On the modelling of downscaled bipolar transistors

Document status and date:
Published: 01/01/1990

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

On the Modelling of
Downscaled Bipolar Transistors

G.A.M. Hurkx
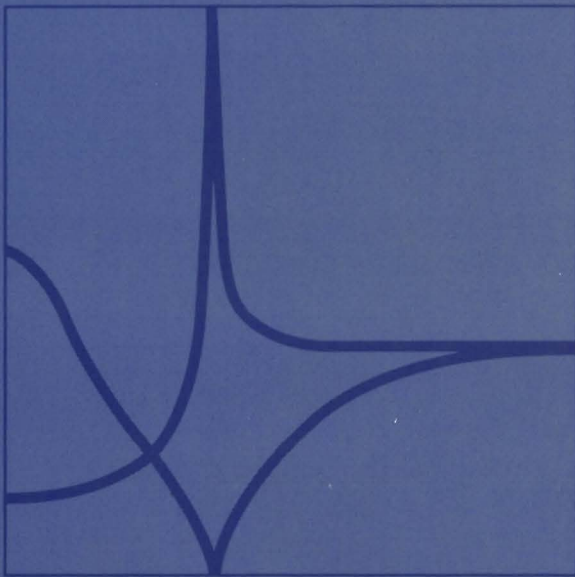
On the Modelling of
Downscaled Bipolar Transistors

On the Modelling of Downscaled Bipolar Transistors

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Eindhoven,
op gezag van de Rector Magnificus,
Prof. ir. M. Tels,
voor een commissie aangewezen
door het College van Dekanen
in het openbaar te verdedigen op
dinsdag 27 november 1990 te 13.30 uur

door

GODEFRIDUS ADRIANUS MARIA HURKX

geboren te Best

Dit proefschrift is goedgekeurd
door de promotoren

prof. dr. F.M. Klaassen
en
prof. dr. R.P. Mertens (KU Leuven)

CONTENTS

Chapter 1

INTRODUCTION

1.1 DOWNSCALING OF BIPOLAR TRANSISTORS

In its evolution from the point-contact transistor, invented by Bardeen and Brittain in 1947 [1.1], to a modern IC building-block the bipolar transistor has been subjected to a wide variety of technological innovations. Not only has the semiconductor material changed from germanium to silicon and GaAs, but innovations in the field of doping techniques, lithography and etching techniques have also had their impact on transistor development. In addition, the improvement in our understanding of device physics played a role in this device development. Perhaps the most striking feature in the development of bipolar transistors is the reduction of the dimensions (downscaling) of the active regions from the sub-millimetre range to the sub-micron range.

The major driving forces behind downscaling are speed improvement, reduction of power consumption and improvement of packing density. While the latter quantities are improved by lateral downscaling, reduction of the vertical dimensions (viz. junction depths) of the transistor is especially important for the improvement of the speed capabilities of the transistor. The speed of the transistor is usually characterized by the bias-dependent, common-emitter cut-off frequency f_T which is, roughly speaking, the frequency at which the current gain of the device has decreased to unity. This cut-off frequency increases with decreasing vertical dimensions of the active regions. This is due to a reduction of the stored minority charge without a corresponding decrease in the collector current. Although all active transistor regions may significantly contribute to the charge storage, the thickness and doping profile of the base region are particularly important for the speed capabilities of the device [1.2]. In addition to a reduction of junction depths a corresponding increase in doping density is required in order to maintain proper electrical characteristics with downscaling.
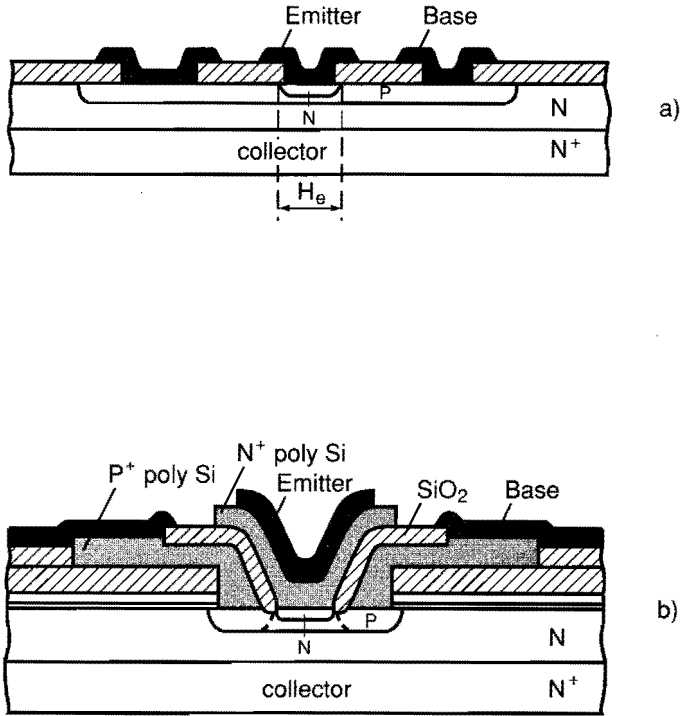
Fig. 1.1. Schematic cross-section of a conventional vertical n-p-n transistor (a) and an advanced self-aligned transistor (b).

Due to the corresponding increase of the electric field at the junction, however, tunnelling effects may adversely affect the reverse and low-forward characteristics of the emitter-base diode. Therefore, in general a trade-off between various electrical parameters appears necessary [1.3].

At present, the state-of-the-art in commercially available vertical n-p-n transistors is a planar structure as sketched in fig. 1.1a. Such a device is at present widely used as an elementary building-block in integrated circuits. The emitter-base and collector-base junction depths are in the sub-micron range, while the lateral dimensions of such a de-

vice are in the order of microns. These lateral dimensions, such as the separation between the base contact and the emitter, are limited by the minimum lithographic rules.

The base current consists of two components, viz. the ideal current and the non-ideal current. The most important component is the ideal current which dominates at medium and high forward base-emitter junction voltages. This component, which is caused by the recombination of electrons and holes in the quasi-neutral regions, is called ideal because it is proportional to exp (qV_{be}/kT), where V_{be} is the junction voltage, T is the absolute temperature, q is the elementary charge and k is Boltzmann's constant. The non-ideal current is proportional to exp (qV_{be}/mkT), with $m \simeq 2$. This component, which dominates at low bias, is caused by recombination in the emitter-base depletion layer. The ideal current in a device as sketched in fig. 1.1a is caused mainly by hole injection from the base into the emitter. These holes recombine either in the bulk or at the contact. For shallow emitters where the recombination length is much larger than the emitter depth the majority of the holes recombine at the contact. Such an emitter is called "transparent". Bulk recombination in the base can usually be neglected.

Since the lateral dimensions of the above transistors are much larger than the vertical dimensions, the hole flow in the emitter will be mainly vertical and injection through the emitter sidewall can be neglected. The electron injection from the emitter into the base will also be mainly vertical. This implies that sidewall effects directly related to minority-carrier injection will be fairly small in these devices.

Recent developments in silicon bipolar transistor technology have also made possible a significant lateral downscaling of bipolar devices [1.4 − 1.6]. Lateral downscaling has several important benefits. First of all, due to the reduction of the junction areas the parasitic junction capacitances are reduced, which increases the cut-off frequency. Because the base current mainly flows laterally in the base region, the base resistance also decreases when the emitter width H_e (see fig. 1.1) and the distance between the emitter and the base contact are reduced. Another very important motive for lateral downscaling is that for VLSI applications it is not only the speed capabilities which are important, but also its power dissipation. Since the power dissipation is proportional

to the collector current, at a given bias voltage the dissipation scales roughly with the emitter area.

The key step towards lateral downscaling has been the introduction of a technique by which the active transistor areas are formed utilizing only one lithographic step. All the other steps in the formation of the active regions are self-aligned to this step. Such a self-aligned scheme, together with the use of sophisticated dry-etching techniques, makes it possible to achieve very compact structures. In addition, the emitter stripe width has decreased due to the improvement of lithographic techniques (e.g. Electron-beam Lithography [1.7,1.8]). In fig. 1.1b such a self-aligned transistor is sketched.

