) into three matrices D, E, F (3.3-2).

$$A = D - E - F \qquad (3.3-2)$$

D denotes the diagonal entries of A; -E denotes a lower triangular matrix which is formed from all sub-diagonal entries of A; and -F denotes an upper triangular matrix which is formed from all super-diagonal entries of A.

With an arbitrary non-singular matrix B which has the same rank as A the linear system (3.3-1) can be rewritten to (3.3-3).

$$B \cdot x + (A-B) \cdot x = b \qquad (3.3-3)$$

One obtains an iterative schema by setting:

$$B \cdot x^{k+1} = b - (A-B) \cdot x^k \qquad (3.3-4)$$

(3.3-4) can be solved for x^{k+1}:

$$x^{k+1} = (I-B^{-1} \cdot A) \cdot x^k + B^{-1} \cdot b \qquad (3.3-5)$$

The iterative scheme (3.3-5) will converge if condition (3.3-6) holds.

$$\varrho(I-B^{-1} \cdot A) < 1 \qquad (3.3-6)$$

(3.3-6) is a necessary and sufficient condition. The various relaxation methods can be won by differently setting matrix B with the matrices obtained by the splitting of A (3.3-2).

The simpliest scheme, the point-Jacobi method, uses D for B. Matrix D is a diagonal matrix and is, therefore, easily invertible.

The Gauss-Seidel method uses D-E for B. The matrix D-E is a lower triangular matrix. Therefore one has only to perform a forward substitution process for its inversion.

The succesive overrelaxation method (SOR) makes use of a parameter ω within the range $]0,2[$. The iteration matrix B is defined:

$$B = D/\omega - E \qquad (3.3-7)$$

As B is again a lower triangular matrix, its inversion is instantly reduced to a substitution.

The major advantage of these iterative methods lies in their simplicity. They are very easy to program and demand only low memory requirement. As already noted, they converge if condition (3.3-4) holds. However, it is difficult to prove that condition. A sufficient condition for convergence is that A is positive definite (3.3.-8) which is the regular case for five-point-star discretized PDEs.

$$x^T \cdot A \cdot x > 0 \text{ for all } x \neq 0 \qquad (3.3-8)$$

Solution of Large Sparse Linear Systems

These point-iterative schemes can by accelerated quite remarkably with the conjugate gradient method or the Chebyshev method. An excellent survey on these topics can be found in /68/.

3.3.3 Strongly Implicit Iterative Methods

The convergence rate of relaxation methods is relatively poor. Therefore, various activities can be observed for the development of more powerful algorithms with the advantages of iterative schemes.

One of the best known algorithms which has been established in semiconductor device analysis is perhaps Stone's strongly implicit procedure /151/. Stone's ideas was to modify the original coefficient matrix A by the addition of a small matrix N so that a factorization of (A+N) involves much less computational effort than the standard decomposition of A and the norm of N is much smaller than the norm of A. Assuming this has been done, the development of an iterative procedure is then fairly straightforward because the equation can be written as:

$$(A+N) \cdot x = (A+N) \cdot x + (b - A \cdot x) \qquad (3.3-9)$$

which suggests the iterative procedure:

$$(A+N) \cdot x^{k+1} = (A+N) \cdot x^{k} + (b - A \cdot x^{k}) \qquad (3.3-10)$$

When the right hand side is known and (A+N) can be factorized easily, (3.3-10) gives an efficient method for directly solving for x^{k+1}. Furthermore, one would intuitively expect a rapid rate of convergence if N is sufficiently small compared to A. We will refrain from explaining in detail the suggestion of Stone of how to choose the perturbation matrix N because this has been done thoroughly in many publications e.g. /59/, /147/, /151/.

There exist a few algorithms which are similiar in terms of underlying ideas compared to Stone's method. The most attractive are perhaps the method of Dupont et al. /50/, the "alternating direction implicit" methods e.g. /16/, /59/, /178/ and the Fourier methods /150/, /177/.

One disadvantge of all strongly implicit methods and also the direct methods is that they cannot be implemented efficiently on a computer with a pipe-line architecture (vector processor). Some comments on that subject have been given in /48/.

Solution of Large Sparse Linear Systems

4. TYPICAL APPLICATIONS OF MINIMOS

4.1 A Didactic Example

It is rather difficult to provide an interesting example for the experienced reader, which is also impressive and easy to understand for readers with general interest in modeling but without specific knowledge of device physics. We have chosen the effects of ion implantation on short channel MOS transistors for the purpose of demonstrating the use of two dimensional simulation. Three devices are calculated whose properties become apparent from the original simulation input decks presented in Fig. 4.1-1. The following discussion of Fig. 4.1-1 shall also demonstrate the ease of using MINIMOS /138/, /139/, /140/, /142/, our simulation program.

The first line is a title line, which is used only to identify the output of the program. The input syntax is totally based on a master key, key and value structure. The next input line which is the "DEVICE" statement, characterizes the device. Specified is an n-channel device (CHANNEL=N) with an, n-doped polysilicon gate (GATE=NPOLY), an oxide thickness of 35 nanometers (TOX=350.E-8), a channel width of 10 micrometers (W=10.E-4) and a channel length of one micrometer (L=1.E-4). The "BIAS" statement specifies the operating point. A drain voltage of 3 volts (UD=3.) and a gate voltage of zero volts (UG=0.) has been chosen. The substrate voltage is assumed to be zero by MINIMOS, if not specified otherwise.

The "PROFILE" statement is used to specify the substrate doping and the source/drain diffusion. In the examples presented here we used the simplest way of defining a doping profile, that is the direct calculation by MINIMOS. Another possibility would be to make use of a technology simulation program like SUPREM, the Stanford University PRocess Engineering Models program /6/, for the more accurate calculation of vertical profile shapes which are fitted in the lateral direction. For our simulation a substrate doping of $10^{15} cm^{-3}$ (NB=1.E15) and a source/drain implantation with phosphorus (ELEM=PH), an implantation dose of $10^{15} cm^{-2}$ (DOSE=1.E15) and an implantation energy of 40keV (AKEV=40) is specified. The implantation is performed through an isolation oxide of 35 nanometers (TOX=350.E-8) and followed by an annealing step at 1000 centigrades (TEMP=1000) for 1200 seconds (TIME=1200).

A Didactic Example

```
ONE-MICRON ANALYSIS (DEVICE 1)
DEVICE    CHANNEL=N  GATE=NPOLY  TOX=350.E-8  W=10.E-4  L=1.E-4
BIAS      UD=3.  UG=0.
PROFILE   NB=1.E15  ELEM=PH  DOSE=1.E15  AKEV=40  TOX=350.E-8
+         TEMP=1000  TIME=1200
END


ONE-MICRON ANALYSIS (DEVICE 2)
DEVICE    CHANNEL=N  GATE=NPOLY  TOX=350.E-8  W=10.E-4  L=1.E-4
BIAS      UD=3.  UG=0.
PROFILE   NB=1.E15  ELEM=PH  DOSE=1.E15  AKEV=40  TOX=350.E-8
+         TEMP=1000  TIME=1200
IMPLANT   ELEM=B  DOSE=3.5E11  AKEV=25  TEMP=925  TIME=1800
END


ONE-MICRON ANALYSIS (DEVICE 3)
DEVICE    CHANNEL=N  GATE=NPOLY  TOX=350.E-8  W=10.E-4  L=1.E-4
BIAS      UD=3.  UG=0.
PROFILE   NB=1.E15  ELEM=PH  DOSE=1.E15  AKEV=40  TOX=350.E-8
+         TEMP=1000  TIME=1200
IMPLANT   ELEM=B  DOSE=3.5E11  AKEV=25  TEMP=925  TIME=1200
IMPLANT   ELEM=B  DOSE=1.5E11  AKEV=100
END
```

Fig. 4.1-1: Some typical input decks for MINIMOS

The second input deck further includes an "IMPLANT" statement which defines a channel implantation with boron (ELEM=B), a dose of $3.5 \cdot 10^{11}$ cm^{-2} (DOSE=3.5E11), an energy of 25keV (AKEV=25), annealed at 925 centigrades (TEMP=925) for 1800 seconds (TIME=1800). The third input deck has an additional "IMPLANT" statement specifying a second, deeper channel implantation with boron (ELEM=B), a dose of $1.5 \cdot 10^{11}$ cm^{-2} (DOSE=1.5E11) and an energy of 100keV (AKEV=100). It is assumed that both channel implantation steps are annealed at the same time. It is fairly well known that the first of these three devices is, owing to the short channel effect, "normally-on" and that the shallow implantation of device 2 effects a threshold shift to obtain a "normally-off" device. Furthermore, the deep implantation of device 3 is necessary to avoid punch through. These effects will now be demonstrated by birds-eye-view- and contour-plots of physically relevant quantities in the interiour of the three model devices.

The calculated doping density distributions for our devices are shown in Figs. 4.1-2, 4.1-3, 4.1-4.

A Didactic Example
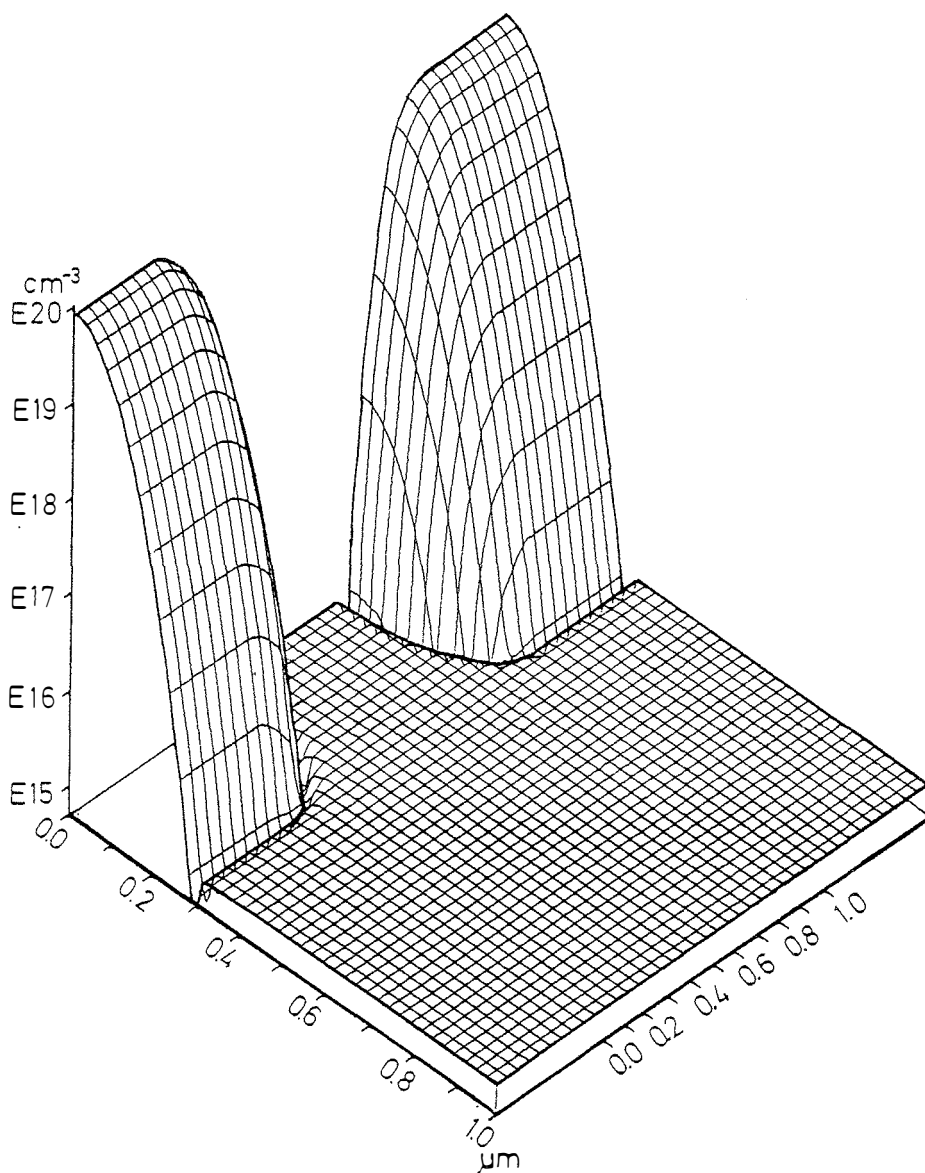
Fig. 4.1-2: Doping profile for device 1

From these figures one can read off the depth of the pn-junctions under source and drain being approximately 300 nanometers. The surface concentration of the source and drain regions is about 10^{20} cm^{-3}. The effective channel length is reduced by the lateral subdiffusion to about 0.6 micrometers. The shallow channel implantation for threshold tailoring can be seen in Figs. 4.1-3, 4.1-4. Additionaly, Fig. 4.1-4 shows the deep implantation for punch through suppression. The threshold voltage is only marginally affected by the deep implantation.

A Didactic Example
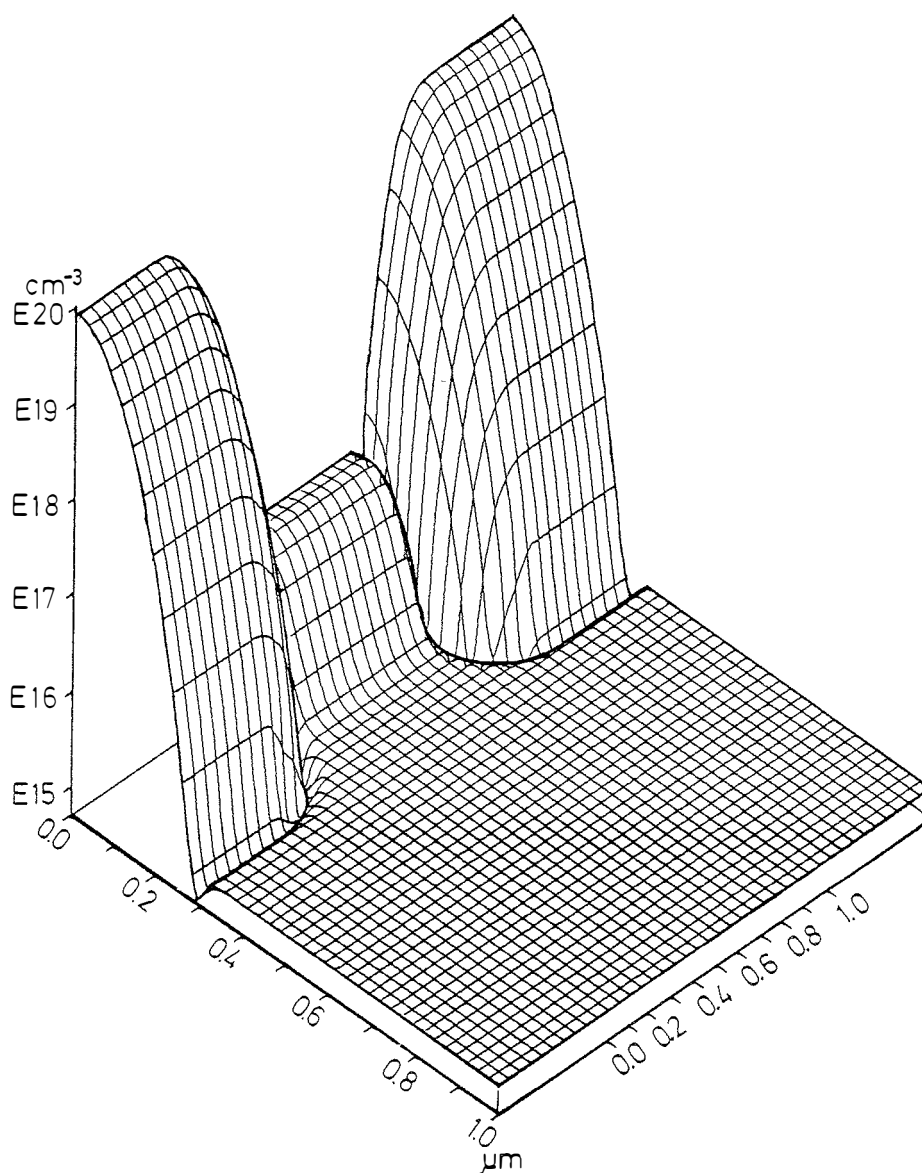
Fig. 4.1-3: Doping profile for device 2

Fig. 4.1-5 shows the distribution of the electric potential for the first device. The drain contact is on the right. In the depletion region of the reverse biased drain-bulk diode the potential decreases monotonously and it is more or less constant in the highly doped source and drain regions. The barrier at the source channel diode is relatively small /168/. Fig. 4.1-6 shows the potential distribution in the second device. The birds-eye-view plot looks very similar to the plot in Fig. 4.1-5.

A Didactic Example
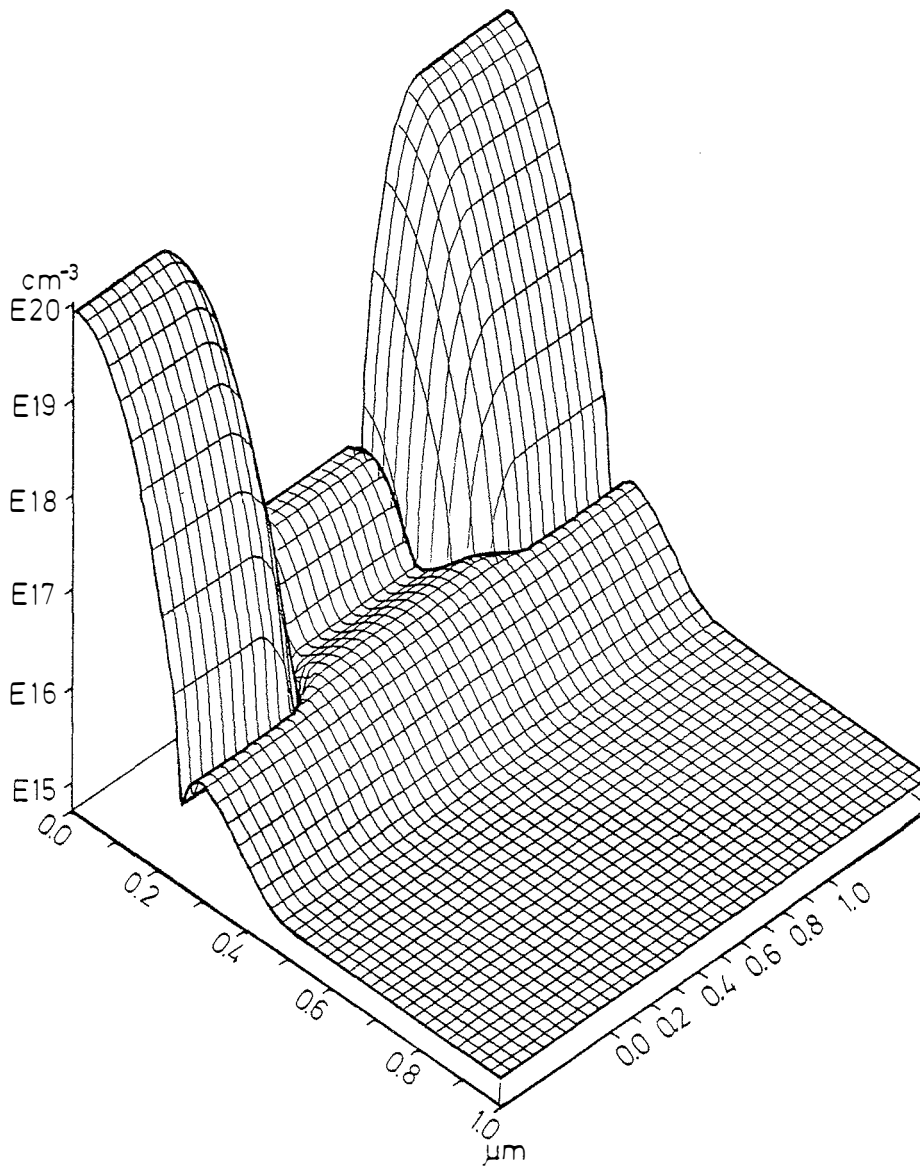
Fig. 4.1-4: Doping profile for device 3

The contour-plot, however, shows quite a pronounced
potential basin directly below the interface. Of even greater
importance than this basin itself is the saddlepoint below the
basin. At this saddlepoint the electric field vanishes and
current only can flow by carrier diffusion. This sort of saddle-
point is, following the proposition of many authors (e.g. /12/,
/87/), a typical indication of the punch-through effect. The
electric field which is induced by the gate is unable to separate
the depletion regions of source and drain.

A Didactic Example
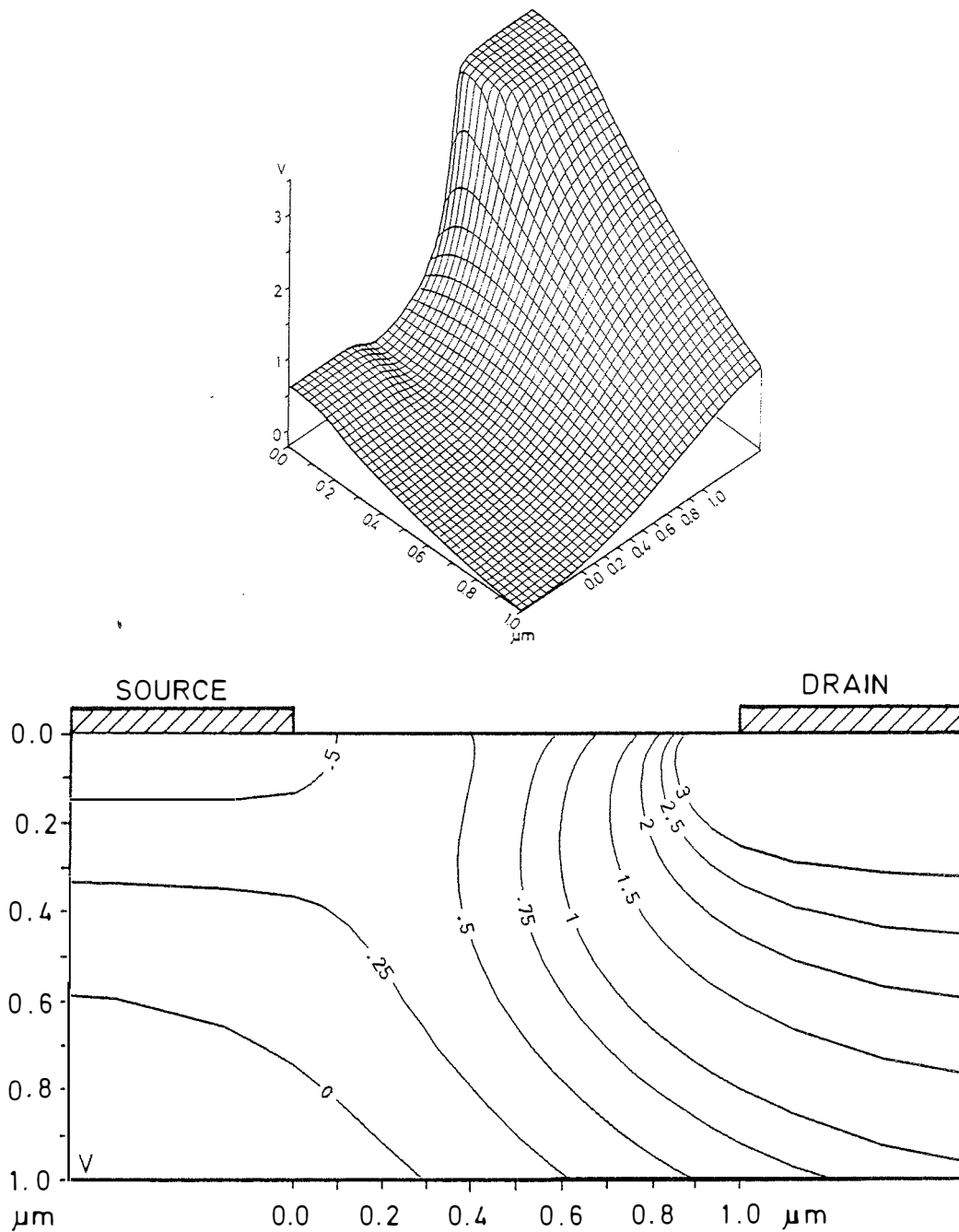
Fig. 4.1-5: Electric potential for device 1

These depletion regions are in contact below the region of control by the gate. As it will become apparent later , the saddlepoint is a reliable indication of the punch-through effect, but it need not exist.