In these self-aligned devices sidewall effects are expected to influence the device characteristics significantly for two main reasons. The first reason is the simple fact that due to lateral downscaling the two-dimensional minority-carrier injection near the emitter periphery plays a relatively large role. The second reason is related to the self-alignment of the process. The proximity of the base contact to the emitter can give rise to additional recombination at this contact [1.9]. In addition, the high base doping concentration at the emitter sidewall can influence the device characteristics.

Next we briefly discuss various new self-aligned processes. These can be divided roughly into three types (see fig. 1.2):
- Double-poly type [1.4]
- Sidewall-contacted base configuration [1.6]
- Single-poly type [1.10,1.11]

All three types are self-aligned, but the major differences lie in the method of contacting the base and, therefore, in the possibilities of reducing the base resistance and the dimensions of the parasitic regions. In both the double-poly structure (fig. 1.2a) and the sidewall-contacted base structure (fig. 1.2b) two deposited polysilicon layers are used to contact the base and the emitter. The extrinsic-base dopant and the emitter dopant are diffused from polysilicon layers, while the intrinsic base is usually formed by ion implantation. In the single-poly concept (fig. 1.2c) only the emitter is diffused from poly. The base is contacted by silicidation of the extrinsic base region. In this way a low extrinsic-base resistance can be obtained. The most sophisticated class of
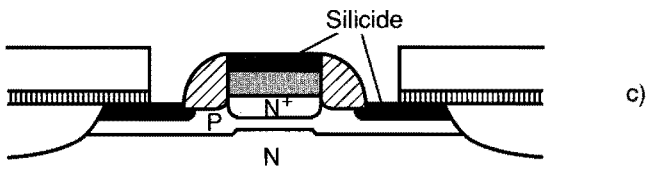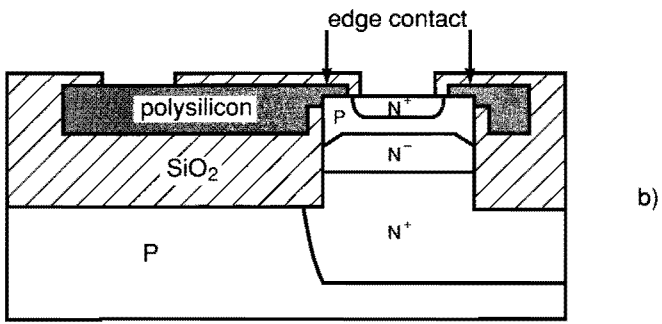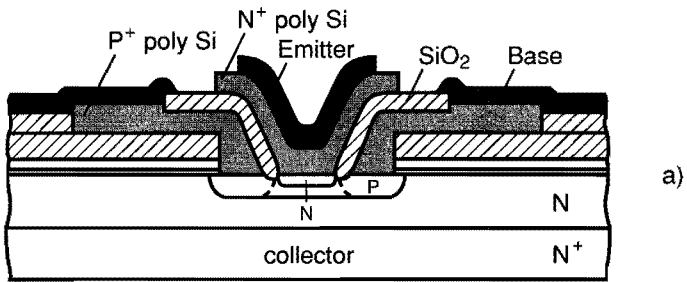
Fig. 1.2. Three types of advanced bipolar processes: (a) double-poly, (b) side-wall-contacted base and (c) single-poly.

devices is the sidewall-contacted base structure (fig. 1.2b). The base resistance of this device is further minimized by lateral diffusion of boron from the polysilicon layers located above the buried oxide. As can be observed from fig. 1.2b, the parasitic regions in such a mesa structure are minimized and the active part of the transistor closely resembles a one-dimensional device.

One of the key technological issues of these self-aligned processes is the control of the lateral diffusion of the extrinsic-base dopant to optimize the link-up between intrinsic base and extrinsic base [1.12]. This control is needed to achieve a low extrinsic-base resistance without introducing peripheral effects such as tunnelling and current-gain reduction due to the increased base doping concentration near the emitter edge. Other important technological issues are the properties of the polysilicon- monosilicon interface in the emitter and the control of very shallow vertical profiles [1.13].

1.2 PHYSICAL MODELLING PROBLEMS RELATED TO DOWNSCALING

Downscaling of bipolar transistors has given rise to several problems in the description of the electrical behaviour of such devices.. These problems can be divided into two categories. One category comprises problems related to numerical device simulations, while the second category comprises compact modelling problems.

In numerical simulations, Poisson's equation

$$\underline{\nabla}^2 \psi = \frac{-q}{\varepsilon_0 \, \varepsilon_r} \, (N_D^+ - N_A^- + p - n), \qquad (1.1)$$

and the continuity equations for electrons (n) and holes (p)

$$\frac{\partial n}{\partial t} = \frac{1}{q} \, \underline{\nabla}.\underline{J}_n - R, \qquad (1.2a)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \, \underline{\nabla}.\underline{J}_p - R, \qquad (1.2b)$$

are solved for a given set of boundary conditions. In (1.1), ψ is the electrostatic potential and N_A^- and N_D^+ are the ionized acceptor and donor concentrations, respectively. ε_0 is the permittivity of vacuum and ε_r is the static dielectric constant of the semiconductor. In (1.2) $\underline{J}_p$ and $\underline{J}_n$ are the hole and electron current densities, while R is the net electron-hole recombination rate. The current densities are given by

$$\underline{J}_n = q\, n\, \mu_n \left[\underline{F} - \frac{kT}{q} \underline{\nabla}(\ln n_{ie}) \right] + q\, D_n \underline{\nabla} n, \qquad (1.3a)$$

$$\underline{J}_p = q\, p\, \mu_p \left[\underline{F} + \frac{kT}{q} \underline{\nabla}(\ln n_{ie}) \right] - q\, D_p \underline{\nabla} p, \qquad (1.3b)$$

where μ_n and μ_p are the mobilities and D_n and D_p are the diffusion constants. $\underline{F}$ ($\underline{F} = -\underline{\nabla}\psi$) is the electric field and n_{ie} is the intrinsic carrier concentration. The current densities consist of a drift term which is proportional to the carrier concentration and a diffusion term which is proportional to the gradient of the carrier concentration. The second term on the right-hand side of (1.3) is the part of the drift term which takes into account the spatial variation of the bandgap due to an inhomogeneous dopant distribution [1.14]. The numerical solution of these equations, for instance for a two-dimensional problem as sketched in fig. 1.3, consists of three steps. Firstly, the space domain is mapped onto a grid of discrete points. Secondly, after application of some discretization scheme (finite differences or finite elements) the problem is transformed into a set of non-linear algebraic equations. Thirdly, this set of equations is solved to yield the unknown quantities ψ,n and p at the discrete points. Obviously, reliable results, even qualitatively, can only be obtained if all the relevant physical phenomena are properly incorporated into the equations.

With downscaling of bipolar devices several new physical effects have come into play. For instance, because of the increased electric field at the junctions (especially at the emitter-base junction) tunnelling effects have become important and must be taken into account in the net recombination rate R [1.15 − 1.17]. Hot-carrier behaviour is also important and this is often modelled by adding two energy-balance
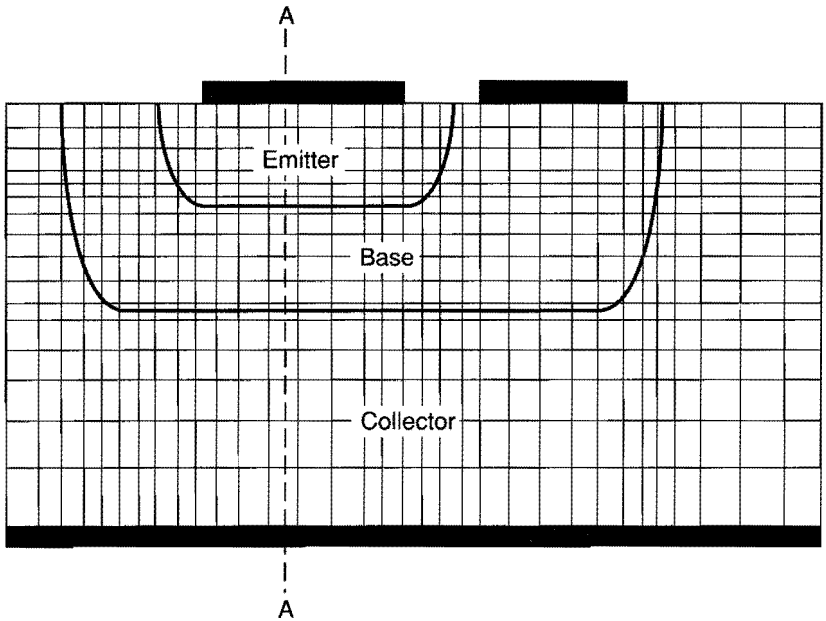
Fig. 1.3. An example of a space domain for the numerical simulation of a vertical n-p-n transistor.