A Didactic Example
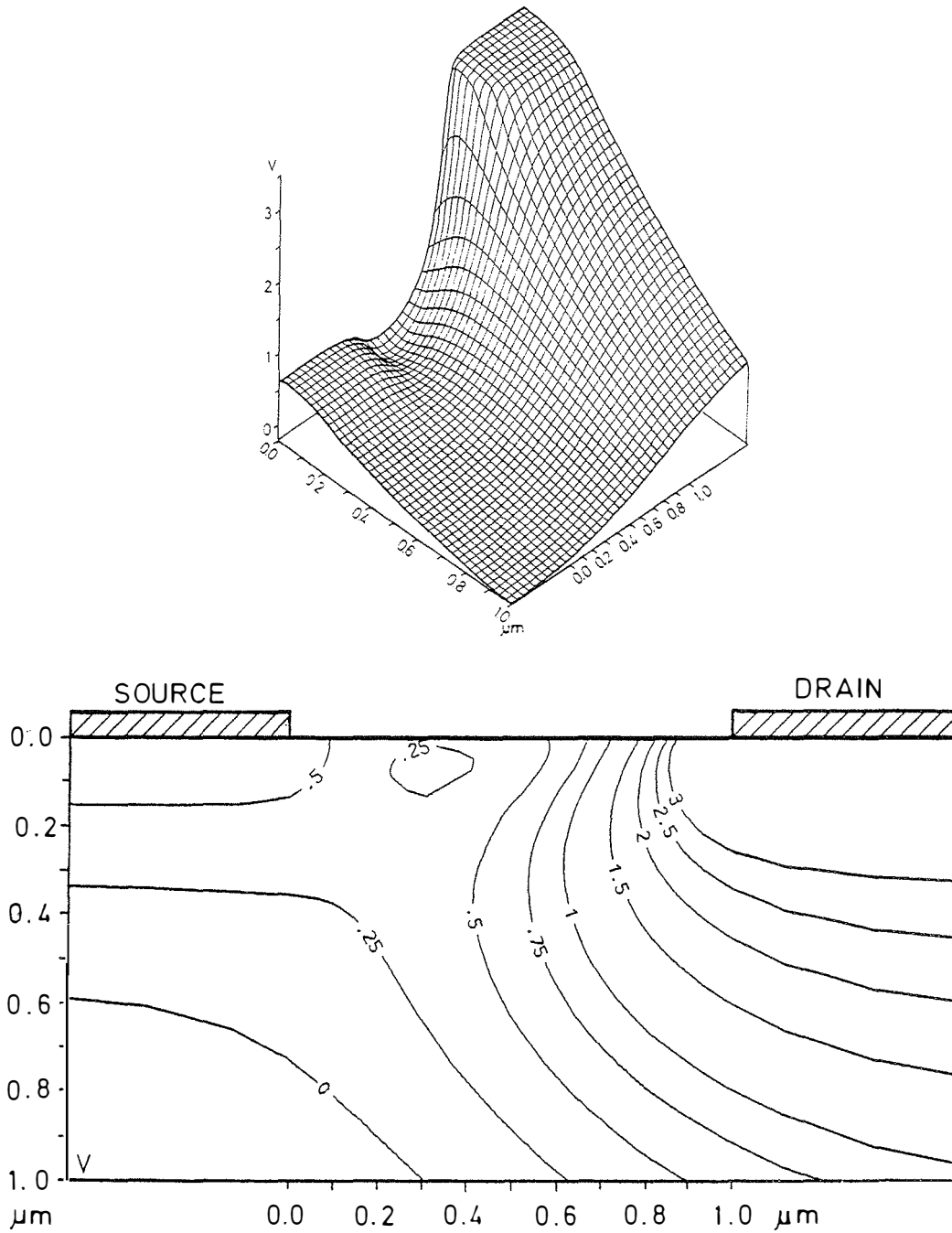
Fig. 4.1-6: Electric potential for device 2

Fig. 4.1-7 shows the potential distribution in the third device. The birds-eye-view plot differs just marginally from the plot in Fig. 4.1-6. But from the contour plot one can see a well pronounced barrier between source and channel which guarantees the "normally-off" behaviour.
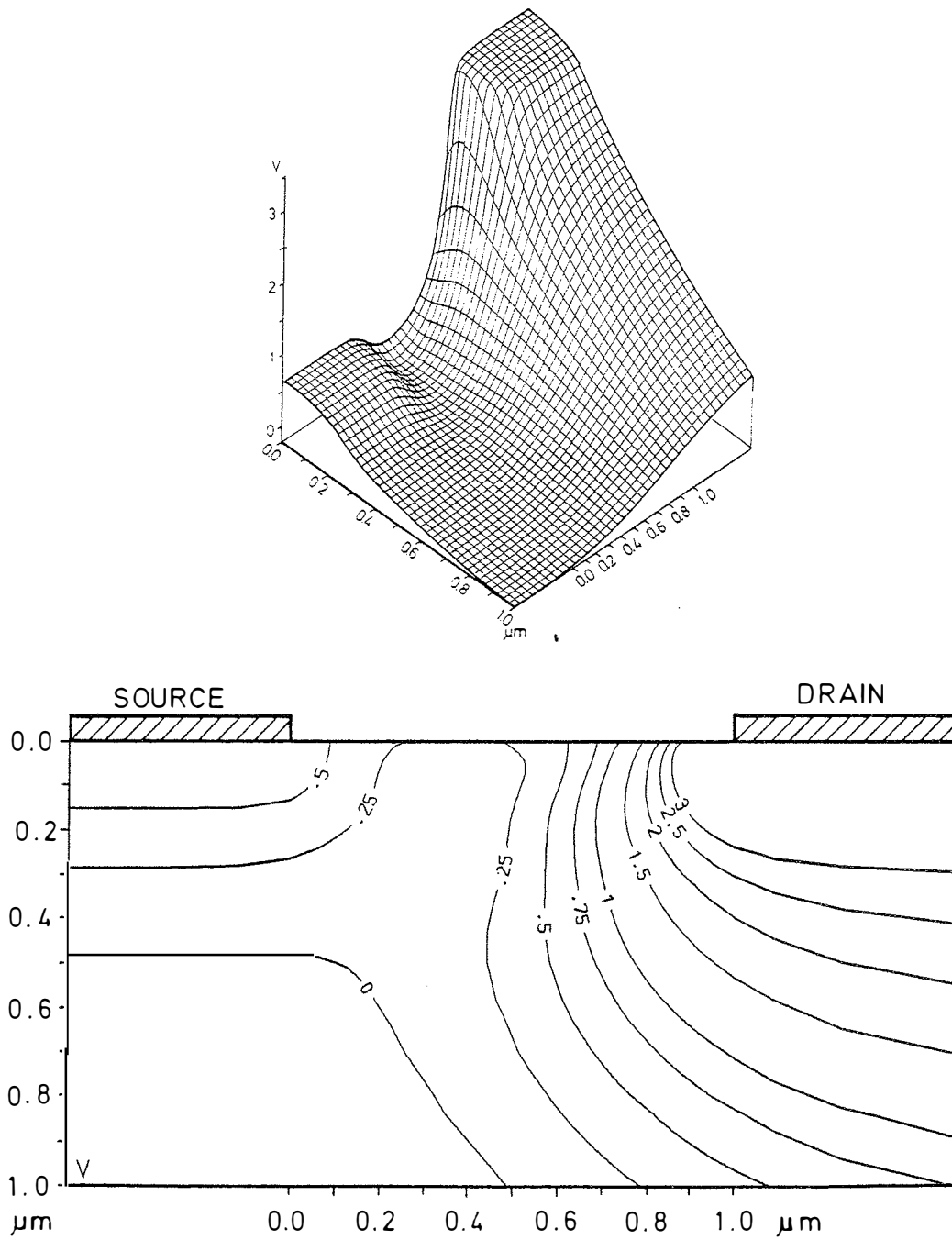
A Didactic Example

Fig. 4.1-7: Electric potential for device 3

A Didactic Example

Fig. 4.1-8 shows the lateral current density distribution in
the first device. For better visibility, the plot on the right
shows the mirror image to give better insight into the channel
region. In the channel near the source the current is forced to
flow at the surface by the transversal component of the field.
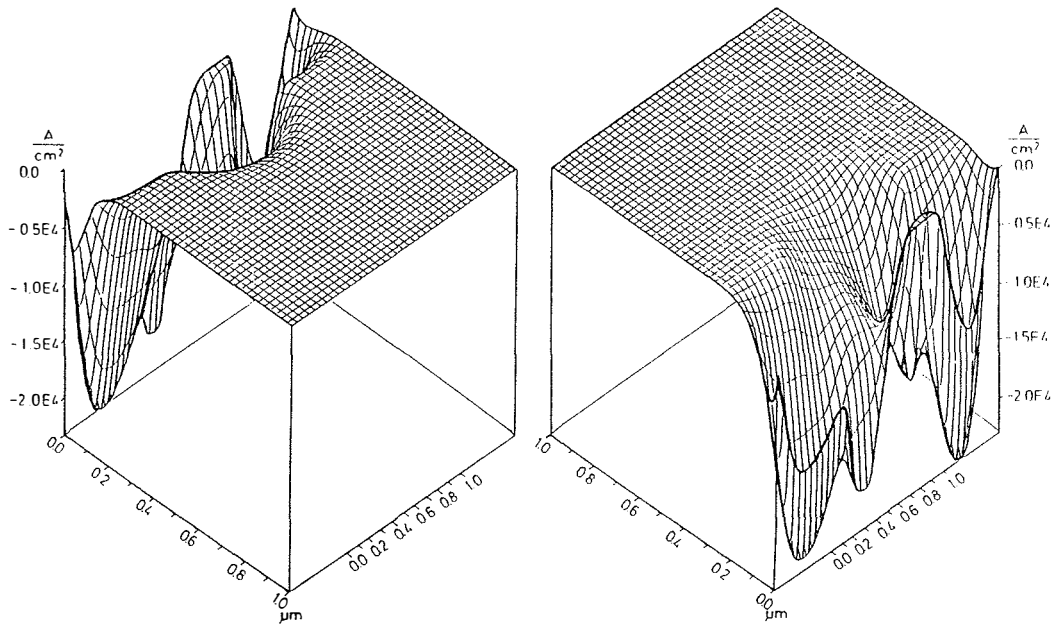
Fig. 4.1-8: Lateral current density for device 1

But already in the middle of the channel one can watch
current spreading caused by the drain influence, a typical short
channel effect.

It also should be noted that the current channel is fairly
wide. The reason for this phenomenon is to be found in a super-
position of an inversion channel and a punch through channel.
The maximum of the lateral current density surprisingly lies
below the contacts. This fact becomes clear when we consider
current continuity. Current can only pass through the contact in
transversal direction. Current flow in the semiconductor,
however, takes place globally in the lateral direction from
source to drain. As current flow is continuous, the lateral

A Didactic Example

current component has to be large below the contacts, because the flux in the channel, which is relatively wide, as mentioned, is large too.

The lateral current distribution for the second device is shown in Fig. 4.1-9. As one can see, this device is operating in the punch through mode. The current flow takes place in a wide channel in the bulk. Surface current does effectively not exist. Furthermore, the maximum of the current density has decreased more than an order of magnitude compared to the first device.

Fig. 4.1-10 shows the lateral current density distribution for the third device. The second channel implantation results in a total suppression of punch through in this operating point. The entire current flows at the semiconductor surface, but the peak value of the current density is about a factor of 200 smaller than in the second device.
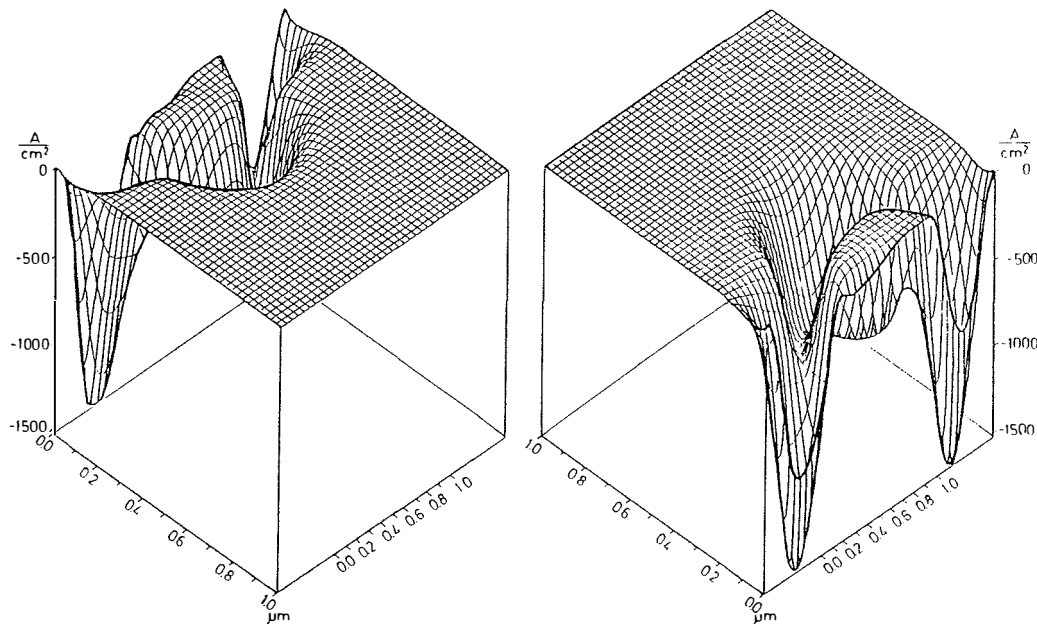
Fig. 4.1-9: Lateral current density for device 2

A Didactic Example
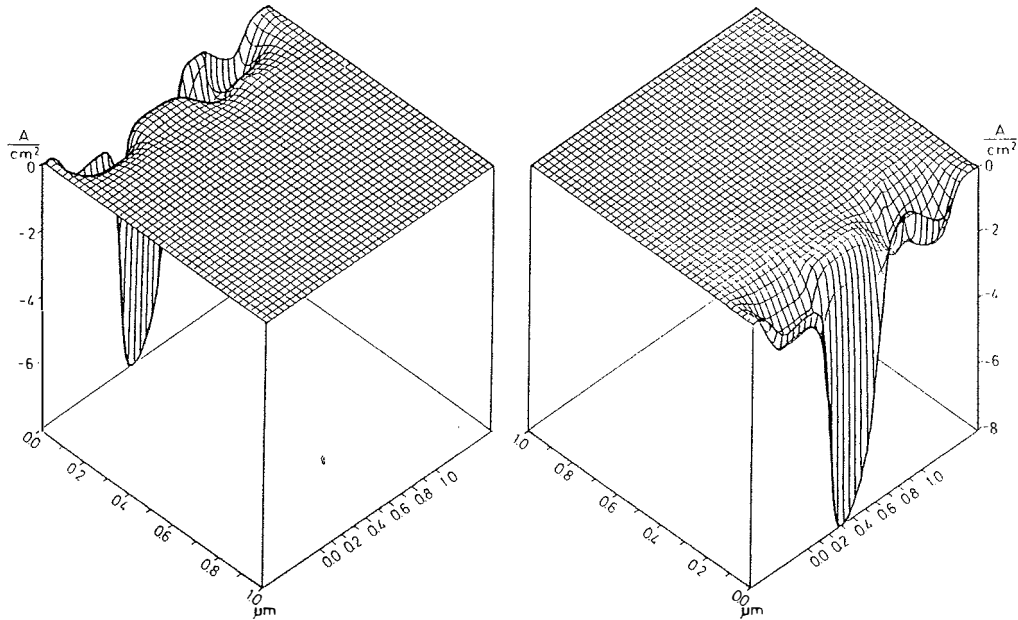
Fig. 4.1-10: Lateral current density for device 3

Current density distributions of this shape are typical for
regularly operating transistors in subthreshold and can be used
as criterion for valuation.

Fig. 4.1-11 shows the subthreshold characteristics for two
different drain voltages. The solid lines denote 100mV, the
dashed lines 3V drain bias. The slope is the same for all three
devices at a drain voltage of 100mV. It is dramatically
decreased at 3V drain bias for devices 1 and 2 by the punch
through current. The shift of the characteristics for different
drain voltages, which is caused by the short channel effect, is a
minimum for the third transistor thus verifying the success of
the channel implantation steps.
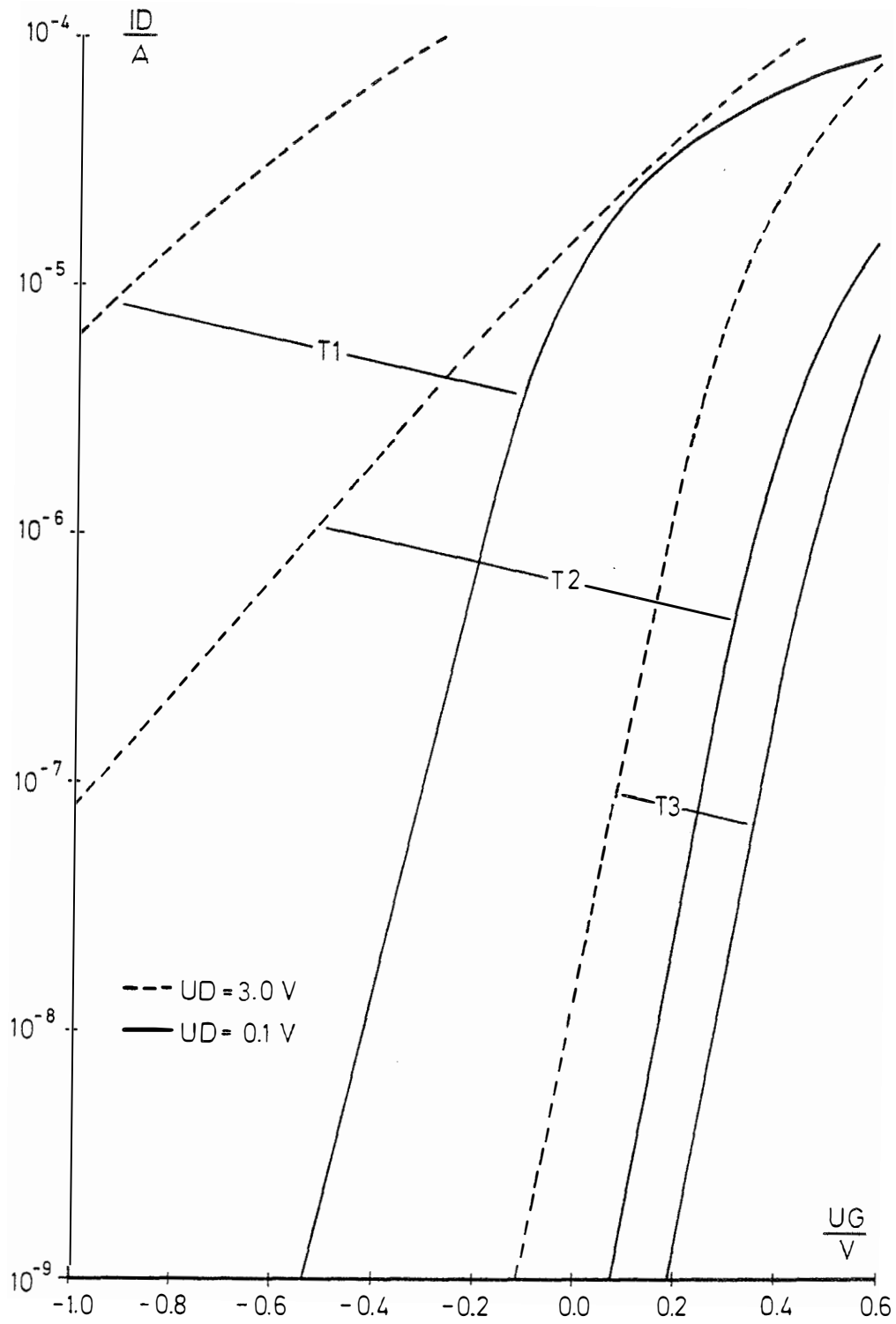
A Didactic Example

Fig. 4.1-11: Subthreshold characteristics

A Didactic Example

4.2 Process Sensitivity

VLSI is evidently connected to the miniaturization of the single transistor. Merely shrinking the physical device dimensions usually poses serious problems concerning device behaviour. Instead, all device parameters have to be scaled (e.g. /43/, /99/) together with the device geometry according to certain rules. In general, lower voltages, heavier doping, shallower junctions and thinner oxides help to maintain applicable device characteristics as channel length is reduced.

Down to about two microns channel length the device behaviour can be controlled excellently by the relevant technological steps (implantation, diffusion, oxidation, photolithography). However, as often observed in experimental investigations, this controlability is no longer ensured for devices with further reduced channel length. Reproducability tends to become worse with decreasing size, posing increasingly severe problems on tracking down the parameters of adjacent transistors, which should behave identically for certain kinds of circuits (e.g. latches).

To verify the increased process sensitivity of scaled devices, we performed an analysis of certain device parameters[1] with MINIMOS /137/. In this chapter the sensitivity of the threshold[1] voltage, which is usually the most important device property[2] for the designer, will be outlined for a well established short channel MOS process to determine the practical limit of miniaturization for a given technology. However, the analysis of threshold sensitivity is just an example for a strategy which is applicable to examine the sensitivity of any device property.

An n-channel silicon gate process with arsenic source/drain doping and a double channel implantation for threshold tailoring and punch-through supression has been chosen. Fig. 4.2-1 shows the doping distribution logarithmically drawn in a quasi-three-dimensional plot for a one micron transistor. The channel implantation is performed with boron as the dopant, a dose of 3.10^{11} cm^{-2} and an energy of 35keV for the shallow layer and a dose of 10^{11} cm^{-2} and an energy of 160keV for the deeper layer.

1) <u>parameter</u> = a variable which one can choose arbitrarily, e.g. L, W, T$_{ox}$.

2) <u>property</u> = a physical attribute which is influenced by choice of parameters, e.g. UT, breakdown voltage, transconductance.
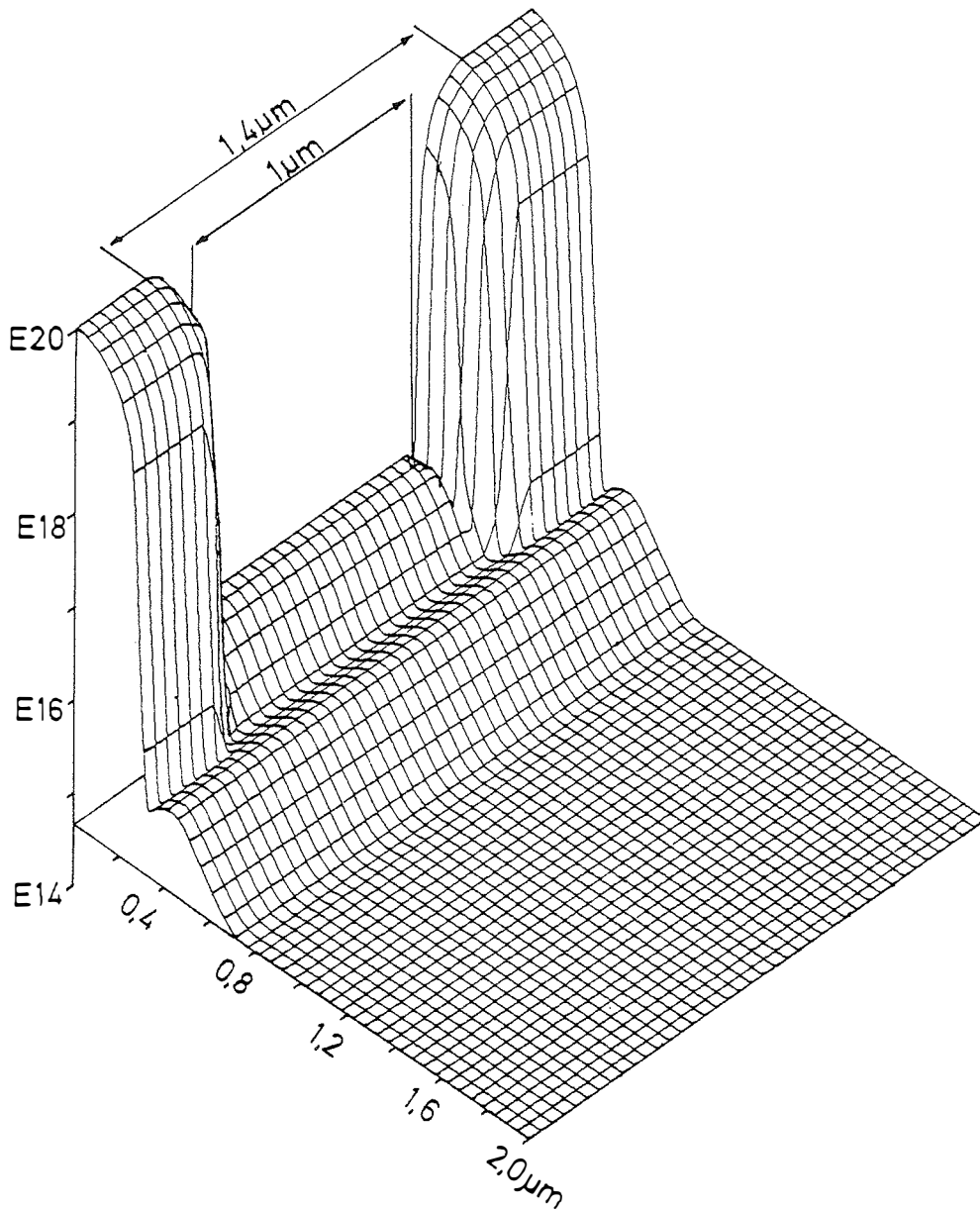
Process Sensitivity

Fig. 4.2-1: Doping profile of the analyzed devices

A junction depth of 320nm and a lateral diffusion of about 200nm is obtained by this process. The extremely steep gradient at the junctions is typical for arsenic. The oxide thickness - the oxide is not drawn in these figures - is about 50nm for these devices. The whole process was designed for two micron lateral dimension.

Process Sensitivity

For an analysis of the behaviour of the threshold voltage
one first has to formulate an adequate definition of the
threshold voltage. The most common definitions are based on the
extrapolation of an output characteristic. However, one drawback
of extrapolation methods lies in their inaccuracy and in the
experimental effort. Mainly owing to these mentioned reasons we
define threshold voltage in the following simple and definite
way: It is that applied gate voltage, at which the device sinks
0.1 microamps times the channel width per channel length. The
channel length is defined as the distance between the
metallurgical junctions. With this definition it is ensured that
no threshold voltage shift versus channel length for long devices
occurs and we can, therefore, directly obtain a quantitative
measure for the influence of the short channel effects. It is
probably necessary at this point to mention that drain bias and
bulk bias are not explicit parameters in our definition of the
threshold voltage. The dependence on those parameters has to be
obtained by certain characteristics, namely: threshold voltage
versus drain bias, threshold voltage versus bulk bias. Our
definition is naturally arbitrary - as arbitrary as any
definition - so one might have to argue about the quantitative
value of the used constant (0.1μA). For devices with a steep
subthreshold characteristic, and only such devices are of
practical relevance, we think that the constant we use is quite
suitable. For devices with a degraded subthreshold
characteristic any definition of a subthreshold characteristic
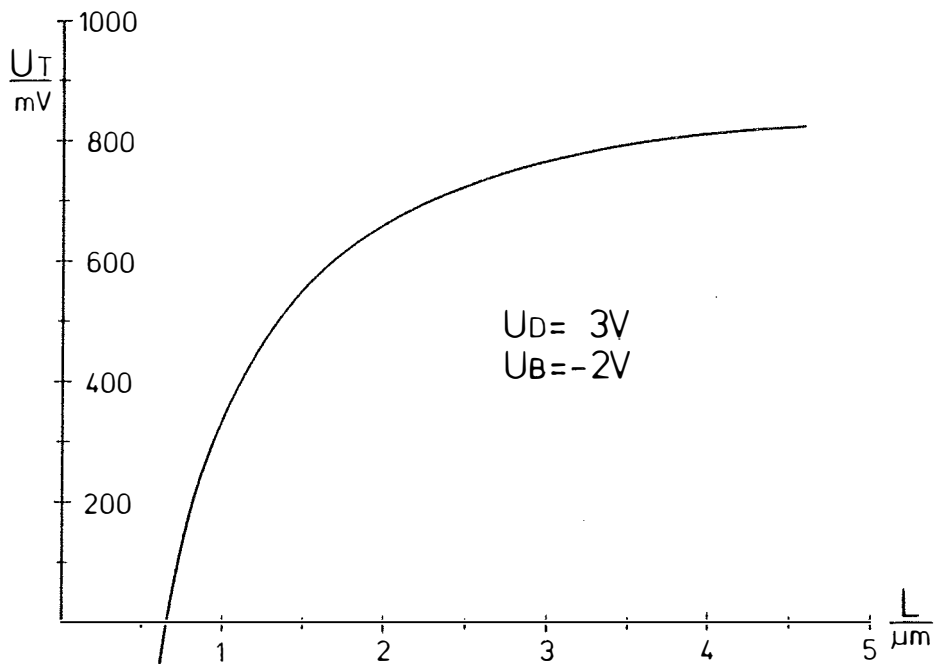becomes meaningless.

Fig. 4.2-2: Threshold voltage versus channel length

Process Sensitivity

Fig. 4.2-2 shows the threshold voltage versus channel length for our devices. An operating point of 3V drain bias and -2V bulk bias has been chosen for the comparison of different channel lengths. To avoid confusion, all the following figures will also refer to this operating point. Fig. 4.2-2 reflects the well known decrease of the threshold voltage with shrinking device length, which becomes dramatic at a length of below one micron.

Usually in papers on short channel MOS transistors a comparison between theoretical curves and selected experimental results is given. Some of them report on statistical measurements (e.g. /42/), but only one paper /184/, to our knowledge, deals explicitly with the sensitivity of an electrical property, namely the threshold voltage. However, with respect to the inherent dependence of most properties on the dispersion of geometry and technology, it seems to be a real necessity to analyse and present these dependences directly. Therefore we carried out numerical investigations to extract the most important sensitivities. A two dimensional simulation program like MINIMOS is excellently suited for numerical investigation of the sensitivity of device properties to dispersion of design and process parameters. First, the physical model parameters of the computer program have to be matched to those corresponding to measured characteristics, i.e. the program has to be "calibrated". This procedure has to be done with non critical transistors with relatively long channels because the measured characteristics should deviate only minimally with inaccuracies in geometry and in technology. This "calibration" procedure should certainly be done for every technology which is to be analyzed numerically as the formulae which are used in a simulation program for modeling the physical parameters (e.g. mobility) are partly heuristic. A few "constants" of those formulae have to be fitted if total agreement of simulation and measurement is desired. It is certainly absurd, and physically invalid, to change the physical model parameters when simulating transistors with just different channel lenghts (for example) because all effects due to changes in the channel length are principally included in the structure of the fundamental semiconductor equations and not in their parameters.

To obtain a sensitivity by computer simulation, one has to vary the interesting parameter (e.g. channel length) in the vicinity of its nominal value and then differentiate after the results (e.g. threshold voltage). This parameter variation must certainly be done within a small range because the validity of linearization on which the whole strategy is based has to be ensured. On the other hand, it is necessary to have a sufficiently large range of parameter variation to avoid cancellation errors at the (numerical) differentiation.

Process Sensitivity

This parameter variation within a small range cannot, in general, be performed experimentally. A minute change of a process parameter which is reproducable piles up tremendous fabrication problems or inherent costs. However, with a fast modeling program the partial derivative of any electrical property with respect to any technological or geometrical parameter can be calculated easily with the outlined strategy. Thus numerical investigations are ideally suited for the performance of sensitivity analysis.
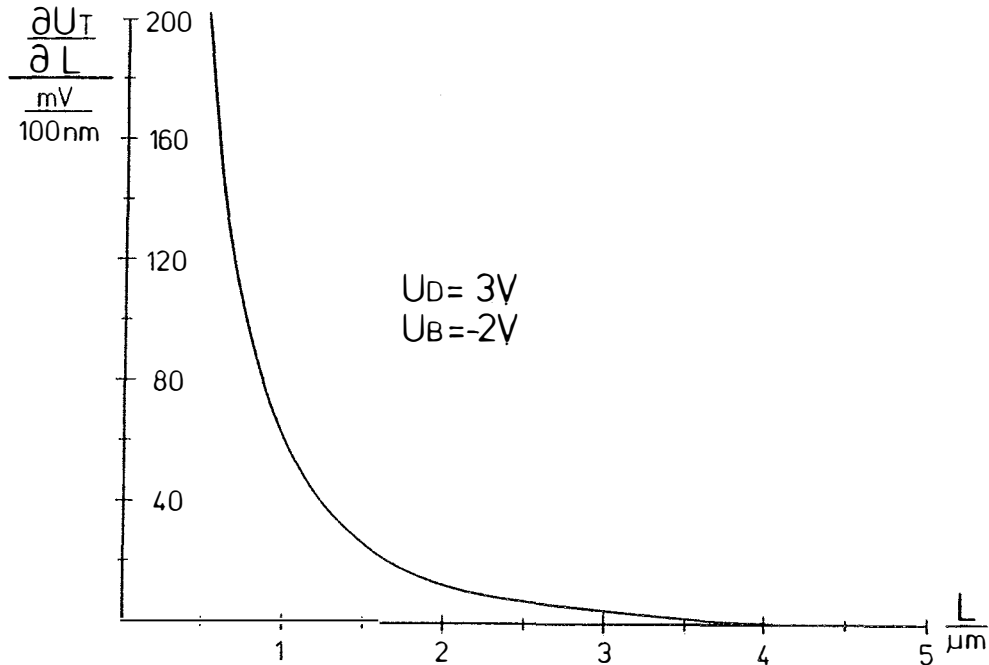
Fig. 4.2-3:   Sensitivity on channel length tolerances

Fig. 4.2-3 shows the partial derivative of the threshold voltage with regard to the channel length versus channel length for our devices; that is, the sensitivity of the threshold voltage on tolerances of the channel length. Assume a transistor with an effective channel length of one micron accurate to ten percent. With this figure one can read an uncertainty of the threshold voltage of +/-60mV.

Fig. 4.2-4 shows the sensitivity of the threshold voltage to the deviation of the oxide thickness. As one probably has not expected at first glance this sensitivity decreases for devices with short channels. This is due to the decreasing influence of the bulk charge with shrinking channel lengths. Note that this figure is qualitatively very similar to the figure showing the

Process Sensitivity

threshold voltage versus channel length (Fig. 4.2-2). This fact can be understood analytically by recalling the simple formula for the threshold voltage:

$$UT \doteq \Phi_{MS} + 2 \cdot \Phi_{F} - (Q_{fs}+Q_{b})/C_{ox}$$

(without short-channel effect)

$$C_{ox} = \varepsilon_{ox}/T_{ox}$$

$$\partial UT/\partial T_{ox} \doteq -(Q_{fs}+Q_{b})/\varepsilon_{ox}$$

$$UT \doteq (\partial UT/\partial T_{ox}) \cdot T_{ox} + const.$$

With an uncertainty of 5% of the oxide thickness, one has an uncertainty of about +/-40mV for a 5 micron device and not even half this value for a 1 micron device. However, one should not
revel in this fact. The decrease of the sensitivity results from the decrease of the controllability of the transistor by the gate.
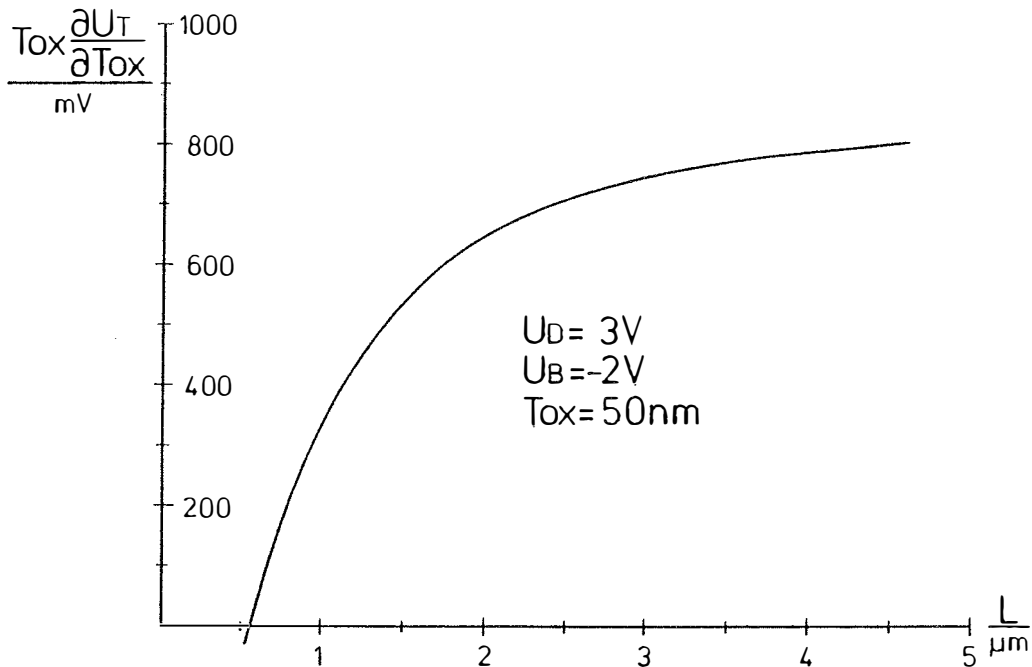
Fig. 4.2-4: Sensitivity on oxide thickness tolerances

Process Sensitivity
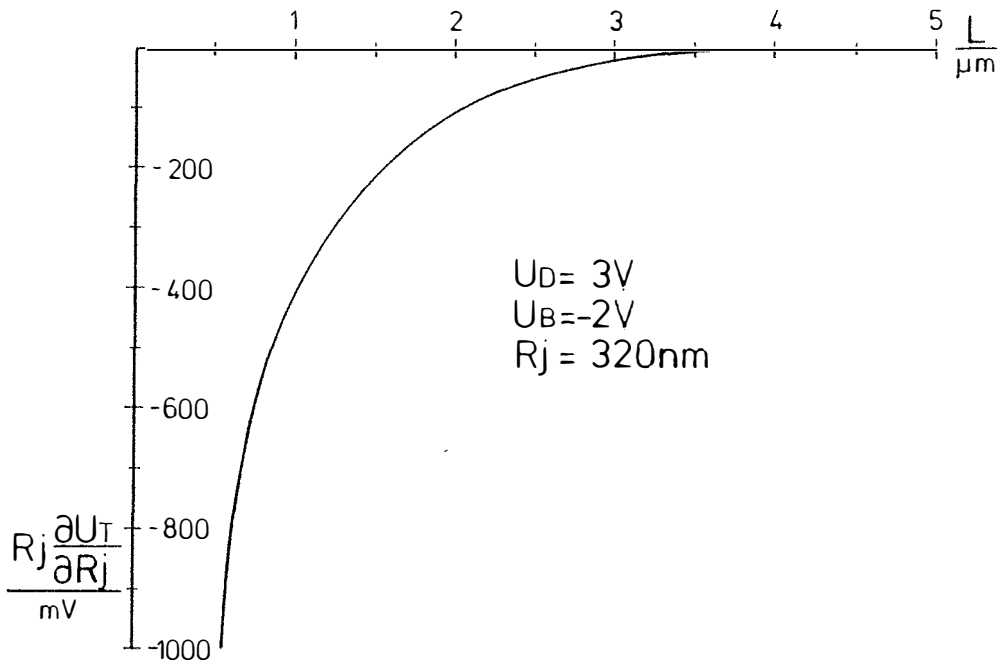
Fig. 4.2-5:   Sensitivity on junction depth tolerances
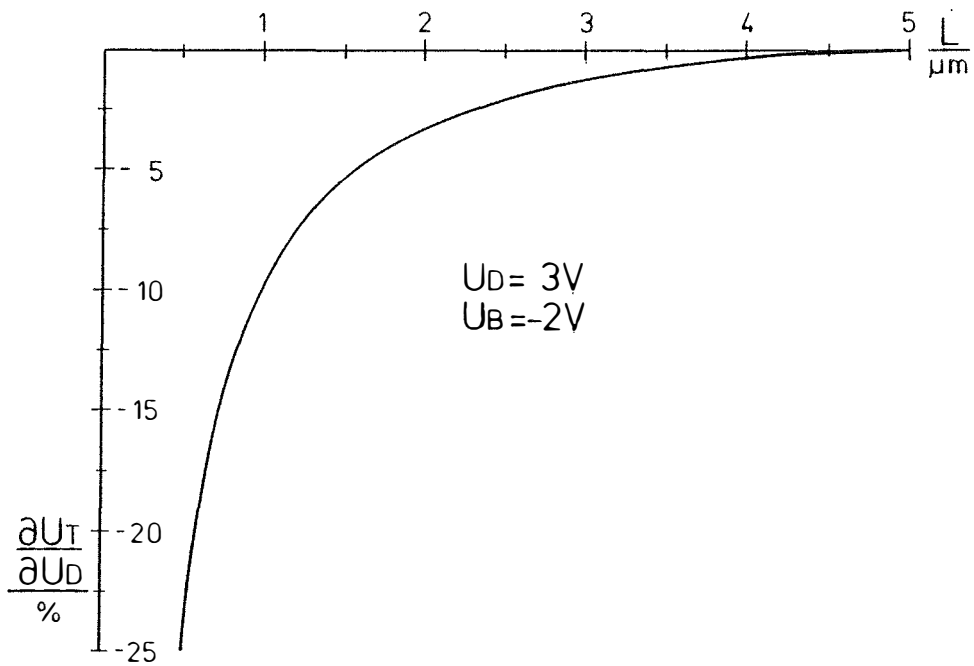
Fig. 4.2-6:   Sensitivity on drain bias variation

Process Sensitivity

Fig. 4.2-5 shows the sensitivity of UT on junction depth tolerances versus channel length. A one micron device with an uncertainty of 10% in the junction depth, thus has an uncertainty of about -/+40mV of the threshold voltage. The underlying physical cause of this sensitivity is the reduction of the channel charge by the depletion regions of source and drain (cf. /168/).

Fig. 4.2-6 shows the sensitivity of UT on drain bias variation. A 300mV change, that is 10% of the applied bias, results in about 30mV change of the threshold voltage for this operating point. Again the modulation of the depletion region of the drain is the relevant physical effect. At first glance it seems to be easy to measure this particular sensitivity in even a minimally equipped laboratory. However, in case of short channel devices just the nominal values of the process and geometry parameters are known for an individual device. The dispersion of these parameters would merely allow to extract bars by statistical measurements which again will make the analysis expensive and time consuming.