equations to the above set of equations to obtain the carrier temperature or by the incorporation of a Monte-Carlo type of calculation into the simulations [1.18]. Furthermore, the description of the minority-carrier mobility and bandgap narrowing are important issues for the modelling of downscaled devices [1.19]. Finally, interface properties and the translation of these properties into appropriate boundary conditions are also important. For instance, due to the presence of interface states the recombination rate at an interface usually deviates from that in the bulk. In addition, the carrier mobility close to the interface may be different from the bulk value. The interfaces between monosilicon and polysilicon or a silicide in the emitter and base regions can have a strong influence on the device behaviour [1.9]. Also the inter-

Chapter 1

faces between monosilicon and insulating materials in the active transistor regions can strongly influence device behaviour [1.20,1.24].

Compact modelling is the description of the electrical transistor behaviour by means of a set of analytical expressions. Such a compact model is used to describe the transistor behaviour in numerical simulations of electrical circuits. The compact model expressions are based on an approximate regional solution of the basic semiconductor equations (1.1) and (1.2). Often the continuity equation for the minority carriers is solved in a certain quasi-neutral region (i.e. a region outside the depletion layers) of the device. The required additional boundary conditions are then obtained from a relation between the minority-carrier density at the edge of the depletion layer and the applied junction voltage (see for instance [1.21]). Obviously, this approach requires an in-depth understanding of the device behaviour. An excellent review of the state-of-the-art in compact modelling can be found in [1.21].

Downscaling-related problems in compact modelling can be divided into problems that are purely related to geometrical effects, i.e. related only to the reduction of the dimensions, both vertical and lateral, and problems related to the incorporation of distinct physical phenomena into compact models. An important example of compact modelling problems purely related to the reduction of the transistor dimensions is given by sidewall effects on the ideal base and collector currents (both d.c. and a.c.) in bipolar transistors. Existing models for these currents are usually based on a one-dimensional (vertical) analysis of the transistor. For instance, along cross-section A-A in fig. 1.3 the minority-carrier injections into both the emitter and base regions are expected to be nearly vertical and a one-dimensional analysis will be sufficient there. However, near the emitter edge the minority-carrier injection will definitely not be vertical and will give rise to sidewall effects on both d.c. and a.c. currents [1.22 − 1.25]. The new self-aligned technologies, especially the proximity of the base contact to the emitter, can also give rise to additional recombination current near the emitter sidewall [1.9,1.26]. Until now, these sidewall effects have been treated merely as lower-order correction terms. However, these effects, which do not scale with the emitter area, are expected to become increasingly important when the emitter area is reduced. Another example of com-

pact modelling problems purely related to geometrical effects is the description of transit times in various regions of the transistor.

Compact modelling problems related to distinct physical phenomena are, for instance, the modelling of avalanche- and tunnelling currents in p-n junctions [1.15 – 1.17,1.27]. These effects have become important because of their strong dependence on the electric field around the junction which increases with increasing doping density. This is particularly the case near the emitter sidewall where the doping density and, hence, the electric field can be very high due to the lateral diffusion of the extrinsic-base dopant.

1.3 A SHORT OVERVIEW OF THE WORK DESCRIBED IN THIS THESIS

This thesis deals with various aspects related to the modelling of currents and charges in downscaled silicon vertical n-p-n transistors.

Chapter two describes the sidewall effects on the ideal d.c. base and collector currents, i.e. geometrical effects on the conventional ideal currents. In the classification as given above this is therefore a compact modelling problem related to geometrical effects. These ideal currents are very important for the transistor behaviour and determine, for instance, the maximum attainable current gain of the device. The purpose of the work described in Chapter two is to reveal the relations between the sidewall effects on the ideal d.c. currents and the main process parameters (such as junction depths, emitter-base contact spacing, contact materials) and to present these relationships in the form of practical expressions which are consistent with 2D numerical simulations. These expressions, which are based on analytical calculations on a simplified structure, may serve as a starting point for a compact model description of these effects. To this end the expressions describing the sidewall effects are formulated using an effective perimeter area so that the relative importance of the corresponding effect can immediately be seen from a comparison of this area with the emitter bottom area. The implications of these sidewall effects on the variation of the current gain with the lateral emitter geometry are also discussed and a comparison is made between the experimentally observed de-

pendence of the current gain on the emitter dimensions and that obtained from the calculations.

Chapter three discusses the modelling of a physical phenomenon which becomes important with increasing doping density around a p-n junction, viz. tunnelling in forward- and reverse-biased junctions. In reverse-biased junctions this phenomenon is usually referred to as band-to-band tunnelling or Zener tunnelling [1.28]. In forward-biased junctions it is known as trap-assisted tunnelling, or as the so-called excess current, because this current is in excess of the normal diode current [1.28]. These effects are important at the emitter-base junction where the electric field can reach values as high as $10^8 V/m$. Band-to-band tunnelling determines the emitter-base leakage current, while the excess current contributes to the non-ideal base current and, therefore, adversely affects the current gain at low and medium bias conditions. Because of the high base dopant concentration at the emitter periphery, tunnelling can cause strong sidewall effects on the reverse and non-ideal forward base current. This particularly applies in modern self-aligned structures, such as those sketched in fig. 1.2, where the proximity of the base contact to the edge of the emitter-base junction may result in a very high base dopant concentration at the emitter periphery. These tunnelling effects may well prove to be one of the ultimate limitations to the downscaling of bipolar devices [1.15]. The primary goal of the work described in this chapter is to obtain model descriptions of these tunnelling effects which are suitable for incorporation into a numerical device simulator and which can provide a basis for a compact model description. This chapter also gives the physical basis for a compact model describing the full I-V characteristics of reverse-biased junctions. Besides tunnelling, this model includes avalanche generation and Shockley-Read-Hall generation. The implications of tunnelling effects on device behaviour, such as sidewall effects on the non-ideal base current, have not yet been investigated but will be the subject of future research.

Chapter four deals with the modelling of the a.c. transistor properties. The work described in this chapter can also be classified as a compact modelling problem related to geometrical effects. This chapter deals mainly with the determination of a.c. quantities (such as the a.c.

base and collector currents and the cut-off frequency) on the basis of a quasi-static analysis. In both compact modelling and numerical device simulations this approach is frequently used to obtain low-frequency a.c. parameters from a d.c. analysis [1.29 − 1.31]. Most treatments of the foundations of this quasi-static approach date from more than a decade ago when the dimensions of the devices were such that the maximum cut-off frequency was almost entirely determined by the base region. As a consequence, most of these treatments are confined to the base region. Nowadays, however, because of downscaling not only the base region but also other regions can contribute significantly to the a.c. properties of the device. In view of this downscaling the foundations of the quasi-static approach are reviewed in this chapter and the relating expressions for the small-signal a.c. base and collector currents are derived from the time-dependent continuity equations. The validity of the approximations involved is also discussed. It will be shown that in the case where the emitter transit time plays a significant role in the cut-off frequency, the approximations on which this approach is based are not fully justified. Section 4.3 gives a mathematically more rigorous low-frequency description of a.c. currents, based on a perturbation analysis of the time-dependent continuity equations. Unlike the conventional quasi-static approach, this method also yields internal a.c. quantities such as minority-carrier densities and current densities. This offers the possibility to add to the usual postprocessing routine for the calculation of the cut-off frequency from a numerical d.c. device simulation, a routine which calculates the phase shift of the collector current in the base region. This so-called excess phase shift can be an important quantity in high-frequency applications. Finally, the use of the quasi-static cut-off frequency as a figure of merit to characterize the high-frequency properties of advanced devices is discussed. Results from quasi-static calculations are also compared with those obtained from numerical small-signal a.c. solutions of the continuity equations.