Fig. 4.2-7 shows the sensitivity of UT on bulk bias variation. A 200mV change, that is again 10%, results in a threshold shift of about 11mV, which is usually not dramatic. (For the practical problem, however, one has to deal with a sum of all uncertainties. Therefore this influence may also become important.) An interesting detail of this figure is the fact that the sensitivity decreases first with shrinking channel length and at a certain length begins to increase rapidly. This behaviour is caused by a superposition of the short channel effect, which decreases this particular sensitivity, and the punch-through effect, which increases the sensitivity. For long channel devices it is fairly simple to estimate this sensitivity analytically:

$$\partial UT/\partial UB \doteq -1/C_{ox} \cdot \partial Q_b/\partial UB$$

with:

$$Q_b \doteq q \cdot (N_b \cdot y_c + Dose)$$

For the partial derivative only y_c has to be considered:

$$\partial UT/\partial UB \doteq -(T_{ox}/y_c) \cdot (\varepsilon_{si}/\varepsilon_{ox})$$

With a value of about two micrometer for y_c one obtains a sensitivity of approximately -7.5 percent which is confirmed by the more exact two-dimensional calculations.
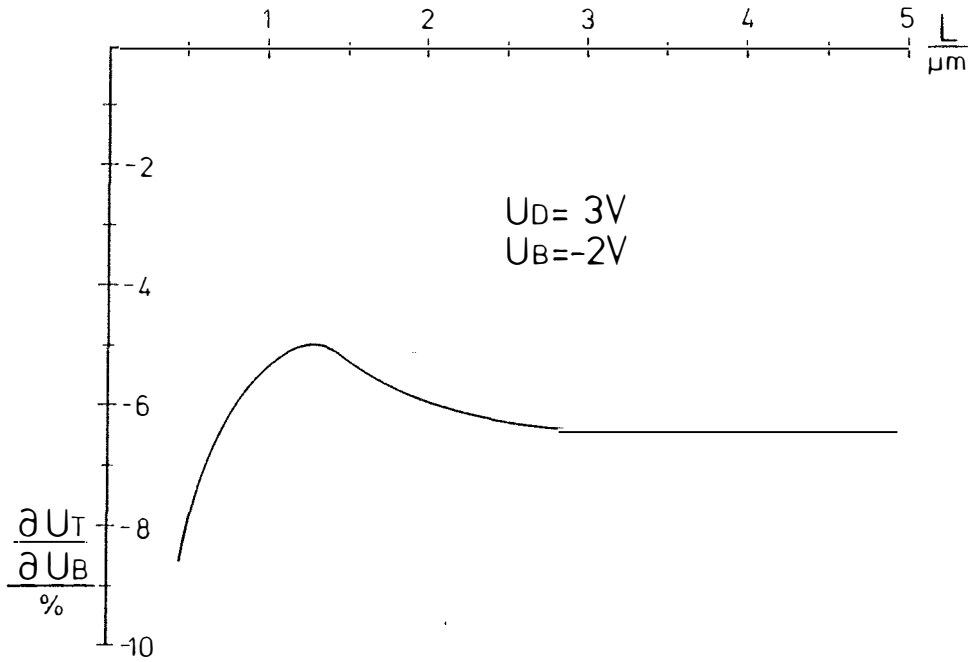
Process Sensitivity

Fig. 4.2-7: Sensitivity on bulk bias variation

$U_D = 3V$
$U_B = -2V$

$$\frac{\partial U_T}{\partial U_B}$$
$\%$

$$\frac{\partial U_T}{\partial Dose}$$
$$\frac{mV}{10^{10} cm^{-2}}$$

$U_D = 3V$
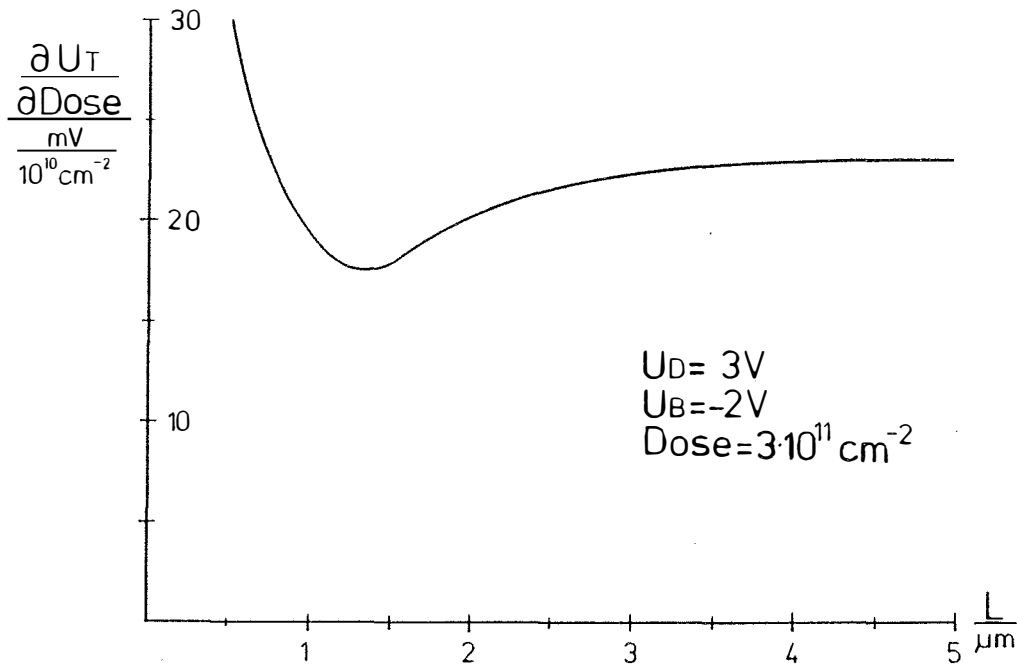$U_B = -2V$
$Dose = 3 \cdot 10^{11} cm^{-2}$

Fig. 4.2-8: Sensitivity on implantation dose tolerances

Process Sensitivity

Fig. 4.2-8 shows the sensitivity of UT on uncertainties of the implantation dose. Qualitatively the superposition of the short channel effect and the punch-through effect is again apparent. The absolute value of this particular sensitivity is low due to the fact that the depletion region below the channel covers the whole implanted region at this operating point. An analytical estimate for the long channel transistor can be obtained in a straightforward way for this sensitivity:

$$\partial UT/\partial Dose \doteq -1/C_{ox} \cdot \partial Q_b/\partial Dose$$

$$= q/C_{ox} = 23 \ mV/10^{10} cm^{-2}$$

Fig. 4.2-9 shows the temperature coefficient of the threshold voltage for our devices. We have, qualitatively, a similar behaviour to that already discussed, namely the superposition of short channel effect and punch through. The absolute value is around -1mV/K. The qualitative behaviour as well as the absolute value of this sensitivity have been verified by fairly complicated experiments /159/.
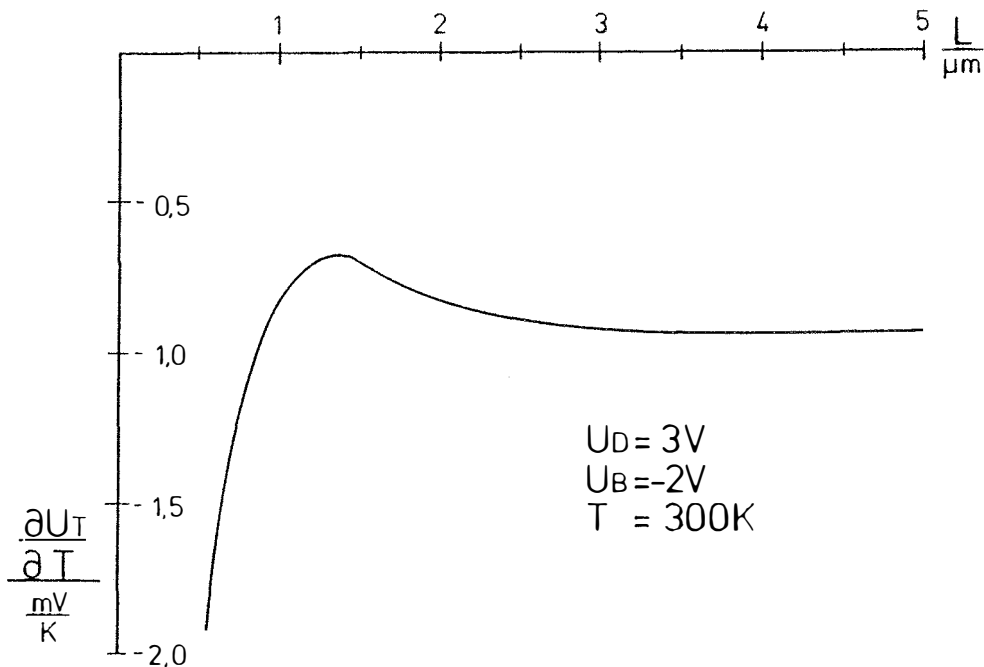

Fig. 4.2-9: Sensitivity on temperature variation

Process Sensitivity

The partial derivatives denote isolated sensitivities on a certain set of parameters. These values show which parameters are the most critical ones. However, in addition, a global sensitivity number indicating the cumulative effect of the isolated sensitivities is useful. The global sensitivity is related to a certain technology and its expected application. It should indicate the limit of channel length reduction. To obtain such a global number typical ranges of deviation of design parameters have to be specified. The table in Fig. 4.2-10 is an example for such a specification. In this example a rather small value of the absolute uncertainty of the channel length (100nm) has been chosen. For long devices this value is unrealistic, but in consideration of a one micrometer technology 100nm absolute uncertainty represents 10 percent relative dispersion, which is relatively large. The tolerances of the remaining parameters in Fig. 4.2-10 , however, represent a good laboratory standard.

Parameter	X	DX	%
L		100nm	
TOX	50nm	2.5nm	5
RJ	320nm	32nm	10
UD	3V	150mV	5
UB	-2V	100mV	5
AKEV	35keV	0.7keV	2
DOSE	$3.10^{11} cm^{-2}$	$6.10^{9} cm^{-2}$	2

Fig. 4.2-10: Desired process and operating tolerances

Process Sensitivity

$$\frac{\sigma}{mV}$$

$$\text{—— } \sigma_D = \frac{\partial U_T}{\partial L}\,\Delta L$$

$$\text{---- } \sigma_W = \sqrt{\sum\left(\Delta x\,\frac{\partial U_T}{\partial x}\right)^2}$$

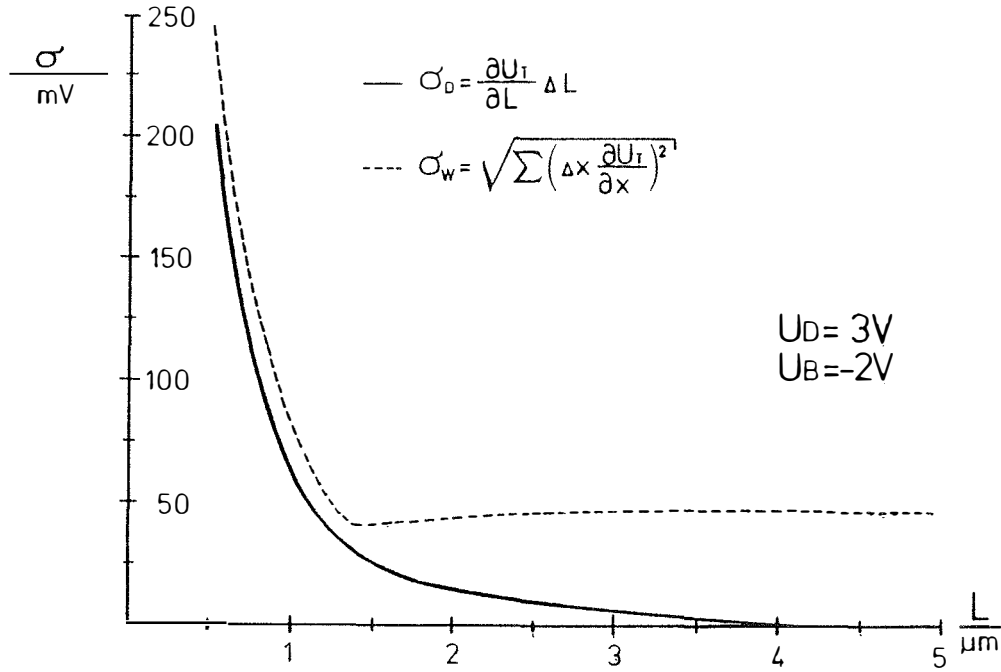$$U_D = 3V$$
$$U_B = -2V$$

$$\frac{L}{\mu m}$$

Fig. 4.2-11: Reproducability of the analyzed devices

Fig. 4.2-11 shows the global threshold voltage sensitivity based on the specifications of Fig. 4.2-10. σ_D denotes the uncertainty of the threshold voltage for identical devices on the same chip. D stands for device. This sensitivity is given by just the length influence, as the other parameters are commonly very homogeneous across one chip. σ_W, W stands for wafer, denotes the uncertainty for identical devices on wafers, which have been fabricated with different charges. Here one has to use a Euclidian norm over all deviations. Note that this value is highly constant down to a certain channel length, but then increases dramatically. The channel length at which the excellently pronounced knee is located, 1.4 microns for our devices, can, therefore, be interpreted as the practical limit of channel length reduction due to threshold uncertainty. Nevertheless should it be noted that the data in Fig. 4.2-10 are to be understood as an example which is mainly of importance for our technology.

Process Sensitivity

4.3 Breakdown Phenomena

To increase the number of functional units per chip, it is necessary to decrease the size of the devices (e.g. channel length and channel width for MOS transistors). As the performance of a device depends strongly on its geometry, any reduction requires certain design rules (c.f. /43/ for MOS devices). However, in recent years devices have been miniaturized without reduction of supply voltage, mainly to stay compatible with existing circuits and to maintain an acceptable signal to noise ratio. Unfortunately problems with punch through and avalanche breakdown arise from the reduction of size. Punch through can be controlled relatively well by technological steps as already outlined in the last chapter for an MOS transistor. But there exists an increasing demand for a transparent description of the physical processes which lead to avalanche breakdown.

Avalanche problems have so far been treated /86/, /165/ in the following manner: First Poisson's equation is solved to obtain a solution for the electrical potential distribution and then the ionization integral is evaluated by integrating the strongly field dependent ionization coefficients over the high field region. As result multiplication factors are obtained which describe the increase of current due to avalanche. Since the carrier densities need not be calculated, this method seems to be very efficient in calculating breakdown voltages. However, any feedback of the increase in carrier densities on the electrical field is, therefore, neglected. A more serious treatment requires the solution of both carrier continuity equations with proper modeling of the generation term /133/, /135/.

In this section calculations for a one micron gate length n-channel MOS transistor are presented. The lateral subdiffusion and the junction depth of the source and drain regions are 0.2 and 0.3 μm, respectively. A deep channel implantation with fairly high dose was supposed to have been performed to suppress punch-through.

Fig. 4.3-1 shows calculated drain and bulk currents versus drain voltage for that transistor. For $U_{GS}=1V$ breakdown is reached at $U_{DS}=5.6V$ whereas 8.4 Volts are necessary to lead the device into breakdown if no gate voltage is applied. On first glance that seems to be paradox, if one considers that $U_{GS}=0V$ certainly causes larger peak values of the electric field. The explanation of this phenomenon lies in the low current level.
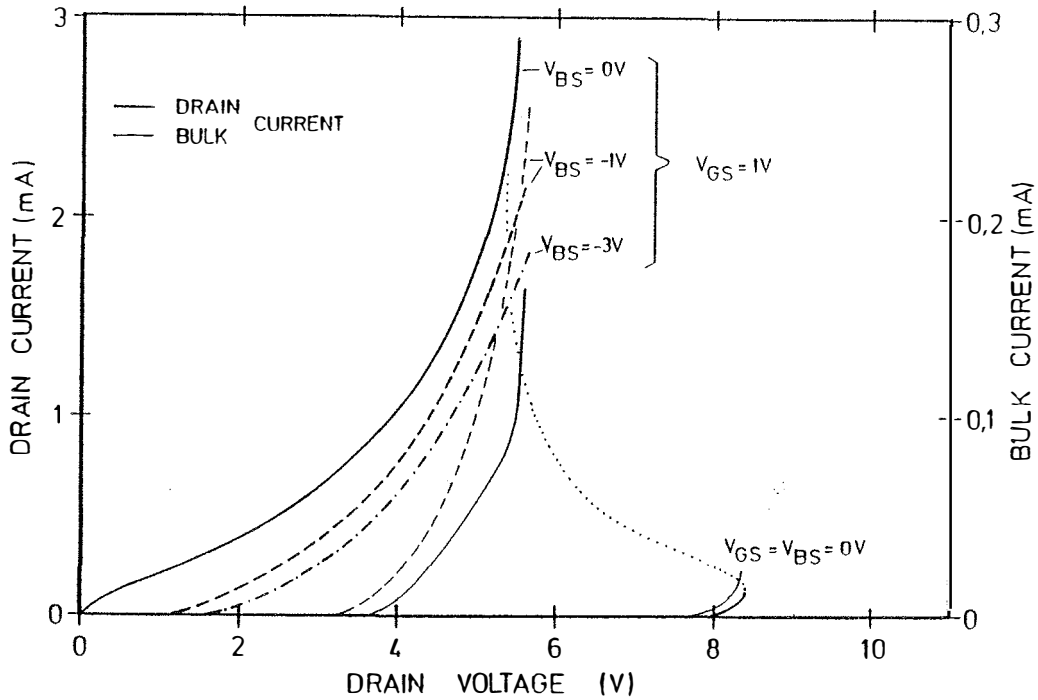
Fig. 4.3-1 Drain and bulk current characteristics

Although the probability of ionization is larger for U_{GS}=0V than for U_{GS}=1V, the generation rate still remains small as there is little current flow causing ionization. With increasing drain voltage the drain current and consequently avalanche generation as well as hole density increase. This additional space charge even lowers the potential barrier between source and bulk. Now an internal feedback mechanism exists which acts as follows: Because of the lower potential barrier the electron current injected by source, and consequently, the avalanche generation increase. Thus the hole density rises even more and, in turn, further lowers the potential barrier. Once the feedback gain becomes unity the node currents rise unlimited unless controlled by external resistors in the current paths. Furthermore, owing to the higher current level, the situation now becomes more and more similar to the situation at larger gate voltages. The I-V characteristic, therefore, has to move towards the U_{GS}=1V characteristic and the drain voltage decreases with increasing drain current. This effect implies negative resistance and is

Breakdown Phenomena

usually called "snap-back". The voltage drop of the hole current at the parasitic resistor of the deep bulk also lowers the potential barrier and thus enhances the feedback gain.

Applying a negative bulk voltage renders breakdown more difficult although it increases the bulk current level. The reason for this lies in the hole density which is decreased by applying a more negative bulk bias which attracts the holes.
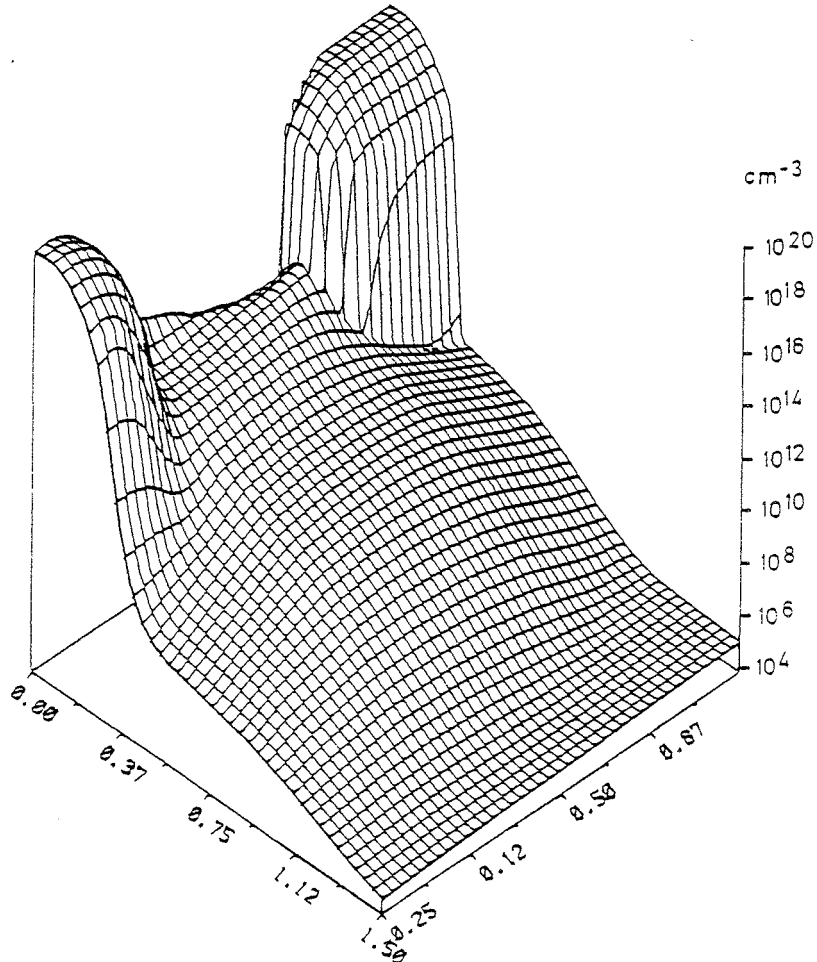
Fig. 4.3-2 Concentration of electrons (U_{GS}=0V, U_{DS}=8V)

There exists an additional feedback mechanism apart from the one just mentioned: The carriers generated by ionization cause again ionization. This effect leads to an "avalanche-like" increase of both carrier densities, and determines the breakdown voltage of a p-n junction. The feedback depends on the ionization ability of both carrier types and is of little

Breakdown Phenomena

importance in our case. For MOS transistors the mechanism described above is much stronger and is active with even vanishing ionization ability of holes.

In the following we should like to discuss internal physical quantities at $U_{GS}=0V$, $U_{DS}=8V$, and $U_{GS}=2V$ $U_{DS}=5.6V$. These operating points have been chosen to explain clearly the physical phenomena which eventually lead to the snap back effect. The computed drain currents are about 20μA and 15mA, respectively. Since the $U_{GS}=2V$ characteristic was out of locus bounds, it is not drawn in Fig. 4.3-1.
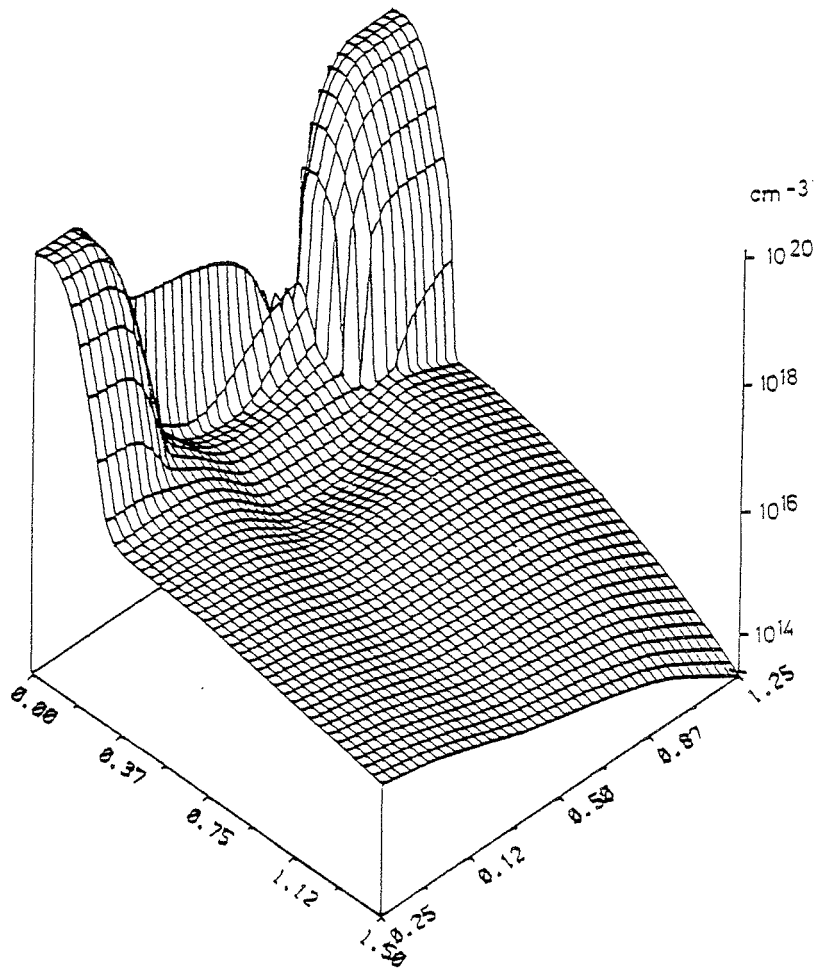
Fig. 4.3-3 Concentration of electrons ($U_{GS}=2V$, $U_{DS}=5.6V$)

Fig. 4.3-2 and Fig. 4.3-3 show the electron distribution for both operating points in a logarithmic scale. At the first operating point, Fig. 4.3-2, the transistor is turned off; there

Breakdown Phenomena

is no inversion layer between the source and drain regions which can be found, as expected, in Fig. 4.3-3 at the second operating point. It should be noted that in Fig. 4.3-3 the electron density does not drop below the intrinsic number in contrast to Fig. 4.3-2. The reason for this is source barrier lowering brought about by the increased hole density.