In chapter five some final remarks are made and an outlook for future work is presented.

REFERENCES

1.1 J. Bardeen and W.H. Brittain, Phys. Rev., 74, p. 230 (1948)

1.2 J.J.H. van den Biesen, Solid-St. Electron., 29, p. 529 (1986)

1.3 D.D. Tang, G.P. Li, C.T. Chuang and T.H. Ning, Proc. IEDM '86, San Francisco, p. 412 (1986)

1.4 T. Sakai, Y. Kobayashi, H. Yamauchi, M. Sato and T. Makino, Jap. J. Appl. Phys., 20, supp. 20-1, p. 155 (1980)

1.5 H. Goto, Proc. ESSDERC '88, Montpellier, p. 471 (1988)

1.6 T. Nakamura, K. Nakazato, T. Miyazaki, T. Okabe and M. Nagata, IEEE J. Solid-State Circuits, SC-20, p. 168 (1985)

1.7 E. Greeneich, D.L. Tolliver and A. Gonzalez, IEEE Trans. Electron Devices, ED-28, p. 1346 (1981)

1.8 Y. Tamaki, F. Murai, Y. Kawamoto, K. Uehara, A. Hayasaki and A. Anzai, IEEE Electron Device Lett., EDL-7, p. 425 (1986)

1.9 M.H. El-Diwany, M.P. Brassington and P. Tuntasood, IEEE Electron Device Lett., EDL-9, p. 247 (1988)

1.10 J.L. de Jong, R.H. Lane, J.G. de Groot, and G.W. Connor, Proc. BCTM'88, Minneapolis, p. 202 (1988)

1.11 T. Hirao, T. Ikeda and Y. Kuramisu, Proc. ESSDERC'87, Bologna, p. 373 (1987)

1.12 D.D. Tang, T-C Chen, C-T Chuang, G.P. Li, J.M.C. Stork, M.B. Ketchen and E. Hackbarth, IEEE Electron Device Lett., EDL-8, p. 174 (1987)

1.13 H. Takemura, T. Kamiya, S. Ohi, M. Sugiyama, T. Tashiro and M. Nakamae, Proc. IEDM'86, San Francisco, p. 424 (1986)

1.14 A.H. Marshak and C.M. van Vliet, Proc. IEEE, 72, p. 148 (1984)

1.15 J.A. Del Alamo and R.M. Swanson, 18th Conf. on Solid State Devices and Materials, Tokyo, p. 283 (1986); IEEE Electron Device Lett., EDL-7, p. 629 (1986)

1.16 G.P. Li, E. Hackbarth and T.C. Chen, IEEE Trans. Electron Devices, ED-35, p. 89 (1988)

1.17 H. Schaber, J. Bieger , B. Benna and T. Meister, Proc. ESSDERC '87, Bologna, p. 365 (1987)

1.18 J.M. Higman, K. Hess, C.G. Hwang and R.W. Dutton, IEEE Trans. Electron Devices, ED-36, p. 930 (1989)

1.19 H.C. de Graaff and G.A.M. Hurkx, Proc. ESSDERC '88, Montpellier, p. 707 (1988)

1.20 M.W. Hillen and J. Holsbrink, Solid-St. Electron., 26, p. 453 (1983)

1.21 H.C. de Graaff and F.M. Klaassen, Compact Transistor Modelling for Circuit Design, Wien: Springer (1990)

1.22 H. M. Rein, Solid-St. Electron., 27, p. 625 (1984)

1.23 D. P. Verret and J. E. Brighton, IEEE Trans. Electron Devices, ED-34, p. 2297 (1987)

1.24 N. Shiono, Jap. J. Appl. Phys., 18, p. 1097 (1979)

1.25 T. Hamasaki, T. Wada, N. Shigyo and M. Yoshimi, IEEE Trans. Electron Devices, ED-35, p. 1620 (1988)

1.26 S. Sawada, Proc. BCTM'88, Minneapolis, p. 206 (1988)

1.27 J.M.C. Stork and R.D. Isaac, IEEE Trans. Electron Devices, ED-30, p. 1527 (1983)

1.28 S.M. Sze, Physics of Semiconductor Devices, New York: Wiley (1981)

1.29 R. Beaufoy and J.J. Sparkes, A.T.E.J., 13, p. 310 (1957)

1.30 J.W. Slotboom, Thesis, Technische Hogeschool, Eindhoven (1977)

1.31 D.J. Roulston, S.G. Chamberlain and J. Sehgal, IEEE Trans. Electron Devices, ED-19, p. 809 (1972)

Chapter 2

SIDEWALL EFFECTS ON THE D.C. BASE AND COLLECTOR CURRENTS

2.1 INTRODUCTION

In this chapter we investigate the sidewall effects on the ideal d.c. base and collector currents in vertical n-p-n transistors. These ideal currents are caused by the minority carrier injections in the neutral base and emitter regions of the device. To investigate the sidewall effects on these ideal currents we consider the self-aligned structure as sketched in fig. 2.1. The main process parameters, such as junction depths (W_e, W_b), spacing between emitter sidewall and base contact (t_b), type of emitter contact (poly or metal) and base contact (poly or silicide), are denoted in this figure.

Sources of sidewall effects on the ideal base current are:

(1) Hole injection from the extrinsic base (i.e. the base region outside the emitter area) into the emitter. Roughly speaking, the injected hole current density becomes larger with decreasing distance between the emitter-base junction and the emitter contact because the gradient of the hole concentration becomes larger with decreasing distance. This implies that the injected hole current density at the sidewall can become larger than that at the bottom. A relatively large sidewall effect is therefore expected for this term.

(2) Recombination at the base contact of injected electrons in the extrinsic base [2.1]. This term becomes important with the use of self-aligned techniques because the spacing between this contact and the emitter-base sidewall junction has been reduced to deep submicron dimensions. The contact material (viz. polysilicon or a silicide) is also important.

(3) Recombination at the oxide-silicon interface, both of injected holes in the emitter and of injected electrons in the extrinsic base [2.2]. With the use of improved oxide formation technologies this term

can be expected to become less important because the oxide quality has improved and the number of interface states has therefore decreased.

The sidewall effect on the collector current originates from electron injection into the extrinsic base. The distance between the emitter-base junction and the reverse-biased collector-base junction is larger for the extrinsic base than for the intrinsic base, so the situation is opposite to that in the emitter and a relatively small sidewall effect is expected.

In sections 2.2 and 2.3 the above sidewall effects are investigated in several ways. Analytical calculations are made for a simple two-dimensional structure. In order to obtain practically useful analytical results the transistor is represented by rectangular regions with a constant doping level. See fig. 2.2 for a sketch of this simplified structure. The
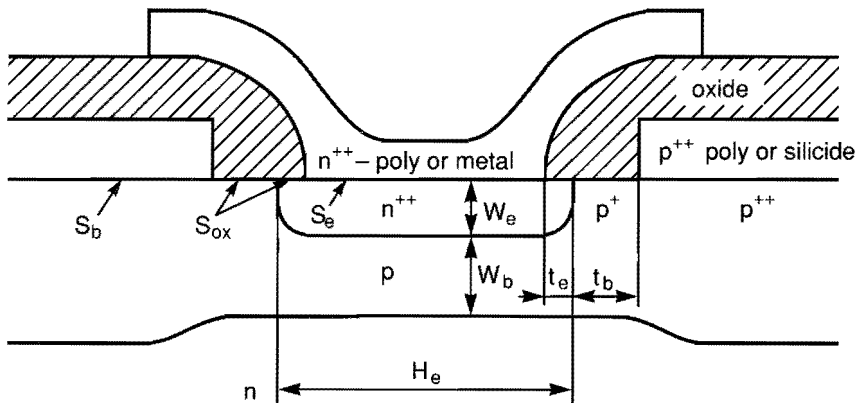
Fig. 2.1. Cross-section of a modern bipolar transistor. The relevant model parameters are also denoted. The interface recombination velocities at the emitter contact, the base contact and at the oxide-silicon interface are denoted by S_e, S_b and S_{ox} respectively.

structure is symmetric about the line $x = 0$. By means of numerical simulations the resulting functional relations between the sidewall effects and the process parameters are tested for more realistic inhomogeneously-doped structures. To this end we have used different doping profiles with different combinations of W_e and W_b (0.05 μm < W_e < 0.3 μm and 0.05 μm < W_b < 0.3 μm) which cover the range of most practical vertical doping profiles used in modern devices. Geometrical scaling rules for the base and collector currents are presented in section 2.4. Based on the calculations presented in sections 2.2 and 2.3, simple expressions are given for the sidewall effects on these currents which may serve as a starting point for a compact model description of these sidewall effects. In section 2.5 ways of measuring the sidewall effects are discussed and a comparison is made between the experimentally observed dependence of the current gain on the lateral emitter dimensions and that obtained by calculations.

2.2 THE BASE CURRENT

2.2.1 Hole injection into the emitter

In order to calculate the hole injection into the emitter we consider the ideal structure as sketched in fig. 2.2. The emitter is represented by a rectangular uniformly-doped region with a constant diffusion length $L_p = \sqrt{\tau_p D_p}$, where τ_p and D_p are the recombination lifetime and diffusion constant, respectively. Since we have assumed a uniformly doped emitter, the electric field is negligibly small outside the space-charge regions and the hole current density consists of a diffusion term only. Expression (1.3b) reduces to

$$\underline{J}_p = - qD_p \underline{\nabla}p. \qquad (2.1)$$

Combination of the stationary hole continuity equation (from (1.2b)) and the expression for the hole diffusion current (2.1) yields the following diffusion equation for the excess hole concentration in the emitter:

$$\underline{\nabla}^2\overline{p}(x,\ y) = L_p^{-2}\overline{p}(x,\ y). \tag{2.2}$$

The excess concentration is given by $\overline{p} = p - p_0$, where p_0 is the equilibrium concentration. In (2.2) we have used the common expression for the recombination rate $R = \overline{p}/\tau_p$. The boundary conditions are as follows:

- The surface recombination rate at the emitter contact can be described by an effective surface recombination velocity S_e, so that the excess hole concentration at the contact satisfies the following equation:

$$- qD_p\left(\frac{\partial\overline{p}}{\partial y}\right)_{y=W_e} = qS_e\,\overline{p}(x,\ W_e). \tag{2.3}$$

- At the emitter surface outside the contact region no normal current density is assumed, so $\partial\overline{p}/\partial y = 0$ there.
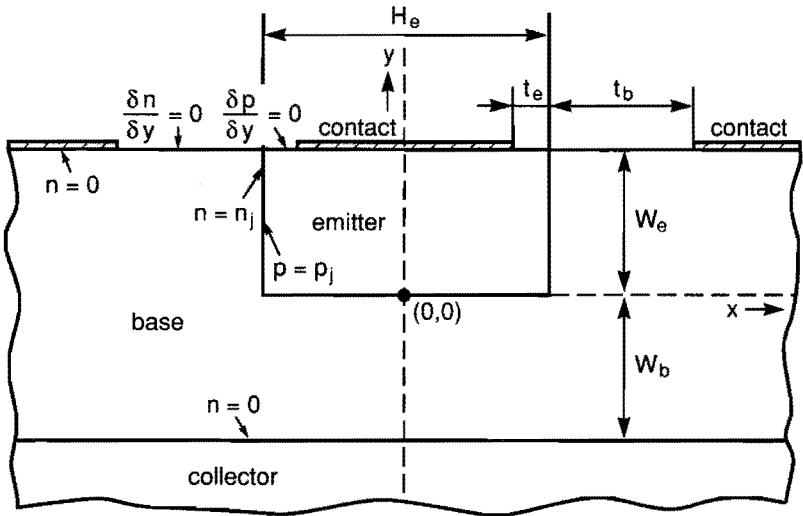
Fig. 2.2. Geometry of the transistor as used for the analytical calculations. The structure is symmetric about the line $x = 0$.

- Along the emitter-base junction the excess hole concentration is constant and its magnitude $\bar{p}_j$ is given by the usual boundary condition $\bar{p}_j = p_0 (\exp (qV_{be}/kT) - 1)$.

For the one-dimensional hole current density J_{p1} the one-dimensional (no x-dependence) solution of (2.2) yields:

$$J_{p1} = q \bar{p}_j \frac{D_p}{L_p} \left[\frac{\sigma + \tanh (W_e/L_p)}{1 + \sigma \tanh (W_e/L_p)} \right], \qquad (2.4)$$

where $\sigma = S_e L_p / D_p$. For a metal contact $S_e \simeq 10^4 m/s$, which implies that for a highly doped emitter σ is in the order of 10, while for a poly layer on top of the emitter $S_e \simeq 10^3 m/s$, which implies that in that case σ is in the order of unity. In both situations the current density becomes independent of L_p in the case of a shallow emitter (i.e. $W_e/L_p << 1$), which means that the current density is determined by surface recombination only. In such a case the emitter is called "transparent" because all injected holes reach the contact. For deep emitters ($W_e >> L_p$) bulk recombination dominates and J_{p1} is virtually independent of σ.

The solution of (2.2) in the two-dimensional case can be obtained by separation of variables [2.3]. The result is:

$$\bar{p}(x,y) = \bar{p}_j \left\{ \sum_{n=0}^{\infty} \left[\frac{A_n \sinh(w[W_e - y]) + B_n \sinh (wy)}{\sinh (wW_e)} \right] \cos (mx) \right.$$

$$\left. + \frac{\cosh (x/L_p)}{\cosh (H_e/2L_p)} \right\}, \qquad (2.5)$$

with

$$m = (n + \frac{1}{2}) \frac{2\pi}{H_e}, \qquad (2.6)$$

and

$$w = (L_p^{-2} + m^2)^{1/2}. \qquad (2.7)$$

In the above expressions W_e and H_e are the emitter thickness and width, respectively. It is shown in appendix 2A that the Fourier coefficients B_n of this expression for $\bar{p}(x, y)$ can be obtained by solving a set of linear equations, while A_n can be solved directly. Once a sufficient number of coefficients has been calculated (usually between 10 and 50), the hole current density can be calculated by using (2.1).