The corresponding hole densities are given in Fig. 4.3-4 and Fig. 4.3-5, respectively. One should bear in mind that all the holes outside the undisturbed bulk region are generated by impact ionization. In agreement with the electron densities the hole density is also much larger for U_{GS}=2V. The large hole density near the source partially compensates the acceptor doping. Thus the potential barrier at the source is lowered and high electron injection from the source region ensues.
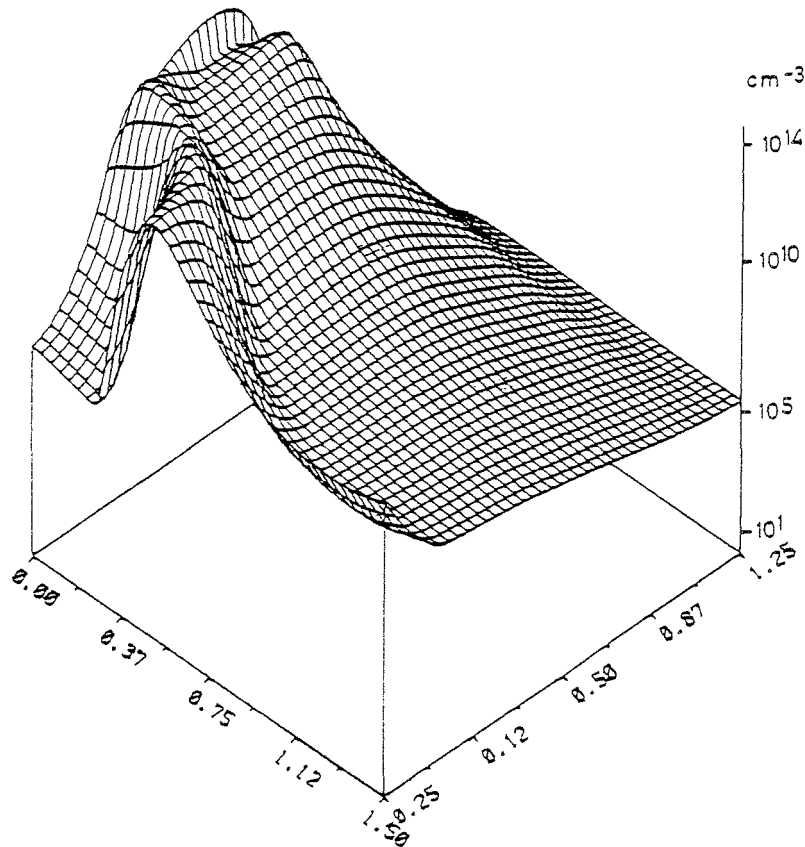
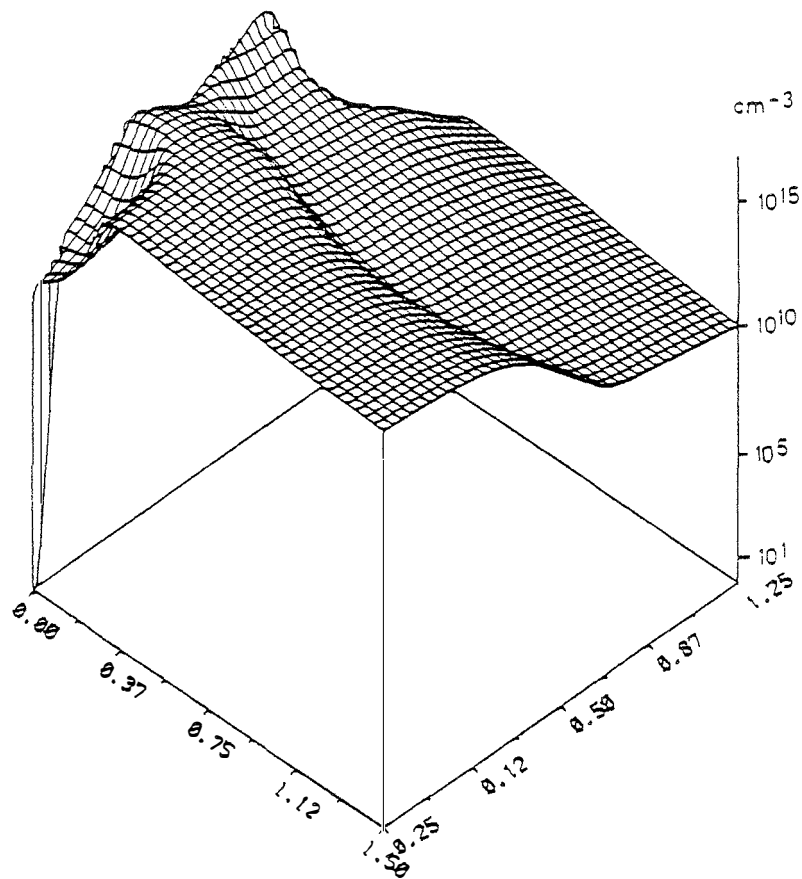Fig. 4.3-4 Concentration of holes (U_{GS}=0V, U_{DS}=8V)

Breakdown Phenomena

Fig. 4.3-5 Concentration of holes $(U_{GS}=2V, \ U_{DS}=5.6V)$

Looking at Fig. 4.3-1 again, we find that the negative resistance branch of the $U_{GS}=0V$ characteristic for large current levels leads into a vertical slope, i.e. the decrease of U_{DS} is stopped. The corresponding drain voltage is called "sustain voltage"; it increases weakly with increasing U_{GS} because a large gate bias smoothes the electric field distribution thus lowering its peak value. The existence of a nearly unique sustain voltage can be explained by heavy recombination as demonstrated in /134/. A good many interesting results concerning avalanche breakdown phenomena have been published in /33/, /57/, /86/, /114/, /133/.

Breakdown Phenomena

4.4 A Simple n-MOS Inverter with Depletion Load

Simple inverters are eminently important for the design of an integrated logic circuit. They consist of a driver transistor and a load which is usually another transistor as these devices can be fabricated more easily than resistors. The driver transistor should be normally off in order to turn it off easily. With the input voltage logically high, this device should turn on and should exhibit high conductivity between source and drain; the voltage drop across this transistor will, therefore, be small and the output voltage will be very low. The load transistor has to combine two features: when the driver transistor is turned on, nearly the total voltage should drop across the load and the current must be moderate. On the other hand, when the driver transistor is turned off, it should be able to supply the output without a significant voltage drop. These features can be achieved by a depletion (normally on) transistor with a smaller channel width than the driver transistor.
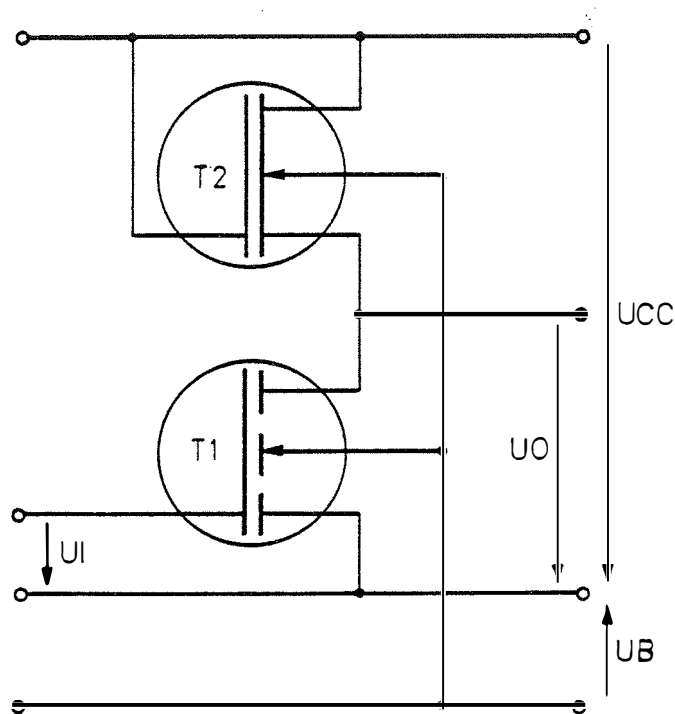

Fig. 4.4-1: A simple n-MOS inverter

Fig. 4.4-1 shows a typical n-MOS inverter with depletion load. In the following we shall analyze a depletion transistor which can be used as depletion load device together with device 3 of chapter 4.1.

A Simple n-MOS Inverter with Depletion Load

To realize a depletion transistor, donor impurities have to
be implanted into the channel. With this technique a conducting
channel between source and drain is created. However, if the
dose and depth of the implantation are moderate, the device can
still be controlled via the gate voltage. In this way the
threshold voltage can be shifted towards a negative value.

Fig. 4.4-2 shows the doping profile of the load device
indicating the shallow implantation at the surface of the
channel. The implantation has been carried out with Antimony, a
dose of 10^{12}cm^{-2}, and an energy of 180 keV. The other device
data are identical with the third transistor of chapter 4.1. The
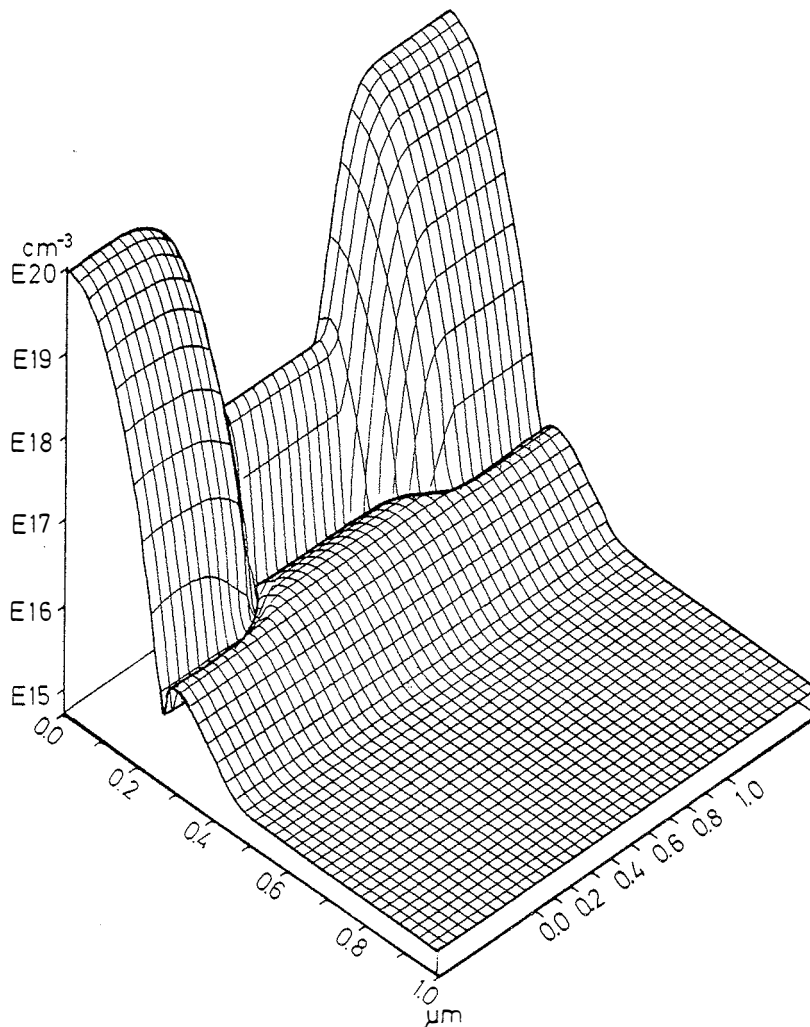shallow Boron implantation, however, has been omitted.

Fig. 4.4-2: Doping profile of depletion transistor

A Simple n-MOS Inverter with Depletion Load

To get a more principal understanding of depletion-mode devices, some internal distributions will be discussed for the same operating point as has been chosen in chapter 4.1 (UD=3V, UG=0V, UB=0V).
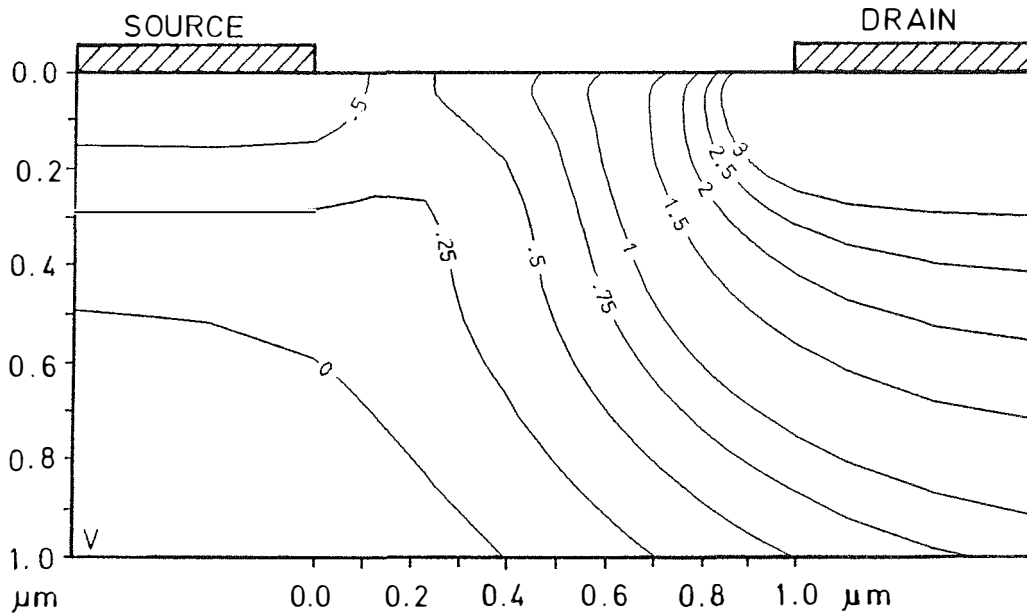

Fig. 4.4-3: Potential distribution in depletion device

The potential distribution is given in Fig. 4.4-3. No effective barrier between source and channel can be seen in contrast to the pictures of chapter 4.1. A source-channel diode does not exist and the built-in potential at the n^+n junction remains small.

Fig. 4.4-4 shows the electron distribution in the load device. An electron channel with its maximum electron density at a depth of about 100nm can be seen. The onset of pinch-off is due to the fact that donor channel implantation was low enough to be easily depleted. Similarly, the channel could be depleted by a negative gate voltage.

In Fig. 4.4-5 the distribution of the lateral current density is presented. The channel is rather wide because of the donor implantation profile; that can be also deduced from the electron density. Thus the load device is able to conduct much more current at the same peak current density than an enhancement device. It is also to note that no indications of punch-through exist.

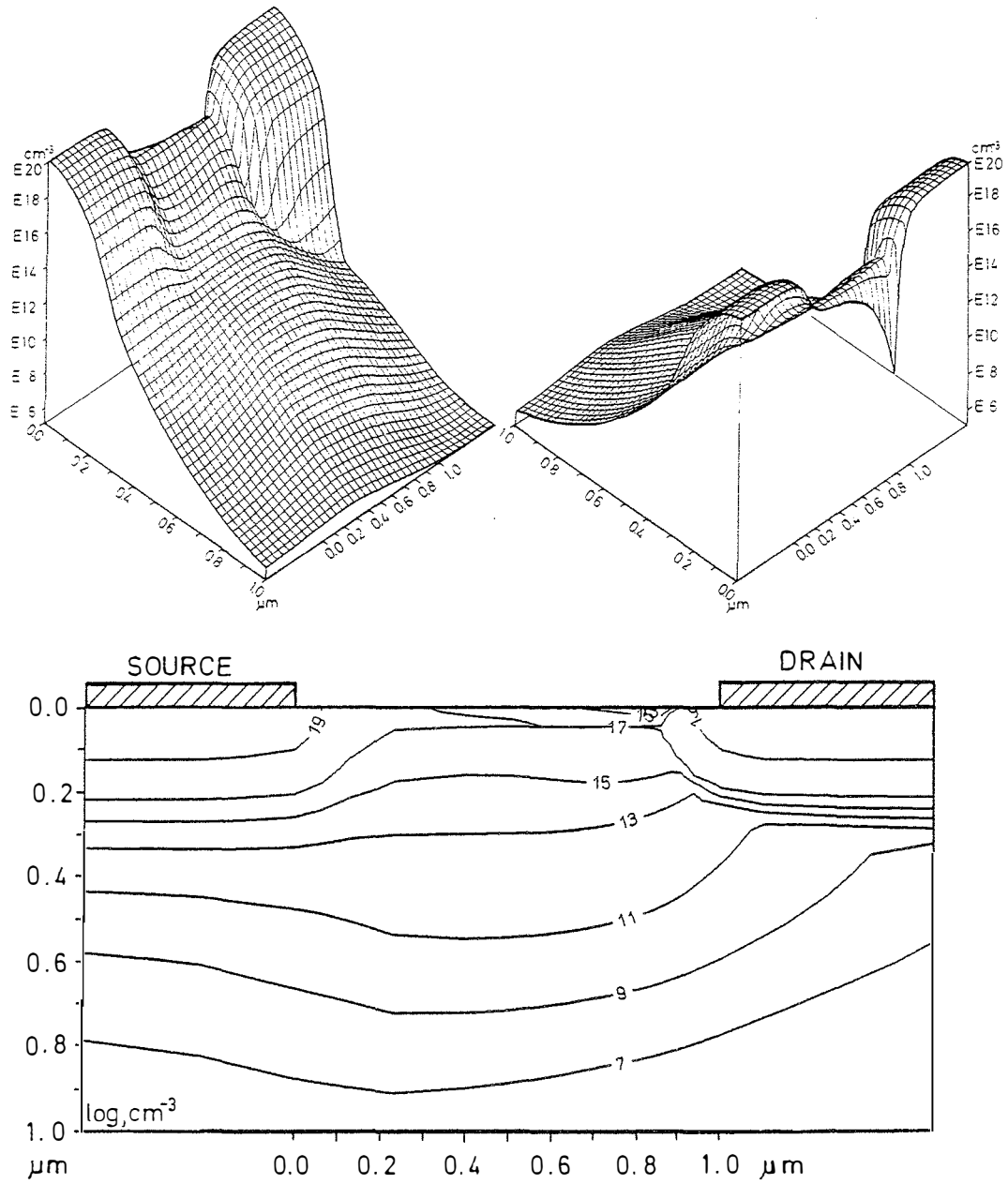A Simple n-MOS Inverter with Depletion Load

Fig. 4.4-4: Electron distribution in depletion device

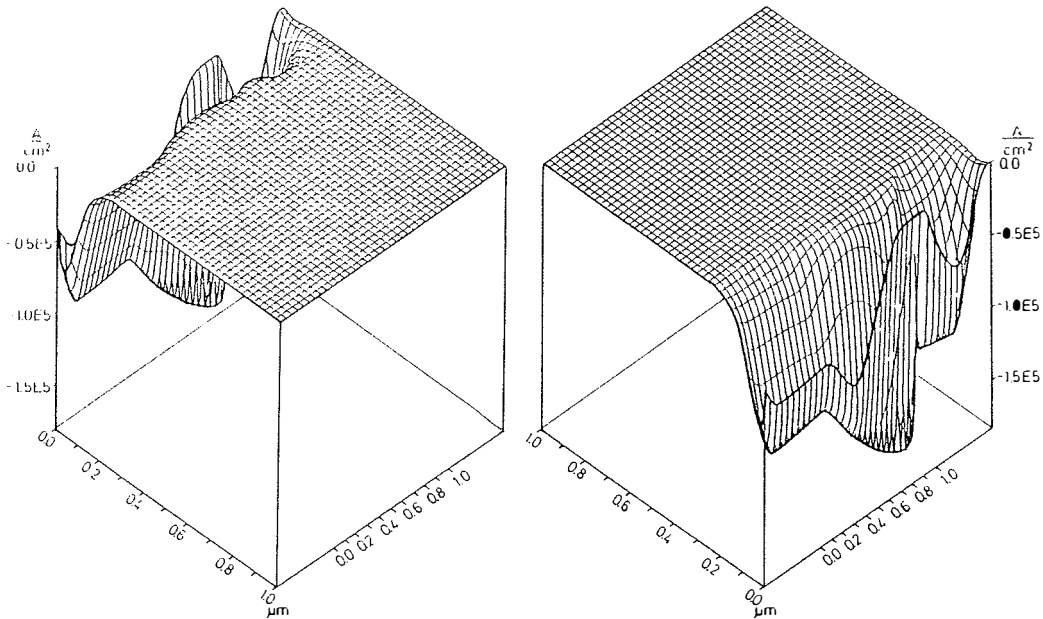A Simple n-MOS Inverter with Depletion Load

Fig. 4.4-5: Lateral current density in depletion device

The transfer function of an n-MOS inverter consisting of two transistors – as just discussed – has been calculated. The channel widths for the enhancement and depletion transistors have been chosen to be 20 μm and 2.5 μm, respectively to get a proper ß ratio /72/. Substrate voltage has been set to -2 V to increase the threshold voltage of the enhancement transistor and to enhance the signal to noise ratio. Fig. 4.4-6 shows the output characteristic of the driver transistor in solid and the load characteristic in dashed lines. As drain and gate voltages are identical for the load transistor and the threshold voltage is negative, this device always operates in the triode region. In the given circuit the load characteristic is linear which is due to the compensation of substrate and drain voltage influence on threshold voltage.