In fig. 2.3 the hole current density J_p at the emitter-base junction, divided by the one-dimensional hole current density J_{p1}, is plotted as a function of the distance l along that junction. The origin in fig. 2.3 corresponds to the point $(H_e /2, W_e)$ in fig. 2.2. The result in fig. 2.3a is calculated for a fairly deep emitter $(W_e = 0.3\mu m)$. This is done for two limiting cases for the emitter width, i.e. A: $H_e << 2L_p$ and B: $H_e >> 2L_p$. The peak of the current density at the surface is due to recombination at the contact edge and, as expected, is found to be strongly dependent on the values of t_e (see fig. 2.2) and S_e. For both an increasing value of S_e and a decreasing value of t_e the gradient in the $x-$ direction of the hole density at the surface between the junction and the contact increases, which leads to an increase in this peak value of the current density. The dip in the current density near the corner originates from the hole injections from both the horizontal and the vertical junctions. These injections influence each other in such a way that within a distance in the order of the hole diffusion length from the emitter corner the gradient of the hole density is lower than in the one-dimensional case. From fig. 2.3a it can also be seen that in the middle of the device (i.e. at the origin in fig. 2.2) J_p equals J_{p1} for case B $(H_e > 2L_p)$, while for case A J_p is considerably lower than J_{p1}. In the latter case a linear scaling rule for the base current with the emitter width is not valid! Fig. 2.3b shows a similar plot for a shallow emitter $(W_e = 0.1\mu m)$. In this case the bottom current density reaches its one-dimensional value within a distance from the emitter sidewall which is much less than the diffusion length. In this case a linear scaling rule for the base current breaks down at a much smaller emitter width than in the case of a deep emitter (fig. 2.3a). This will be discussed further at
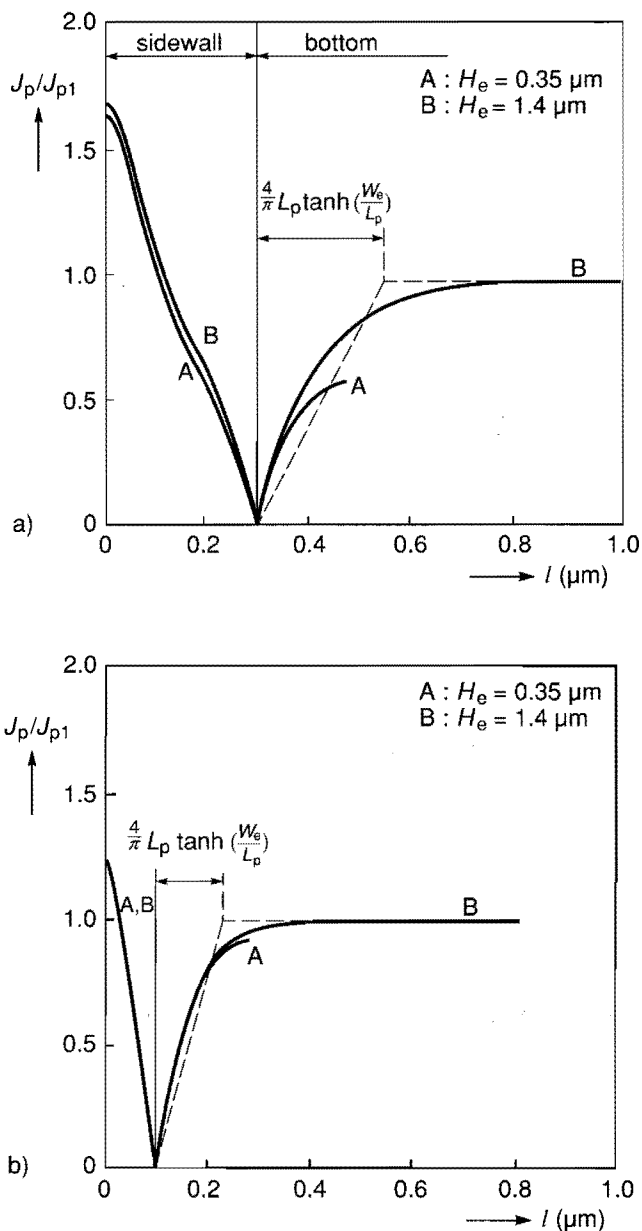
Fig. 2.3. Normalized hole current density along the emitter-base junction for two emitter widths. In case (a), $W_e = 0.3\ \mu m$, $S_e = 10^4\ m/s$. In case (b), $W_e = 0.1\ \mu m$ and $S_e = 10^3\ m/s$. In both cases the hole diffusion length L_p is $0.25\ \mu m$, $t_e = W_e/4$ and $D_p = 10^{-4}\ m^2/s$.

Sidewall effects on the d.c. base and collector currents 21
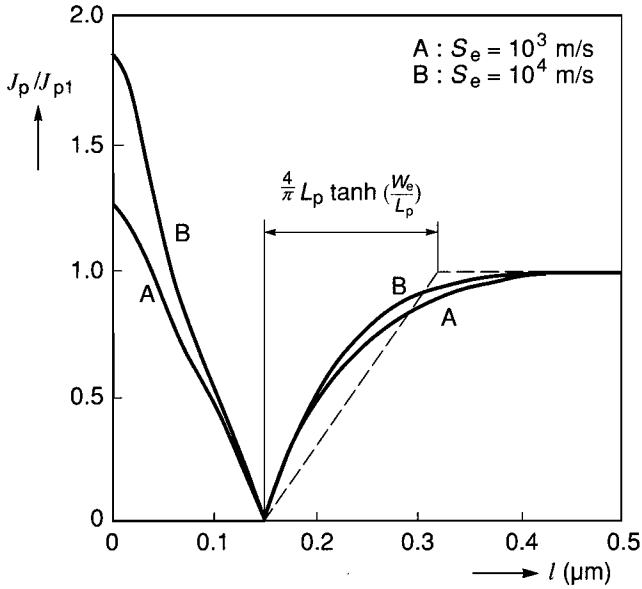
Fig. 2.4. Normalized hole current density along the emitter-base junction for two values of the surface recombination velocity. $W_e = 0.15 \, \mu m$ and $H_e = 1.4 \, \mu m$. The other parameters are equal to those in fig. 2.3a.

a later stage. In fig. 2.4 J_p is plotted for a $0.15 \mu m$ deep emitter for two values of the surface recombination velocity. From this figure we see that a reduction of S_e greatly reduces the value of the current density through the sidewall near the surface, while the shape of the bottom current density is more or less unaffected. The reason for this is that the injected current density through the junction is virtually independent of the surface recombination velocity when $S_e >> D_p/d$, where d is the distance between the junction and the contact. For $10^3 < S_e < 10^4 m/s$ this is true for the injection through the bottom in this case but not for the lateral injection through the sidewall at the surface. The physical interpretation of this effect is that when

$S_e > D_p / d$ the hole injection into the emitter is limited by the hole transport, while for $S_e < D_p / d$ the limiting factor is surface recombination.

Since the above analytical results are based on the assumptions of homogeneously-doped regions, numerical calculations on structures with more realistic doping profiles have been made. These calculations show that the hole current distribution along the emitter-base junction behaves in roughly the same way as the analytically calculated current distribution. As an example in figs. 2.5 and 2.6 the numerically calculated hole current density along the emitter-base junction is given for two cases. In these figures the current densities are normalized to their one-dimensional (vertical) value. The doping profiles in figs. 2.5b and 2.6b are along cross-section A-A as given in fig. 1.3. In figs. 2.5 and 2.6 we can recognize the two striking features of the hole current density, i.e. the peak in the current density at the surface and the dip in the current density near the corner of the emitter-base junction. This dip is less pronounced than in the analytical results. This is due to the fact that in the numerical calculations we have a rounded corner instead of a sharp corner. The dip in the current density makes a negative contribution to the sidewall effect on the hole injection current I_{be}, while the peak of the current density at the surface makes a positive contribution. For a washed emitter, where t_e is determined only by the lateral diffusion of the emitter dopant, the net result is, rather surprisingly, a small sidewall effect. In fig. 2.6 we can again observe the fact that for a relatively deep emitter the current density at the bottom does not reach the one-dimensional value J_{p1} when $H_e < 2L_p$.