The transfer characteristic is given in Fig. 4.4-7. The low-level is 0.2 Volt, the high level is 3 Volt and is identical with the supply voltage because the load transistor does not produce a voltage drop without current flow. The noise margin at the low-level is very low (0.2 Volt) which is a well known problem with miniaturized logic circuits. The reason for this unfortunate phenomenon is to find in the low threshold voltage of the driver transistor. The noise margin at the high-level is 1.36 Volts and is certainly sufficient. Voltage amplification is -2.5 and should be acceptable. However, the performance of the

A Simple n-MOS Inverter with Depletion Load

inverter could definitely be improved by optimizing the device parameters. This offers a broad field of application for the present model.

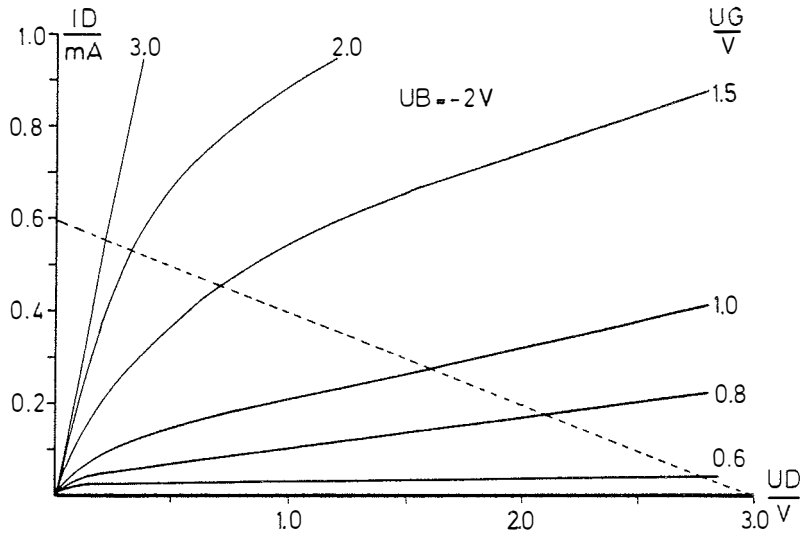
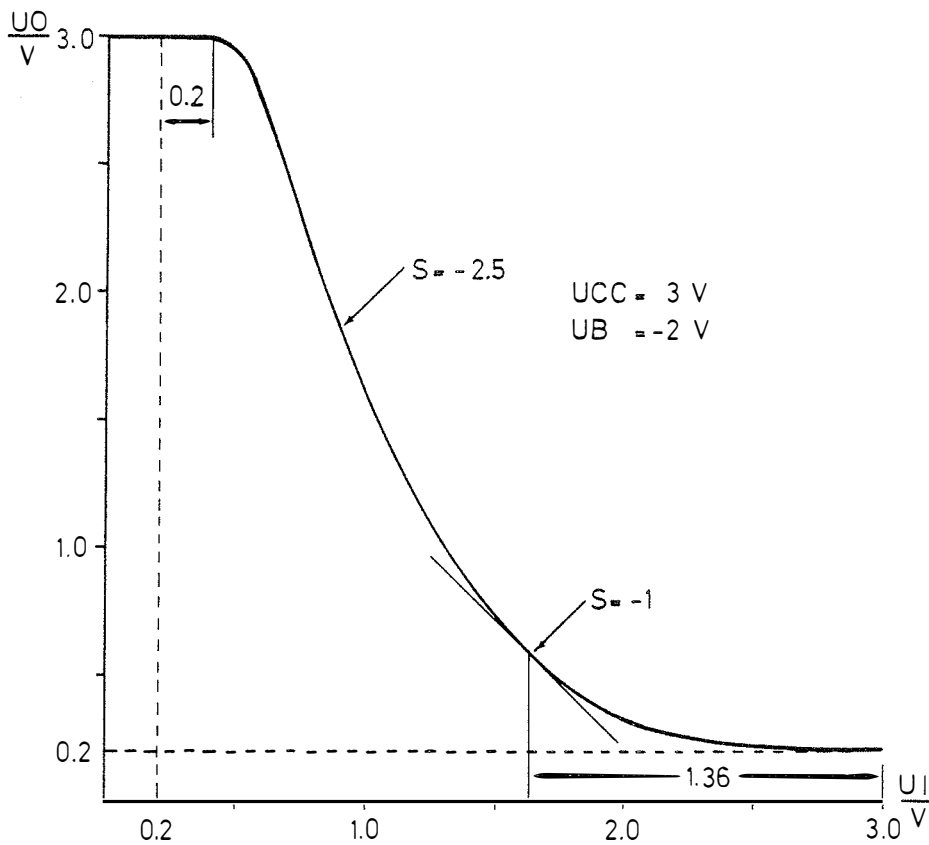Fig. 4.4-6: Output charateristics of drive and load device

Fig. 4.4-7: Transfer characteristic of n-MOS inverter

A Simple n-MOS Inverter with Depletion Load

4.5 A CMOS Inverter

In this chapter we shall discuss a simulation of a CMOS inverter which is the basic cell of any integrated circuit in CMOS technology. Of primary interest in this technology are power consumption, propagation delay and impact ionization, the latter mainly because of the disadvantageous "Latch-Up" which is induced by substrate current. The device data for the inverter which will be discussed here are summarized in Fig. 4.5-1.

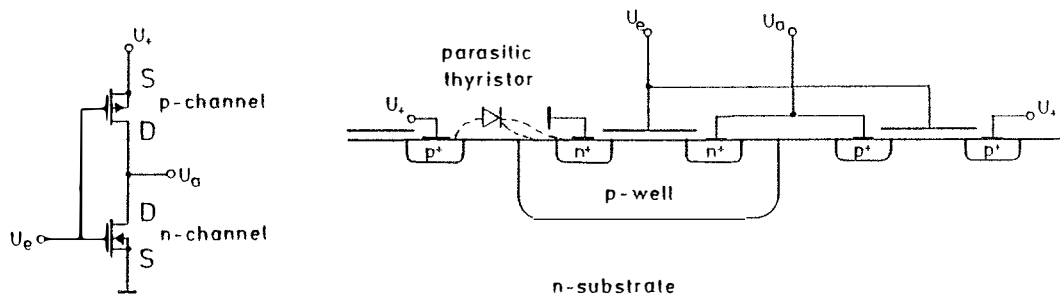|	n-channel	p-channel	
r_j	0.36	0.36	μm
L	2.6	2.6	μm
L_{eff}	2.2	2.2	μm
W	20	20	μm
t_{ox}	30	30	nm
Gate	n-poly Si	p-poly Si	
Doping	p-well $2.4 \cdot 10^{15}$	substrate $6 \cdot 10^{14}$	cm^{-3}
	channel implantation		
Dose	$2 \cdot 10^{12}$ (B)	$1.3 \cdot 10^{12}$ (As)	cm^{-3}
Energy	40	180	keV

Fig. 4.5-1: Process data of CMOS devices

Fig. 4.5-2: The basic diagram of the inverter

The basic diagram of the CMOS Inverter and its principal technological realisation are given in Fig. 4.5-2. Using the process data of Fig. 4.5-1, threshold voltages of about 1.2V (n-channel transistor) and -1.2V (p-channel transistor) are achieved. As these values are rather large, good signal-to-noise behavior can be expected in contrast to the inverter discussed in the previous chapter.

The output diagrams for both independent transistors have been calculated within the range $0 < U_{DS} < 5V$ for $U_{GS}=2$, 3, 4, and 5 Volts and are given Fig. 4.5-3. Although the threshold voltages of both transistors are equal (their absolute values), the current in the n-channel transistor is much larger than in the p-channel transistor (for given drain and gate bias) which has to be contributed to the lower mobility of holes. The difference in the current values is less pronounced for high drain voltages because the saturation velocities of both carrier types only differ slightly.
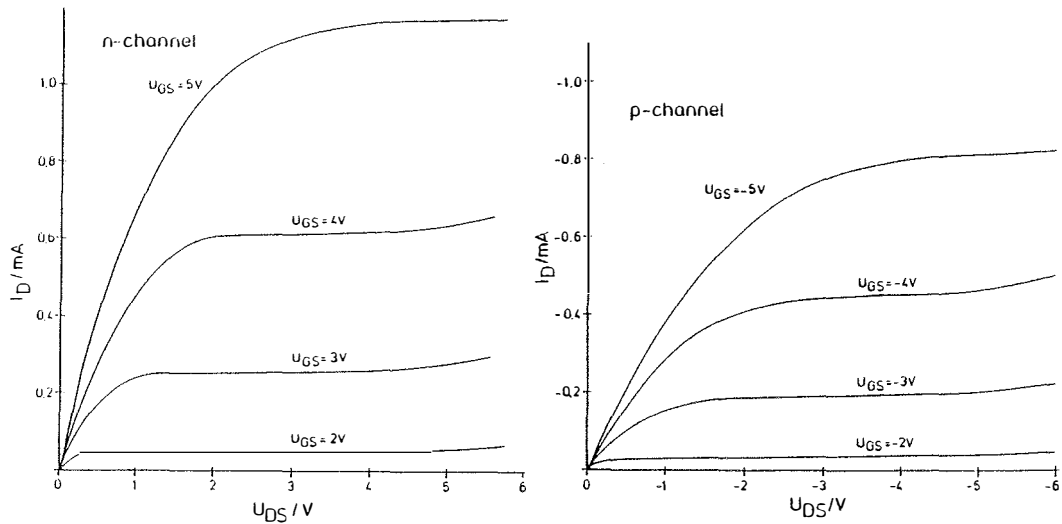

Fig. 4.5-3:   Output diagrams for both transistors

The output resistance in the pentode regime is very large, thus demonstrating that both transistors exhibit long channel behavior which is caused by the large channel doping. The long-channel behavior can also be seen in the potential distribution in Fig. 4.5-4 for both transistors for a subthreshold case ($U_{GS}=\pm 0.5V$, $U_{DS}=\pm 5V$). The surface potential is constant along the channel up to the pinch-off region. This is a typical subthreshold behavior of long channel devices. The pinch-off region extends more towards source in the depth than at the surface as the doping level decreases with increasing distance from the surface. In the p-MOST the pinch-off region is also more extended than in the n-MOST which is due to the different doping levels.

A CMOS Inverter

$U_{GS} =$	.50 V
$U_{DS} =$	5.00 V
$U_{BS} =$	0.00 V

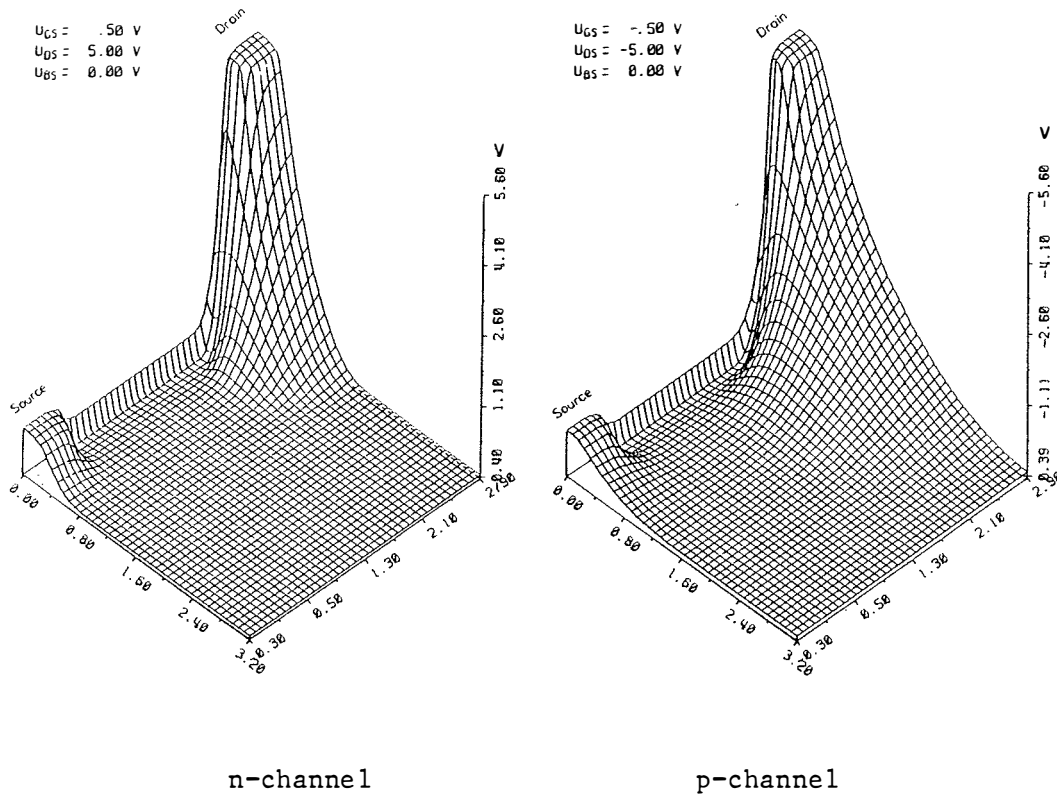| $U_{GS} =$ | -.50 V |
| $U_{DS} =$ | -5.00 V |
| $U_{BS} =$ | 0.00 V |

n-channel   p-channel

Fig. 4.5-4: Potential distribution

The transfer characteristic for the given transistor pair has been calculated and is shown in Fig. 4.5-5. Because of different carrier mobilities in both transistors this characteristic is shifted by about 90mV from the symmetry-line. As expected, noise immunity is very good. Specifying $U_a < 0.5V$ for the L-level and $U_a > 4.5V$ for the H-level it can be deduced from Fig. 4.5-5 that these levels are obtained at $U_e > 2.5V$ and $U_e < 2.22V$, respectively. Voltage amplification is very large in the active region ($v_u = -50$) because both transistors operate in the pentode regime with very large output resistance.
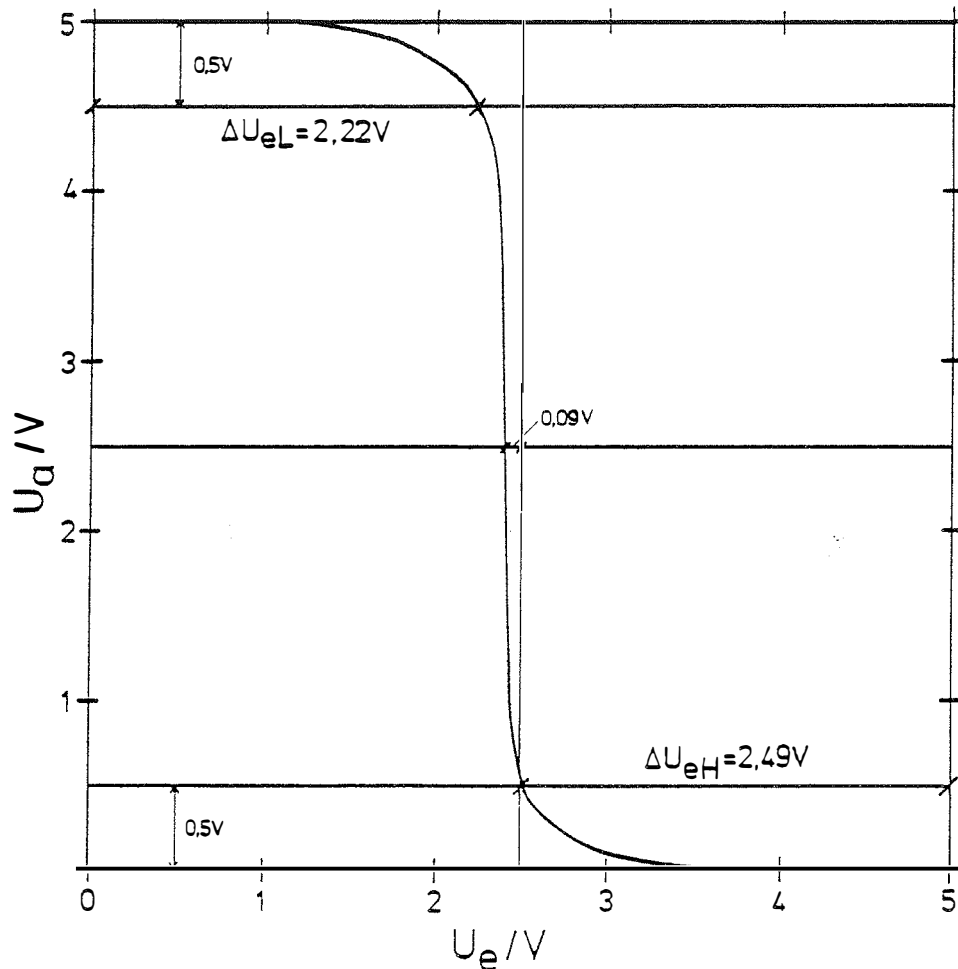
A CMOS Inverter

Fig. 4.5-5: Transfer characteristic

Fig. 4.5-6 shows the supply current as function of the input voltage. The very low stand-by current is the main power of the CMOS technology and has to be contributed to the excellent subthreshold behavior of the simulated transistors. However, the benefit of little power consumption can only be utilized in the static or low-frequency operation as parasitic capacitances have to be charged and discharged during any logic sweep. Another component contributing to increased power consumption in the dynamic operation is the current flux through the transistors during the change of the input voltage.
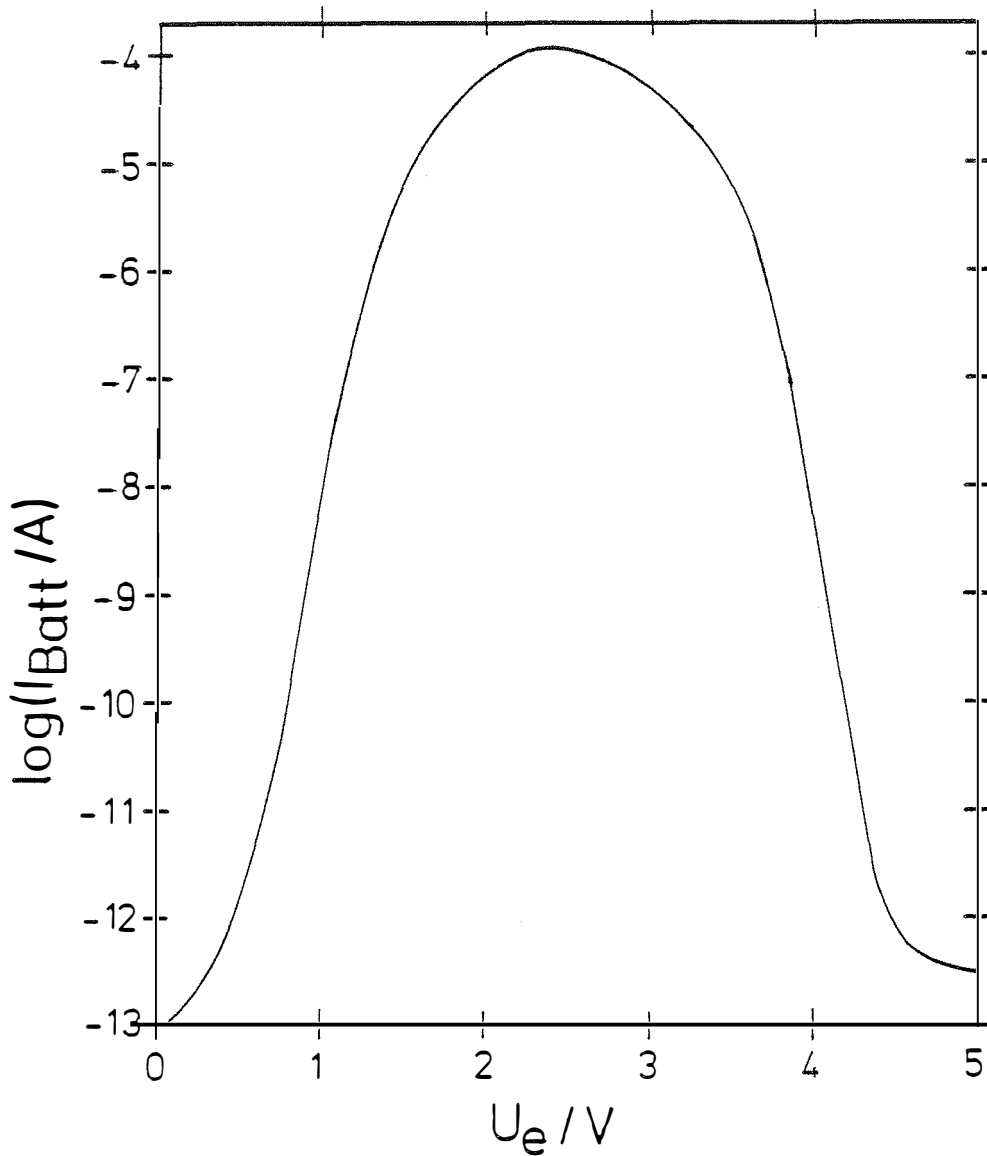
A CMOS Inverter

Fig. 4.5-6: Supply current versus input voltage

Now let us estimate the propagation delay for the chosen inverter. As our analysis is only static, no exact calculation can be presented here. However, the transistors are usually so fast that switching speed is determined mainly by parasitic capacitances rather than by the rise times of the node currents of the transistors. So let us neglect any transient behavior of the transistors themselves. If we apply an ideal L-H voltage jump on input, the p-channel transistor is turned off instantly and can, therefore, be ignored (cf. Fig. 4.5-7). The $u_a(t)$ characteristic is given by the integral equation

A CMOS Inverter

$$u_a(t) = u_a(0) - \int_0^t \frac{i_d(t)}{C_a} \, dt \qquad\qquad (4.5.1)$$

The fall time t_f, defined as the time after which the L-level at the output has been reached, can be found by numerical integration of the inverse of the output characteristic for $U_{GS}=5V$ (Fig. 4.5-5). t_f is proportional to the output capacitance and is $t_f=4.3ns$ for $C_a=1pF$. The delay is larger for the L-H swing at the output because of the smaller conductance of the p-channel transistor ($t_r=7.6ns$).
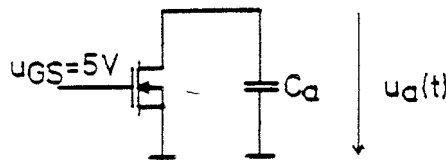

Fig. 4.5-7: Simplification for estimation of t_f

Figure 4.5-8 demonstrates the substrate currents in both transistors (p-well current in the n-MOST) as function of the input voltage. For low input voltage the n-MOST is turned off, the supply voltage totally drops along this transistor and U_{DS} vanishes for the p-MOST. When the input voltage is increased, the transfer current rises exponentially (subthreshold regime of the n-channel transistor) and the same applies to the p-well current. However, when the input voltage approaches about 2 Volts, the output voltage, which is the drain-to-source voltage of the n-channel device, and consequently the p-well current start to decrease. At about 2.4Volts the output voltage drops rapidly and the p-channel device now yields substrate current and the p-well current vanishes. Increasing the input voltage further, the p-channel finally comes into the subthreshold region (U_e >3.8V) and the substrate current again decreases. The p-MOST exhibits larger leakage current than the n-MOST as its doping level is in general smaller. However, as dark space phenomena /123/ are not accounted for in the model, the substrate current level may be considerably smaller than computed.

A CMOS Inverter
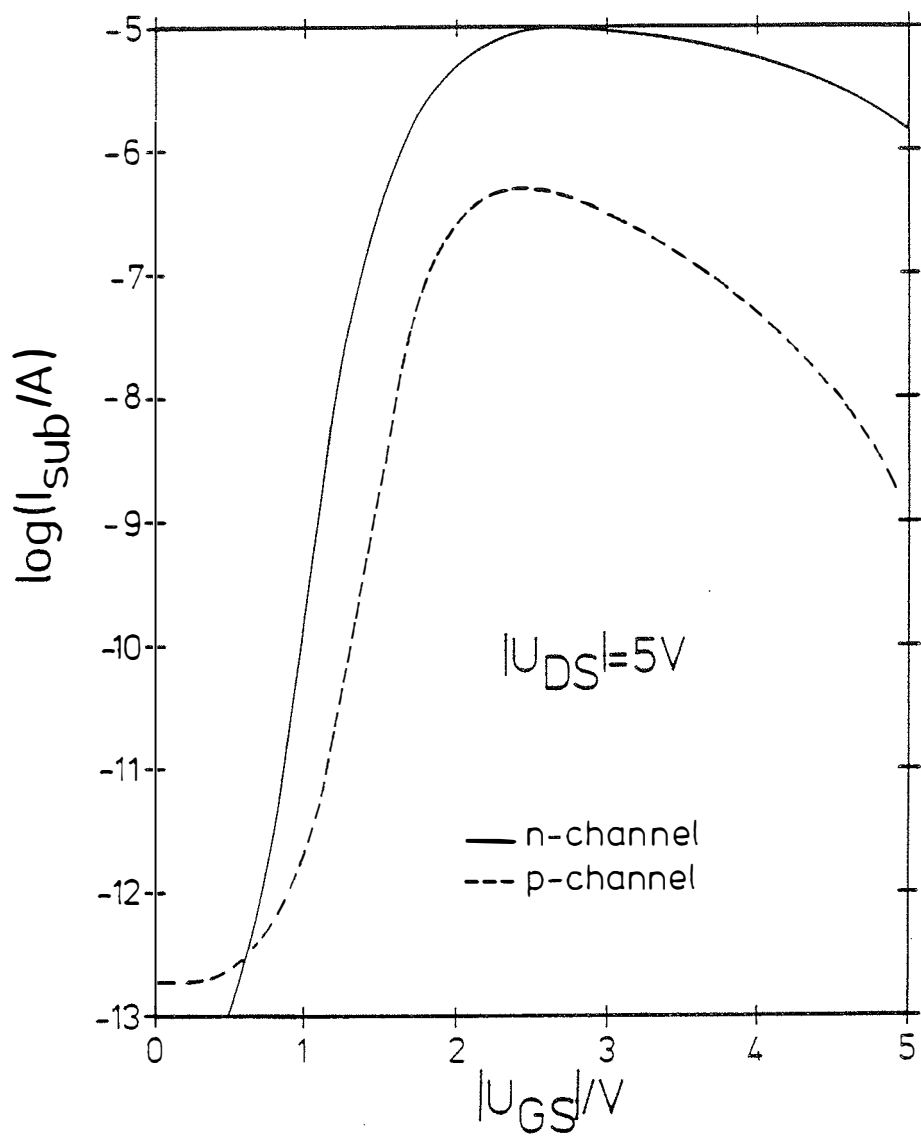
Fig. 4.5-8: Substrate currents

Fig. 5.5-9 shows substrate currents for a fixed drain voltage ($U_{DSn}=U_{SDp}=5V$). These operating points are not part of the transfer characteristic and cannot occur in the steady state case. However, they may occur during a logic sweep. As holes exhibit less ionization activity than electrons, substrate currents in the p-MOST are smaller by at least one decade compared to the substrate currents in the n-MOST.

A CMOS Inverter
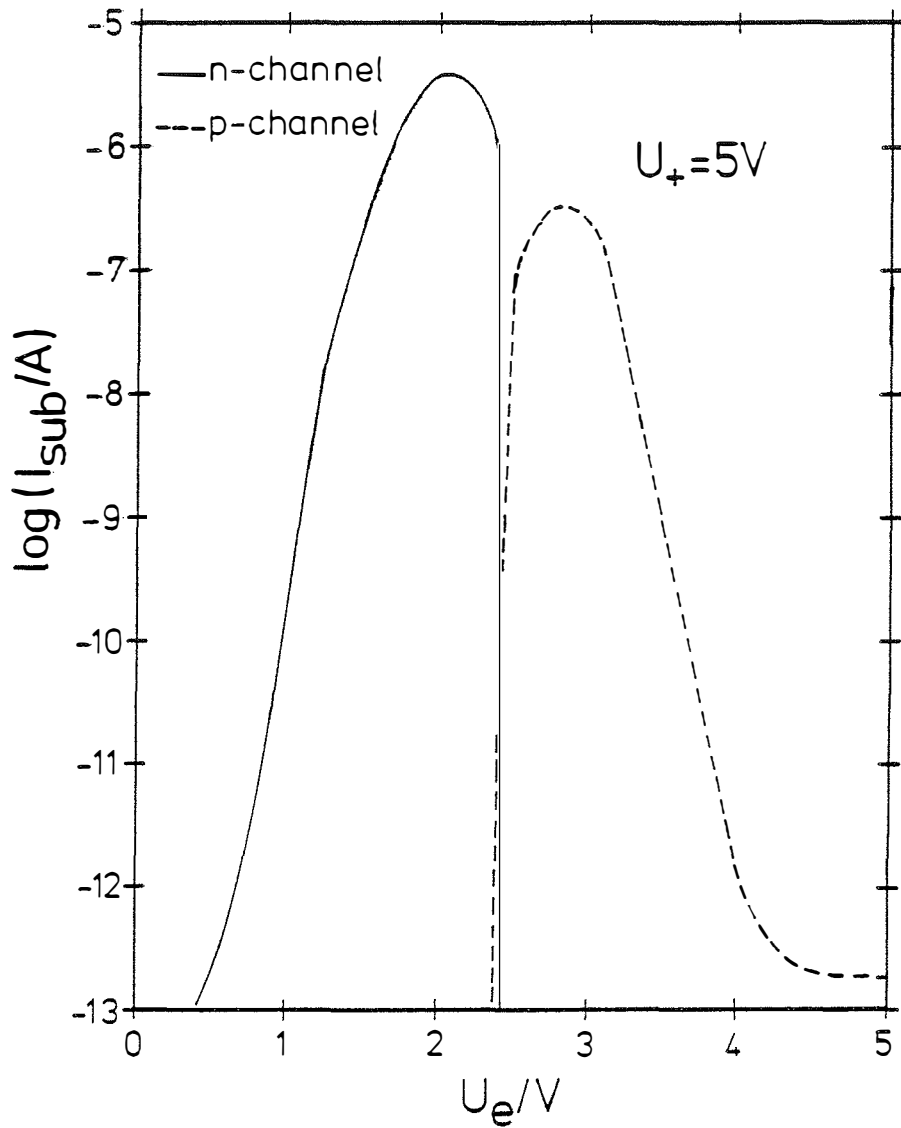
Fig. 4.5-9: Substrate currents for fixed drain voltage

Substrate current is important in CMOS technology as it initiates current injection form source into substrate. Thus the parasitic n^+pnp^+ thyristor (which consists of n-MOST source, p-well, n-substrate, and p-MOST source of a nearby inverter, cf. Fig. 4.5-1) may be triggered. The so called "latch-up" may result in the burn out of the entire circuit. Critical values of the substrate current for triggering the parasitic thyristor are about $10^{-5}A$-$10^{-4}A$ /125/. Contacts to the the p-well and the substrate result in improved latch-up performance /125/.

A CMOS Inverter

5. CONCLUSION

In this paper we tried to sketch the state of the art in modeling MOS transistors with numerical methods. The underlying physics has been discussed and the importance of increasingly sophisticated numerical methods has been briefly outlined. It became evident that only progress in basic semiconductor physics will lead to the development of models which are capable of simulating device behaviour more reliably and which will match the technological advances of the recent device miniaturisation. One highly important objective of a model, its ability to predict the performance of a new device prior to having built the actual device, can only be reached if the physical parameters of the basic equations are analyzed even more thoroughly. This possibly implies a complete re-evaluation of some commonly accepted assumptions and approximations and it also seems to be the only way to get rid of the enormous amount of fitting parameters and the heuristic formulae which just simulate more or less precisely some complex physical phenomena. The inherent physics has to be analyzed very carefully before one can begin to synthesize a model which is able to simulate reality better. The power of a numerical model to predict device behaviour has been demonstrated using our MOS-transistor simulation program MINIMOS.

However, still much effort in analysis and simulation will have to be spent to make device miniaturisation and integration keep pace with the speed of recent days.

ACKNOWLEDGMENT

REFERENCES

/1/   Adachi T.,   Yoshii A.,   Sudo T.,   "Two-Dimensional
       Semiconductor Analysis Using Finite-Element Method", IEEE
       Trans. Electron Devices, Vol.ED-26, pp.1026-1031, 1979.

/2/   Adler M.S., "A Method for Achieving and Choosing Variable
       Density Grids in Finite Difference Formulations and the
       Importance of Degeneracy and Band Gap Narrowing in Device
       Modeling", Proc. NASECODE I Conf., pp.3-30, 1979.

/3/   Adler M.S., "A Method for Terminating Mesh Lines in Finite
       Difference Formulations of the Semiconductor device
       Equations", Solid-State Electron., Vol.23, pp.845-853,
       1980.

/4/   Agajanian A.H., "A Bibliography on Semiconductor Device
       Modeling", Solid-State Electron., Vol.18, pp.917-929, 1965.

/5/   Anderson C.L., Crowell C.R., "Threshold Energies for
       Electron-Hole Pair Generation by Impact Ionization in
       Semiconductors", Physical Review, Vol.B5, pp.2267-2272,
       1972.

/6/   Antoniadis D.A., Hansen S., Dutton R.W., "SUPREM II - a
       Program for IC Process Modeling and Simulation", Stanford
       Technical Report No.5019-2, 1978.

/7/   Arora N.D., Hauser J.R., Roulston D.J., "Electron and Hole
       Mobilities in Silicon as a Function of Concentration and
       Temperature", IEEE Trans. Electron Devices, Vol.ED-29,
       pp.292-295, 1982.

/8/   Baccarani G., Mazzone A.M., "On the Diffusion Current in
       Heavily Doped Silicon", Solid-State Electron., Vol.18,
       pp.469-470, 1975.

/9/   Bank R.E., Rose D.J., "Parameter Selection for Newton-Like
       Methods Applicable to Nonlinear Partial Differential
       Equations.", SIAM J.Numer.Anal., Vol.17, pp.806-822, 1980.

/10/   Baraff G.A., "Distribution Functions and Ionization Rates
       for Hot Electrons in Semiconductors", Physical Review,
       Vol.128, pp.2507-2517, 1962.

/11/   Baraff G.A., "Maximum Anisotropy Approximation for
       Calculating Electron Distributions; Application to High
       Field Transport in Semiconductors", Physical Review,
       Vol.133, pp.A26-33, 1964.

/12/   Barnes J.J., Shimohigashi K., Dutton R.W., "Short-Channel
       MOSFET's in the Punchthrough Current Mode", IEEE Trans.
       Electron Devices, Vol.ED-26, pp.446-453, 1979.

/13/   Barnes J.J., Lomax R.J., "Transient 2-Dimensional
       Simulation of a Submicrometre Gate-Length M.E.S.F.E.T.",
       Solid-State Communications, Vol.13, pp.93-95, 1973.

/14/   Bell D.A., "Improved Formulation of Gummel's Algorithm for
       solving the 2-Dimensional Current-flow Equations in
       Semiconductor Devices", Electron. Lett., Vol.8, No.22,
       pp.536-538, 1972.

/15/ Benedek P., Silvester P., "Capacitance of Parallel Rectangular Plates Separated by a Dielectric Sheet", IEEE Trans. Microwave Theory and Techniques, Vol.MTT-20, pp. 504-510, 1972.

/16/ Birkhoff G., "The Numerical Solution of Elliptic Equations", SIAM, Philadelphia, 1971.

/17/ Blakey P.A., "Comments on "Effects of Carrier Diffusion on the Small-Signal Behavior of IMPATT Diodes"", IEEE Trans. Electron Devices, Vol.ED-27, p.299, 1980.

/18/ Blatt F.J., "Physics of Electronic Conduction in Solids", McGraw-Hill, New York, 1968.

/19/ Brown G.W., Lindsay B.W., "The Numerical Solution of Poisson's Equation for Two-Dimensional Semiconductor Devices", Solid-State Electron., Vol.19, pp.991-992, 1976.

/20/ Buturla E.M., Cottrell P.E., Grossman B.M., Salsburg K.A., "Finite-Element Analysis of Semiconductor Devices: The FIELDAY Program", IBM J. Res. Dev., Vol.25, pp.218-231, 1981.

/21/ Buturla E.M., Cotrell P.E., "Simulation of Semiconductor Transport Using Coupled and Decoupled Solution Techniques", Solid-State Electron., Vol.23, pp.331-334, 1980.

/22/ Buturla E.M., Cotrell P.E., Grossman B.M., Salsburg K.A., Lawlor M.B., McMullen C.T. "Three-Dimensional Finite Element Simulation of Semiconductor Devices", Proc. International Solid-State Circuits Conf., pp.76-77, 1980.

/23/ Canali C., Majni G., Minder R., Ottaviani G., "Electron and Hole Drift Velocity Measurements in Silicon and Their Empirical Relation to Electric Field and Temperature", IEEE Trans. Electron Devices, Vol.ED-22, pp.1045-1047, 1975.

/24/ Caughey D.M., Thomas R.E., "Carrier Mobilities in Silicon Empirically Related to Doping and Field", Proc. IEEE, Vol.52, pp.2192-2193, 1967.

/25/ Chamberlain S.G., Husain A., "Three-Dimensional Simulation of VLSI MOSFET's", Proc. International Electron Devices Meeting, pp.592-595, 1981.

/26/ Chwang R., Chung-Whei Kao, Crowell C.R., "Normalized Theory of Impact Ionization and Velocity Saturation in Nonpolar Semiconductors via a Markov Chain Approach", Solid-State Electron., Vol.22, pp.599-620, 1979.

/27/ Chynoweth A.G., "Ionization Rates for Electrons and Holes in Silicon", Physical Review, Vol.109, pp.1537-1540, 1958.

/28/ Chynoweth A.G., "Uniform Silicon p-n Junctions. 2. Ionization Rates for Electrons", J. Appl. Phys., Vol.31, pp.1161-1165, 1960.

/29/ Coe D.J., Brockman H.E., Nicholas K.H., "A Comparison of Simple and Numerical Two-Dimensional Models for the Threshold Voltage of Short-Channel MOSTs", Solid-State Electron., Vol.20, pp.993-998, 1977.

References

/30/ Coen R.W., Muller R.S., "Velocity of Surface Carriers in Inversion Layers on Silicon", Solid-State Electron., Vol.23, pp.35-40, 1980.

/31/ Conradt R., "Auger-Rekombination in Halbleitern", in: Festkörperprobleme XII, pp.449-464, Vieweg, Braunschweig, 1972.

/32/ Cotrell P.E., Buturla E.M., "Two-Dimensional Static and Transient Simulation of Mobile Carrier Transport in a Semiconductor", Proc. NASECODE I Conf., pp.31-64, 1979.

/33/ Crowell C.R., Sze S.M., "Temperature Dependence of Avalanche Multiplication in Semiconductors", Appl. Phys. Lett., Vol.9, pp.242-244, 1966.

/34/ Curtis A.R., Reid J.K., "The Choice of Step Lengths When Using Differences to Approximate Jacobian Matrices", J. Inst. Maths Applics, Vol.13, pp.121-126, 1974.

/35/ D'Avanzo D.C., "Modeling and Characterization of Short-Channel Double Diffused MOS Transistors", Stanford Technical Report No. G-201-6, 1980.

/36/ Dalal V.L., "Avalanche Multiplication in Bulk n-Si", Appl. Phys. Lett., Vol.15, pp.379-381, 1969.

/37/ Dang L.M., Konaka M., "A Two-Dimensional Computer Analysis of Triode-Like Characteristics of Short-Channel MOSFET's", IEEE Trans. Electron Devices, Vol.ED-27, pp.1533-1539, 1980.

/38/ De La Moneda F.H., "Threshold Voltage from Numerical Solution of the Two-Dimensional MOS Transistor", IEEE Trans. Circuit Theory, Vol.CT-20, pp.666-673, 1973.

/39/ De Mari A., "An Accurate Numerical Steady-State One-Dimensional Solution of the P-N Junction", Solid-State Electron., Vol.11, pp.33-58, 1968.

/40/ De Mari A., "An Accurate Numerical One-Dimensional Solution of the P-N Junction under Arbitrary Transient Conditions", Solid-State Electron., Vol.11, pp.1021-2053, 1968.

/41/ Debye P.P., Conwell E.M., "Electrical Properties of N-Type Germanium", Physical Review, Vol.93, pp.693-706, 1954.

/42/ Demoulin E., Greenfield J.A., Dutton R.W., Chatterjee P.K., Tasch A.F., "Process Statistics of Submicron MOSFET's", Proc. International Electron Devices Meeting, pp.34-37, 1979.

/43/ Dennard R.H., Gaensslen F.H., Yu H.N., Rideout V.L., Bassous E., Le Blanc A.R., "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions", IEEE J. Solid-State Circuits, Vol.SC-9, pp.256-268, 1974.

/44/ Deuflhard P., "A Modified Newton Method for the Solution of Ill-Conditioned Systems of Nonlinear Equations with Application to Multiple Shooting", Numer. Math., Vol.22, pp.289-315, 1974.

References

/45/ Deuflhard P., Heindl G., "Affine Invariant Convergence Theorems for Newton's Method and Extensions to Related Methods", SIAM J.Numer.Anal., Vol.16, pp.1-10, 1979.

/46/ Dongarra J.J., Moler C.B., Bunch J.R., Stewart G.W., "LINPACK", SIAM, Philadelphia, 1979.

/47/ Dorkel J.M., Leturcq Ph., "Carrier Mobilities in Silicon Semi-Empirically Related to Temperature, Doping and Injection Level", Solid-State Electron., Vol.24, pp.821-825, 1981.

/48/ Duff I.S., "A Survey of Sparse Matrix Research", Proc. IEEE, Vol.65, pp.500-535, 1977.

/49/ Duff I.S., "Practical Comparison of Codes for the Solution of Sparse Linear Systems", A.E.R.E. Harwell, Oxfordshire, 1979.

/50/ Dupont T., Kendall R.D., Rachford H.H., "An Approximate Factorization Procedure for Solving Self-Adjoint Elliptic Difference Equations", SIAM, J. Num. Anal., Vol.5, pp.559-573, 1968.

/51/ Dutton R.W., Hansen S.E., "Process Modeling of Integrated Circuit Device Technology", Proc. IEEE, Vol.69, pp.1305-1320, 1981.

/52/ Dziewior J., Schmid W., "Auger Coefficients for highly doped and highly excited Silicon", Appl. Phys. Lett., Vol.31, pp.346-348, 1977.

/53/ Engl W.L., Manck O., Wieder A.W., "Device Modeling", in: Process and Device Modeling for Integrated Circuit Design, pp.3-17, Noordhoff, Leyden, 1977.

/54/ Engl W.L., Dirks H., "Functional Device Simulation by Merging Numerical Building Blocks", Proc. NASECODE II Conf., pp.34-62, 1981.

/55/ Engl W.L., Dirks H., "Models of Physical Parameters", in: Introduction to the Numerical Analysis of Semiconductor Devices and Integrated Circuits, pp.42-46, Boole Press, Dublin, 1981.

/56/ Engl W.L., Dirks H., "Numerical Device Simulation Guided by Physical Approaches", Proc. NASECODE I Conf., pp.65-93, 1979.

/57/ Fair R.B., Sun R.C., "Threshold-Voltage Instability in MOSFET's Due to Channel Hot-Hole Emission", IEEE Trans. Electron Devices, Vol.ED-28, pp.83-94, 1981.

/58/ Forsythe G.E., Wasow W.R., "Finite Difference Methods for Partial Differential Equations", Wiley, New York, 1960.

/59/ Fox L., "Finite-Difference Methods in Elliptic Boundary-Value Problems", in: The State of the Art in Numerical Analysis, pp.799-881, Academic Press, London, 1977.

/60/ Frey J., "Physics Problems in VLSI Devices", in: Introduction to the Numerical Analysis of Semiconductor Devices and Integrated Circuits, pp.47-50, Boole Press, Dublin, 1981.

References

/61/ Frey J., "Transport Physics for VLSI", in: Introduction to the Numerical Analysis of Semiconductor Devices and Integrated Circuits, pp.51-57, Boole Press, Dublin, 1981.

/62/ Gaensslen F.H., Jaeger R.C., "Temperature Dependent Threshold Behaviour of Depletion Mode MOSFET's", Solid-State Electron., Vol.22, pp.423-430, 1979.

/63/ Gansner M.,Ilegems M.,Schwob P.,Dutoit M., "Modelisation des Structures Microelectroniques de Petites Dimensions", Proc. Journees d'Electronique, pp.93-105, 1980.

/64/ Gibbons J., Johnson W.S., Mylroie S.W., "Projected Range Statistics", Halstead Press, Strandsberg, 1975.

/65/ Grant W.N., "Electron and Hole Ionization Rates in Epitaxial Silicon at High Electric Fields", Solid-State Electron., Vol.16., pp.1189-1203, 1973.

/66/ Greenfield J.A., Dutton R.W., "Nonplanar VLSI Device Analysis Using the Solution of Poisson's Equation", IEEE Trans. Electron Devices, Vol.ED-27, pp.1520-1532, 1980.

/67/ Greenfield J.A., Hansen S.E., Dutton R.W., "Two-Dimensional Analysis for Device Modeling", Stanford Technical Report No. G-201-7, 1980.