In the rest of this section we examine the above observations in greater detail. First we will look at the bottom part of the hole injection current in the case where the influence of surface recombination on the bottom current is negligible, i.e. for fully non-transparent emitters. For the two limiting cases as shown in fig. 2.3, viz. $H_e >> 2L_p$ and $H_e << 2L_p$ simple expressions for bottom hole current as a function of the emitter width will be presented. After that we will take a closer look at the sidewall part of the hole injection current. Finally, we investigate the geometry dependence of the hole current in the case
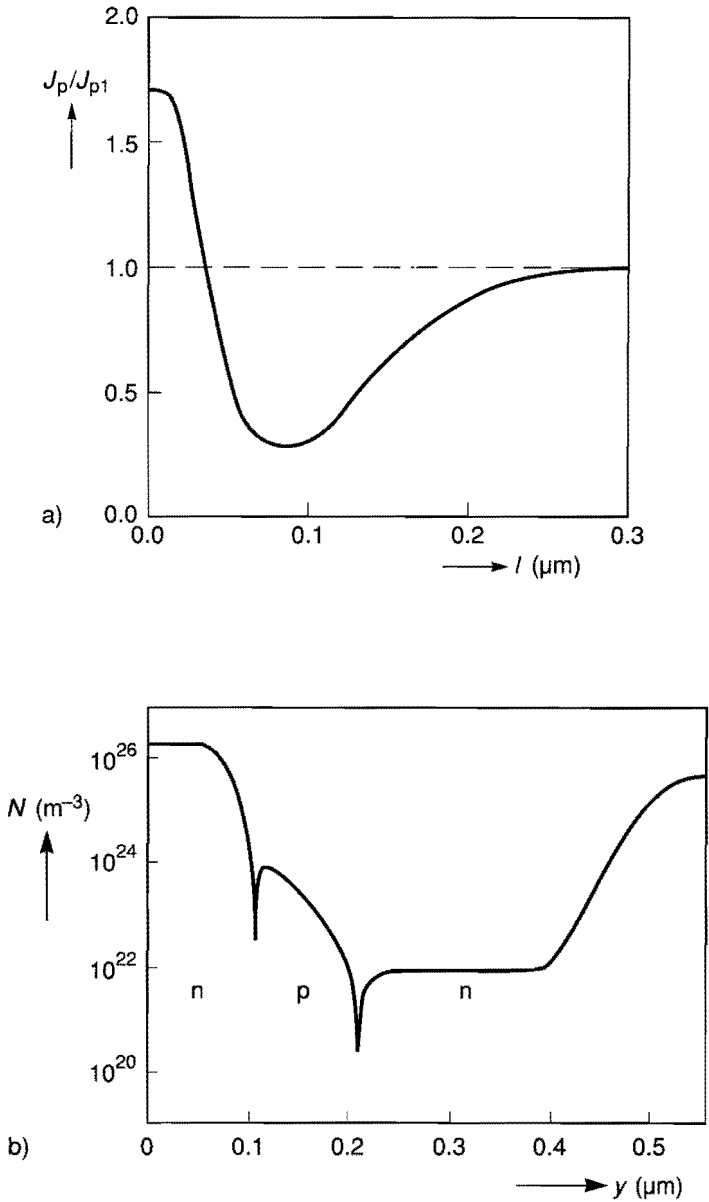
Fig. 2.5. Numerically calculated normalized hole current density along the emitter-base junction. $W_e = 0.1\ \mu m$, $t_e = 0.05\ \mu m$ and $H_e = 0.4\ \mu m$ while $S_e = 10^4\ m/s$. The doping profile in the depth direction of the intrinsic part of the transistor is given in b). $y = 0$ is at the emitter surface.
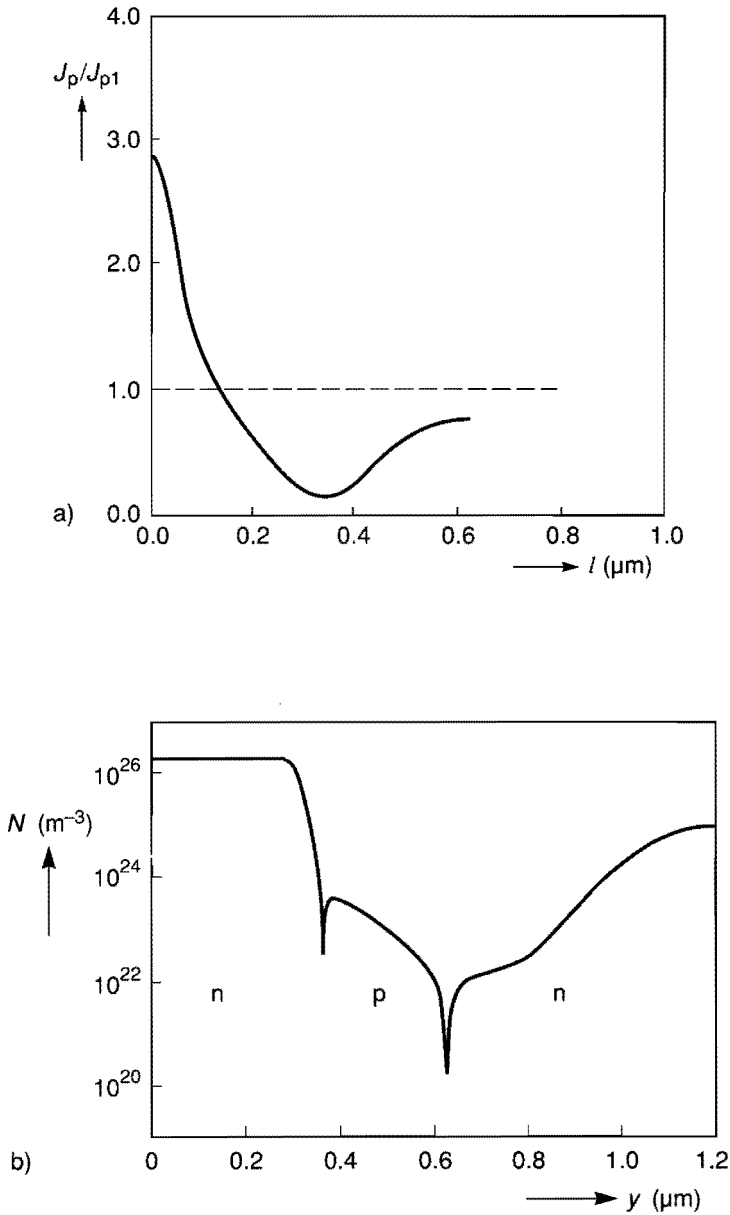
Fig. 2.6. Numerically calculated normalized hole current density along the emitter-base junction. $W_e = 0.35\ \mu m$, $t_e = 0.1\ \mu m$ and $H_e = 0.35\ \mu m$ while $S_e = 10^4\ m/s$. The doping profile in the depth direction of the intrinsic part of the transistor is given in b).

Sidewall effects on the d.c. base and collector currents 25

where the emitter is very shallow and can be considered to be transparent.

First we look at the bottom part of the hole current. This part of the hole current I_{beb} (per unit length in the third direction) can be obtained from the following relation:

$$I_{beb} = - 2qD_p \int_0^{He/2} \left(\frac{\partial \overline{p}}{\partial y} \right)_{y=0} dx. \tag{2.8}$$

When the influence of surface recombination on the bottom current is negligible ($\sigma < < \tanh (W_e/L_p)$ or $W_e > > L_p$) simple approximate expressions for I_{beb} can be obtained. When $\sigma < < \tanh (W_e/L_p)$ the non-diagonal elements M_{nj} ($n \neq j$) in (2A.11) vanish and $B_n = Z_n / M_{nn}$. In the case where $W_e > > L_p$ the coefficients B_n are much smaller than A_n and can be neglected. Substitution of the resulting expressions for A_n and B_n in (2.5) and subsequent substitution of the partial derivative of this expression in (2.8) gives for the bottom part of the hole injection current

$$I_{beb} = J_{p1} H_e \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{R_n}{(2n + 1)^2} , \tag{2.9}$$

with

$$R_n = \frac{\tanh \left[\sqrt{1 + (mL_p)^2} \; \dfrac{W_e}{L_p} \right]}{\sqrt{1 + (mL_p)^2} \; \tanh \left(\dfrac{W_e}{L_p} \right)} , \tag{2.10}$$

where m is given by (2.6). The term R_n approximately equals unity for $mL_p \tanh (W_e/L_p) < < 1$, while for $mL_p \tanh (W_e/L_p) > > 1$ this term is approximately inversely proportional to n.

For wide emitters ($H_e > 2L_p$), we can approximate (2.9) by replacing the term R_n by a step function which is equal to 1 for $mL_p \tanh (W_e/L_p) < 1$ and equals zero for $mL_p \tanh (W_e/L_p) > 1$. The current then becomes

$$I_{beb} = J_{p1} H_e \frac{8}{\pi^2} \sum_{n=0}^{N} \frac{1}{(2n+1)^2} \simeq J_{p1}[H_e - \frac{4}{\pi} L_p \tanh(\frac{W_e}{L_p})] , (2.11)$$

where $N = \text{trunc}[H_e/(2\pi L_p \tanh (W_e/L_p))]$. This result can be found in [2.4] by using the asymptotic expression for the first derivative of the digamma function. For $H_e > 6L_p$ the error introduced by the approximations leading from (2.9) to (2.11) is less than 2%. An illustrative approximation of (2.11) is given in figs. 2.3 and 2.4, where the approximate value of I_{beb} as obtained from (2.11) is given by the corresponding area below the dashed line.

For narrow emitters ($H_e << 2L_p$), it holds that $mL_p \tanh (W_e/L_p) >> 1$ for all values of n. In that case we can approximate (2.9) by replacing the term $\sqrt{1 + (mL_p)^2}$ in (2.10) by mL_p and take into account only the first term of the series. This yields

$$I_{beb} = \simeq \frac{8}{\pi^3} J_{p1} \frac{H_e^2}{L_p} \frac{\tanh (\pi W_e/H_e)}{\tanh (W_e/L_p)} . \qquad (2.12)$$

In fig. 2.7 the approximations (2.11) and (2.12) are plotted as a function of H_e/L_p for $W_e/L_p = 1$. To show the validity of these approximations relation (2.9) is also plotted. The values of I_{beb} are normalized to $J_{p1} H_e$.