/68/ Grimes R.G., Kincaid D.R., Young D.R., "ITPACK 2A - A Fortran Implementation of Adaptive Accelerated Iterative Methods for Solving Large Sparse Linear Systems", University of Texas, Austin, Report. No. CNA-164, 1980.

/69/ Gummel H.K., "A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations", IEEE Trans. Electron Devices, Vol.ED-11, pp.455-465, 1964.

/70/ Hachtel G.D., Mack M.H., O'Brien R.R., "Semiconductor Analysis Using Finite Elements-Part 2: IGFET and BJT Case Studies", IBM J. Res. Dev., Vol.25, pp.246-260, 1981.

/71/ Hachtel G.D., Mack H.H., O'Brien R.R., Speelpennig B., "Semiconductor Analysis Using Finite Elements-Part 1: Computational Aspects", IBM J. Res. Dev., Vol.25, pp.232-245, 1981.

/72/ Hamilton D.J., Howard W.G., "Basic Integrated Circuit Engineering", McGraw-Hill Kogakusha, Tokyo, 1975.

/73/ Hart J.F., Cheney E.W., Lawson C.L., Maehly H.J., "Computer Approximations", Wiley, New York, 1968.

/74/ Haug A., "Strahlungslose Rekombination in Halbleitern (Theorie)", in: Festkörperprobleme XII, Vieweg, Braunschweig, 1972.

/75/ Hauser J.R., "Threshold Energy for Avalanche Multiplication in Semiconductors", J. Appl. Phys., Vol.37, pp.507-509, 1966.

/76/ Heimeier H.H., "A Two-Dimensional Numerical Analysis of a Silicon N-P-N Transistor", IEEE Trans. Electron Devices, Vol.ED-20, pp.708-714, 1973.

References

/77/ Hess K., "Ballistic Electron Transport in Semiconductors", IEEE Trans. Electron Devices, Vol.ED-28, pp.937-940, 1981.

/78/ Heywang W., Pötzl H.W., "Bandstruktur und Stromtransport", Springer, Berlin, 1976.

/79/ Jacoboni C., Canali C., Ottaviani G., Quaranta A.A., "A Review of Some Charge Transport Properties of Silicon", Solid-State Electron., vol.20, pp.77-89, 1977.

/80/ Jaggi R., Weibel H., "High-Field Electron Drift Velocities and Current Densities in Silicon", Helv.Phys.Acta., Vol.42, pp.631-632, 1969.

/81/ Johnson L.W., Riess R.D., "An Error Analysis for Numerical Differentiation", J. Inst. Maths Applics, Vol.11, pp.115-120, 1973.

/82/ Kahng D., Atalla M.M., "Silicon-Silicon Dioxide Field Induced Surface Devices", IRE-AIEE, Solid-State Device Res. Conf., 1960.

/83/ Kani K., "A Survey of Semiconductor Device Analysis in Japan", Proc. NASECODE I Conf., pp.104-119, 1979.

/84/ Kennedy D.P., Murley P.C., "Steady State Mathematical Theory for the Insulated Gate Field Effect Transistor", IBM J. Res. Dev., Vol.17, pp.2-12, 1973.

/85/ Kennedy D.P., O'Brien R.R., "Two-Dimensional Mathematical Analysis of a Planar Type Junction Field-Effect Transistor", IBM J. Res. Dev., Vol.13, pp.662-674, 1969.

/86/ Kotani N., Kawazu S., "A Numerical Analysis of Avalanche Breakdown in Short-Channel MOSFETS", Solid-State Electron., Vol.24, pp.681-687, 1981.

/87/ Kotani N., Kawazu S., "Computer Analysis of Punch-Through in MOSFET's", Solid-State Electron., Vol.22, pp.63-70, 1979.

/88/ Lee C.A., Logan R.A., Batdorf R.L., Kleimack J.J., Wiegmann W., "Ionization Rates of Holes and Electrons in Silicon", Physical Review, Vol.134, pp.A761-773, 1964.

/89/ Lee H., Sansbury J.D., Dutton R.W., Moll J.L., "Modeling and Measurement of Surface Impurity Profiles of Laterally Diffused Regions", IEEE J. Solid-State Circuits, Vol.SC-13, pp.455-461, 1978.

/90/ Lee H., "Two-Dimensional Impurity Diffusion Studies: Process Models and Test Structures for Low-Concentration Boron Diffusion", Stanford Technical Report No.G201-8, 1980.

/91/ Leguerre R., "Approximate Values of the Multiplication Coefficient in One-sided Abrupt Junctions", Solid-State Electron., Vol.19, pp.875-881, 1976.

/92/ Li S.S., Thurber W.R., "The Dopant Density and Temperature Dependence of Electron Mobility and Resistivity in n-Type Silicon", Solid-State Electron., Vol.20, pp.609-616, 1977.

References

/93/ Liu S., Hoefflinger B., Pederson D.O., "Interactive Two-Dimensional Design of Barrier Controlled MOS Transistors", IEEE Trans. Electron Devices, Vol.ED-27, pp.1550-1558, 1980.

/94/ Loeb H.W., Andrew R., Love W., "Application of 2-Dimensional Solutions of the Shockley-Poisson Equation to Inversion-Layer M.O.S.T. Devices", Electron. Lett., Vol.4, pp.352-354, 1968.

/95/ Longo H.E., "Ladungsträgermultiplikation bei MIS-Transistoren", Zeitschrift f. Angew. Physik, Vol.29, p.166, 1970.

/96/ Manck O., Engl W.L., "Two-Dimensional Computer Simulation for Switching a Bipolar Transistor out of Saturation", IEEE Trans. Electron Devices, Vol.ED-24, pp.339-347, 1975.

/97/ Marhsak A.H., Shrivastava R., "Law of the Junction for Degenerate Material with Position-Dependent Band Gap and Electron Affinity", Solid-State Electron., Vol.22, pp.567-571, 1979.

/98/ Marsal D., "Die Numerische Lösung partieller Differential-gleichungen", Bibliographisches Institut, Mannheim, 1976.

/99/ Masuda H., Nakai M., Kubo M., "Characteristics and Limitation of Scaled-Down MOSFET's due to Two-Dimensional Field-Effect", IEEE Trans. Electron Devices, Vol.ED-26, pp.980-986, 1979.

/100/ Matsumoto H., Sawada K., Asai S., Hirayama M., Nagasawa K., "Effect of Long Term Stress on IGFET Degradations Due to Hot Electron Trapping", Trans. Electron Devices, Vol.ED-28, pp.923-928, 1981.

/101/ McKay K.G., "Avalanche Breakdown in Silicon", Physical Review, Vol.94, pp.877-884, 1954.

/102/ McKay K.G., McAfee K.B., "Electron Multiplication in Silicon and Germanium", Physical Review, Vol.91, pp.1079-1084, 1953

/103/ Mertens R.P., Van Meerbergen J.L., Nijs J.F., Van Overstraeten R.J., "Measurement of the Minority-Carrier Transport Parameters in Heavily Doped Silicon", IEEE Trans. Electron Devices, Vol.ED-27, pp.949-955, 1980.

/104/ Meyer G.H., "On Solving Nonlinear Equations with a One-Parameter Operator Imbedding", SIAM J.Numer.Anal., Vol.5, pp.739-752, 1968.

/105/ Miller S.L., "Avalanche Breakdown in Germanium", Physical Review, Vol.99, pp.1234-1241, 1955.

/106/ Miller S.L., "Ionization Rates for Holes and Electrons in Silicon", Physical Review, Vol.105, pp.1246-1249, 1957.

/107/ Mock M.S., "A Time-Dependent Numerical Model of the Insulated-Gate Field-Effect Transistor", Solid-State Electron., Vol.24, pp.959-966, 1981.

References

/108/ Mock M.S., "A Two-Dimensional Mathematical Model of the Insulated-Gate Field-Effect Transistor", Solid-State Electron., Vol.16, pp.601-609, 1973.

/109/ Mock M.S., "Convergence and Accuracy in Stationary Numerical Models", in: Introduction to the Numerical Analysis of Semiconductor Devices and Integrated Circuits, pp.58-62, Boole Press, Dublin, 1981.

/110/ Moglestue C., "A Monte-Carlo Particle Model Study of the Influence of the Doping Profiles on the Characteristics of Field-Effect Transistors", Proc. NASESCODE II Conf., pp.244-249, 1981.

/111/ Moglestue C., Beard S.J., "A Particle Model Simulation of Field Effect Transistors", Proc. NASECODE I Conf., pp.232-236, 1979.

/112/ Moll J.L., Overstraeten R. van, "Charge Multiplication in Silicon p-n Junctions", Solid-State Electron., Vol.6, pp.147-157, 1963.

/113/ Moll J.J, Meyer N.I., "Secondary Multiplication in Silicon", Solid-State Electron., Vol.3, pp.155-158, 1961.

/114/ Müller W., Risch L., Schütz A., "Analysis of Short Channel MOS Transistors in the Avalanche Multiplication Regime" IEEE Trans. Electron Devices, to be published.

/115/ Nakagawa A., "One-Dimensional Device Model of the npn Bipolar Transistor Including Heavy Doping Effects under Fermi Statistics", Solid-State Electron., Vol.22, pp.943-949, 1979.

/116/ Nussbaum A., "Inconsistencies in the Original Form of the Fletcher Boundary Conditions", Solid-State Electron., Vol.21, pp.1178-1179, 1978.

/117/ Ochi S., Okabe T., Yoshida I., Yamaguchi K., Nagata K., "Computer Analysis of Breakdown Mechanism in Planar MOSFET's", IEEE Trans. Electron Devices, Vol.ED-27, pp.399-400, 1980.

/118/ Ogawa T., "Avalanche Breakdown and Multiplication in Silicon pin Junctions", Jap. J. Appl. Phys., Vol.4, pp.473-484, 1965.

/119/ Oh S.Y., Ward D.E., Dutton R.W., "Transient Analysis of MOS Transistors", IEEE Trans. Electron Devices, Vol.ED-27, pp.1571-1578, 1980.

/120/ Oka H., Nishiuchi K., Nakamura T., Ishikawa H., "Computer Analysis of a Short-Channel BC MOSFET", IEEE Trans. Electron Devices, Vol.ED-27, pp.1514-1520, 1980.

/121/ Oka H., Nishiuchi K., Nakamura T., Ishikawa H., "Two-Dimensional Numerical Analysis of Normally-Off Type Buried Channel MOSFET's", Proc. International Electron Devices Meeting, pp.30-33, 1979.

/122/ Okuto Y., Crowell C.R., "Energy Conservation Considerations in the Characterization of Impact Ionization in Semiconductors", Physical Review, Vol.B6, pp.3076-3081, 1972.

References

/123/ Okuto Y., Crowell C.R., "**Ionization Coefficients in Semiconductors: A Nonlocalized Property**", Physical Review, Vol.B10, pp.4284-4296, 1974.

/124/ Ortega J.M., Rheinboldt W.C., "**Iterative Solution of Nonlinear Equations in Several Variables**", Academic Press, New York, 1970.

/125/ Peyne R.S., Grant W.N., Bertram W.J., "**Elimination of Latch-Up in Bulk CMOS**", Proc. International Electron Devices Meeting, pp.248-251, 1980.

/126/ Rheinboldt W.C., "**Methods for Solving Systems of Nonlinear Equations**", SIAM, Philadelphia, 1974.

/127/ Rokus A., "**Zur numerischen Lösung linearer pentagonaler Gleichungssysteme hohen Ranges**", Diplomarbeit, Technische Universität Wien, 1980.

/128/ Runge H., "**Distribution of Implanted Ions under Arbitrarily Shaped Mask Edges**", Phys. Status Solidi, Vol.(a)39, pp.595-599, 1977.

/129/ Ryssel H., Ruge I., "**Ionenimplantation**", Teubner, Stuttgart, 1978.

/130/ Sabnis A.G., Clemens J.T., "**Characterization of the Electron Mobility in the inverted (100) SI-Surface**", Proc. International Electron Devices Meeting, pp.18-21, 1979.

/131/ Sayle W.E., Lauritzen P.E., "**Avalanche Ionization Rates Measured in Silicon and Germanium at Low Electric Fields**", IEEE Trans. Electron Devices, Vol.ED-18, pp.58-66, 1971.

/132/ Scharfetter D.L., Gummel H.K., "**Large-Signal Analysis of a Silicon Read Diode Oscillator**", IEEE Trans. Electron Devices, Vol.ED-16, pp.64-77, 1969.

/133/ Schütz A., Selberherr S., Pötzl H.W., "**A Two-Dimensional Model of the Avalanche Effect in MOS Transistors**", Solid-State Electron., Vol.25, pp.177-183, 1982.

/134/ Schütz A., Selberherr S., Pötzl H.W., "**Analysis of Breakdown Phenomena in MOSFET's**", IEEE Trans. Computer-Aided-Design of Integrated Circuits, Vol.CAD-1, pp.77-85 1982.

/135/ Schütz A., Selberherr S., Pötzl H.W., "**Numerical Analysis of Breakdown Phenomena in MOSFET's**", Proc. NASECODE II Conf., pp.270-274, 1981.

/136/ Seeger K., "**Semiconductor Physics**", Springer, Wien, 1973.

/137/ Selberherr S., Schütz A., Pötzl H., "**Investigation of Parameter Sensitivity of Short Channel MOSFETS**", Solid-State Electron., Vol.25, pp.85-90, 1982.

/138/ Selberherr S., Fichtner W., Pötzl H.W., "**MINIMOS - a Program Package to Facilitate MOS Device Design and Analysis**", Proc. NASECODE I Conf., pp.275-279, 1979.

/139/ Selberherr S., Schütz A., Pötzl H., "**MINIMOS - Zweidimensionale Modellierung von MOS-Transistoren**", Elektronikschau, Vol.9, pp.18-23, Vol.10, pp.54-58, 1980.

References

/140/ Selberherr S., Schütz A., Pötzl H.W., **"MINIMOS — a Two—Dimensional MOS Transistor Analyzer"**, IEEE Trans. Electron Devices, Vol.ED-27, pp.1540-1550, 1980.

/141/ Selberherr S., Guerrero E., **"Simple and Accurate Representation of Implantation Parameters by Low Order Polynomals"**, Solid-State Electron., Vol.24, pp.591-593, 1981.

/142/ Selberherr S., Schütz A., Pötzl H., **"Two—Dimensional MOS Transistor Modelling"**, European Electronics, Vol.1, pp.20-30, 1981.

/143/ Selberherr S., **"Zweidimensionale Modellierung von MOS—Transistoren"**, Dissertation, Technische Universität Wien, 1981.

/144/ Sherman A.H., **"On Newton—Iterative Methods for the Solution of Systems of Nonlinear Equations"**, SIAM J.Numer.Anal., Vol.15, pp.755-771, 1978.

/145/ Shockley W., **"Problems Related to p—n Junctions in Silicon"**, Solid-State Electron., Vol.2, pp.35-67, 1961.

/146/ Slotboom J.W., **"The pn—Product in Silicon"**, Solid-State Electron., Vol.20, pp.279-283, 1977.

/147/ Smith G.D., **"Numerical Solution of Partial Differential Equations: Finite Difference Methods"**, Clarendon Press, Oxford, 1978.

/148/ Smith R.A., **"Semiconductors"**, Cambridge University Press, Cambridge, 1978.

/149/ Spirito P., **"Avalanche Multiplication Factors in Ge and Si Abrupt Junctions"**, IEEE Trans. Electron Devices, Vol.ED-21, pp.226-231, 1974.

/150/ Stoer J., Bulirsch R., **"Einführung in die Numerische Mathematik II"**, Springer, Berlin, 1978.

/151/ Stone H.L., **"Iterative Solution of Implicit Approximations of Multidimensional Partial Differential Equations"**, SIAM J.Numer.Anal., Vol.5, pp.530-558, 1968.

/152/ Sun E., Moll J., Berger J., Alders B.,, **"Breakdown Mechanism in Short—Channel MOS Transistors"**, Proc. International Electron Devices Meeting, pp.478-482, 1978.

/153/ Sun S.C., Plummer J.D., **"Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces"**, IEEE Trans. Electron Devices, Vol.ED-27, pp.1497-1508, 1980.

/154/ Sutherland A.D., **"A Two—Dimensional Computer Model for the Steady—State Operation of MOSFET's"**, US. Army Electronics Command Res. and Dev. Techn. Rep., ECOM-75-1344-F, 1977.

/155/ Sutherland A.D., **"An Algorithm for Treating Interface Surface Charge in the Two—Dimensional Discretization of Poisson's Equation for the Numerical Analysis of Semiconductor Devices such as MOSFET's"**, Solid-State Electron., Vol.23, pp.1085-1087, 1980.

References

/156/ Sutherland A.D., "On the Use of Overrelaxation in Conjuncton with Gummel's Algorithm to Speed the Convergence in a Two-dimensional Computer Model for MOSFET's", IEEE Trans. Electron Devices, Vol.ED-27, pp.1297-1298, 1980.

/157/ Sze S.M., Gibbons G., "Avalanche Breakdown Volatages of Abrupt and Linearly Graded p-n Junctions in Ge, Si, GaAs, and GaP", Appl. Phys. Lett., Vol.8, pp.111-113, 1966.

/158/ Sze S.M., "Physics of Semiconductor Devices", Wiley, New York, 1969.

/159/ Takacs D.,Schwabe U.,Bürker U., "The Influence of Temperature on the Tolerances of MOS-Transistors in a One-Micrometer Technology", Proc. International Electron Devices Meeting, pp.569-573, 1980.

/160/ Tang J.Y., Shichijo H., Hess K., Iafrate G.J., "Band-Structure Dependent Impact Ionization in Silicon and Gallium Arsenide", Journal de Physique, pp.C7:63-69, Montpellier, 1981.

/161/ Temple V.A.K., Adler M.S., "Calculation of the Diffusion Curvature Related Avalanche Brekdown in High-Voltage Planar p-n Junctions", IEEE Trans. Electron Devices, Vol.ED-22, pp.910-916, 1975.

/162/ Thornber K.K., "Applications of Scaling to Problems in High-Field Electronic Transport", J. Appl. Phys., Vol.52, pp.279-290, 1981.

/163/ Thornber K.K., "Relation of Drift Velocity to Low-Field Mobility and High-Field Saturation Velocity", J. Appl. Phys., Vol.51, pp.2127-2136, 1980.

/164/ Tielert R., "Two-Dimensional Numerical Simulation of Impurity Redistribution in VLSI Processes", IEEE Trans. Electron Devices, Vol.ED-27, pp.1479-1483, 1980.

/165/ Toyabe T., Yamaguchi K., Asai S., Mock M., "A Numerical Model of Avalanche Breakdown in MOSFET's", IEEE Trans. Electron Devices, Vol.ED-25, pp.825-832, 1978.

/166/ Toyabe T., Asai S., "Analytical Model of Threshold Voltage and Breakdown Voltage of Short-Channel MOSFET's derived from Two-Dimensional Analysis", IEEE Trans. Electron Devices, Vol.ED-26, pp.453-461, 1979.

/167/ Troutman R.R., "Low-Level Avalanche Multiplication in IGFET's", IEEE Trans. Electron Devices, Vol.ED-23, pp.419-425, 1976.

/168/ Troutman R.R., "VLSI Limitations from Drain-Induced Barrier Lowering.", IEEE Trans. Electron Devices, Vol.ED-26, pp.461-469, 1979.

/169/ Van de Wiele F., Van Overstraeten R., De Man H., "Graphical Method for the Determination of Junction Parameters and of Multiplication Parameters", Solid-State Electron., Vol.13, pp.25-36, 1970.

References

/170/ Van Overstraeten R., De Man H., **"Measurement of the Ionization Rates in Diffused Silicon p-n Junctions"**, Solid-State Electron., Vol.13, pp.583-608, pp.1970.

/171/ Van Overstraeten R.J., De Man H.J., Mertens R.P., **"Transport Equations in Heavy Doped Silicon"**, IEEE Trans. Electron Devices, Vol.ED-20, pp.290-298, 1973.

/172/ Van Roosbroeck W.V., **"Theory of Flow of Electrons and Holes in Germanium and Other Semiconductors"**, Bell Syst. Techn. J., Vol.29, pp.560-607, 1950.

/173/ Van Vliet K.M., **"On Fletcher's Boundary Conditions"**, Solid-State Electron., Vol.22, pp.443-444, 1979.

/174/ Van Vliet K.M., **"The Shockley-Like Equations for the Carrier Densities and the Current Flows in Materials with a Nonuniform Composition"**, Solid-State Electron., Vol.23, pp.49-53, 1980.

/175/ Vandorpe D., Borel J., Merckel G., Saintot P. **"An Accurate Two-Dimensional Numerical Analysis of the MOS Transistor"**, Solid-State Electron., Vol.15, pp.547-557, 1972.

/176/ Vandorpe D., Xuong N.H., **"Mathematical 2-Dimensional Model of Semiconductor Devices"**, Electron. Lett., Vol.7, pp.47-50, 1971.

/177/ Varga R.S., **"Matrix Iterative Analysis"**, Prentice-Hall, Englewood Cliffs, 1962.

/178/ Wachspress E.L., **"Iterative Solution of Elliptic Systems"**, Prentice-Hall, Englewood Cliffs, 1966.

/179/ Wada T., Dang R.L.M., **"Modification of ICCG Method for Application to Semiconductor Device Simulators"**, Electron. Lett., Vol.18, pp.265-266, 1982.

/180/ Wolf H.F., **"Semiconductors"**, Wiley, New York, 1971.

/181/ Wolff P.A., **"Theory of Multiplication in Silicon and Germanium"**, Physical Review, Vol.95, pp.1415-1420, 1954.

/182/ Yamaguchi K., **"Field-Dependant Mobility Model for Two-Dimensional Numerical Analysis of MOSFET's"**, IEEE Trans. Electron Devices, Vol.ED-26, pp.1068-1074, 1979.

/183/ Yamaguchi K., Toyabe T., Kodera H., **"Two-Dimensional Analysis of Triode-Like Operation of Junction Gate FET's"**, IEEE Trans. Electron Devices, Vol.ED-22, pp.1047-1049, 1975.

/184/ Yokoyama K.Y., Yoshii A., Horiguchi S., **"Threshold Sensitivity Minimization of Short-Channel MOSFET's by Computer Simulation"**, IEEE Trans. Electron Devices, Vol.ED-27, pp.1509-1514, 1980.

/185/ Yoshii A., Horiguchi S., Sudo T., **"A Numerical Analysis for Very Small Semiconductor Devices"**, Proc. International Solid-State Circuits Conf., pp.80-81, 1980.

/186/ Yoshii A., Kitazawa H., Tomizawa M., Horiguchi S., Sudo T., **"A Three-Dimensional Analysis of Semiconductor Devices"**, IEEE Trans. Electron Devices, Vol.ED-29, pp.184-189, 1982.

References