From the above analysis it follows that the quantity $L_p \tanh (W_e/L_p)$ is an important parameter for the bottom current, because within a distance of about $L_p \tanh (W_e/L_p)$ from the sidewall the bottom current density is strongly diminished as compared with the one-dimensional solution. This implies that when H_e is in the range of $2L_p \tanh (W_e/L_p)$ or less, the bottom current is much less than $H_e J_{p1}$, which is the value of I_{beb} resulting from a simple one-dimensional analysis. Furthermore, from (2.12) it can be seen that when the emitter
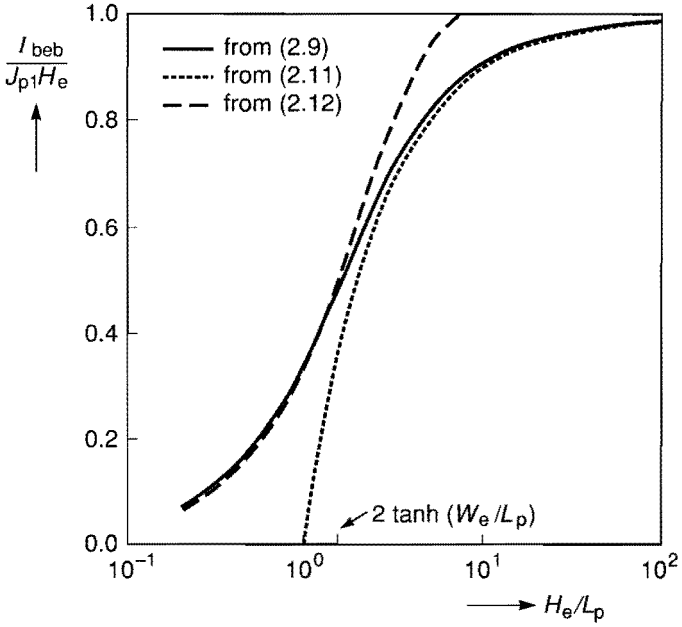
Fig. 2.7. Expressions (2.9), (2.11) and (2.12) as a function of H_e/L_p for $W_e/L_p = 1$. The values of I_{beb} are normalized to $J_{p1}H_e$.

width is in this range the bottom current does not scale linearly with H_e and, hence, such a linear scaling rule of the base current with the emitter area current is no longer valid.

The sidewall part of the hole injection current I_{bes} is given by

$$I_{bes} = 2qD_p \int_0^{We} \left(\frac{\partial \overline{p}}{\partial x} \right)_{x=He/2} dy. \tag{2.13}$$

For large values of the surface recombination velocity the base current through the sidewall I_{bes} strongly depends on the distance between the sidewall junction and the contact edge, especially when $W_e << L_p$. This
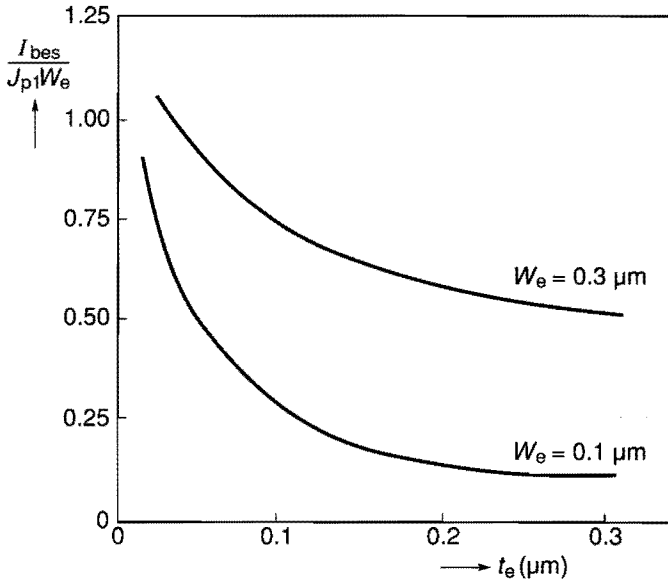
Fig. 2.8. Normalized sidewall current density as a function of the distance between the sidewall junction and the contact edge t_e for $W_e = 0.1\ \mu m$ and $W_e = 0.3\ \mu m$. The other parameters are equal to those of fig. 2.3a, while $H_e = 4\ \mu m$.

is illustrated in fig. 2.8. The normalized sidewall current $I_{bes}/(J_{p1}\ W_e)$ can be considered as the ratio of the average current density through the sidewall and J_{p1}. From fig. 2.8 we observe that for these cases the average current density through the sidewall is approximately equal to or less than the average current density through the bottom. When $W_e < L_p$ and $t_e > L_p$ (the lower curve of fig. 2.8 for $t_e \simeq 0.3 \mu m$) this ratio is much less than unity because in that case the sidewall current is mainly determined by bulk recombination, while the bottom current is mainly determined by surface recombination. In [2.5] an approximate analytical expression for this sidewall current is given.

Sidewall effects on the d.c. base and collector currents
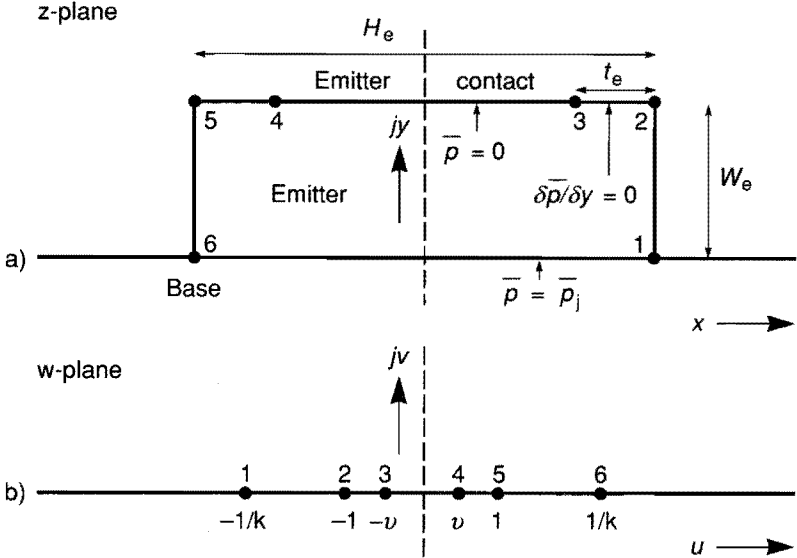
Fig. 2.9. Contours of the emitter region in the complex z-plane (a) and in the complex w-plane (b). The quantities k and v are determined from the dimensions of the region in the z-plane.

In modern emitters which are almost transparent ($W_e << L_p$) bulk recombination is less important. If we neglect bulk recombination, the diffusion equation (2.2) reduces to the Laplace equation

$$\underline{\nabla}^2 \bar{p}\,(x,\,y) = 0. \tag{2.14}$$

When both an infinite surface recombination velocity and a homogeneous dope is assumed, the hole current can be determined analytically by conformal mapping, using a Schwarz-Christoffel transformation [2.17]. This technique has the advantage that closed analytical expressions for the current can be found instead of infinite series such as in a Fourier expansion technique. For the geometry as given in fig. 2.9a the solution can be obtained by mapping the emitter region onto the upper half plane in fig. 2.9b. The analysis is given in appendix 2B and

yields expression (2B.12) for I_{be}. In the case where $H_e >> W_e$ this expression reduces to:

$$I_{be} = J_{p1} \left\{ H_e + \frac{4W_e}{\pi} \ln\left[\mathrm{csch}\left(\frac{\pi t_e}{2W_e} \right) \right] \right\}, \qquad (2.15)$$

where we can observe the linear scaling of the hole current with the emitter width. However, this rule breaks down when $H_e < 2W_e$. For washed emitters the ratio t_e / W_e is approximately constant, so the sidewall effect is proportional to W_e.

For non-washed emitters, where t_e is much larger than in the case of a washed emitter, the situation is different. For $t_e / W_e >> 1$ the
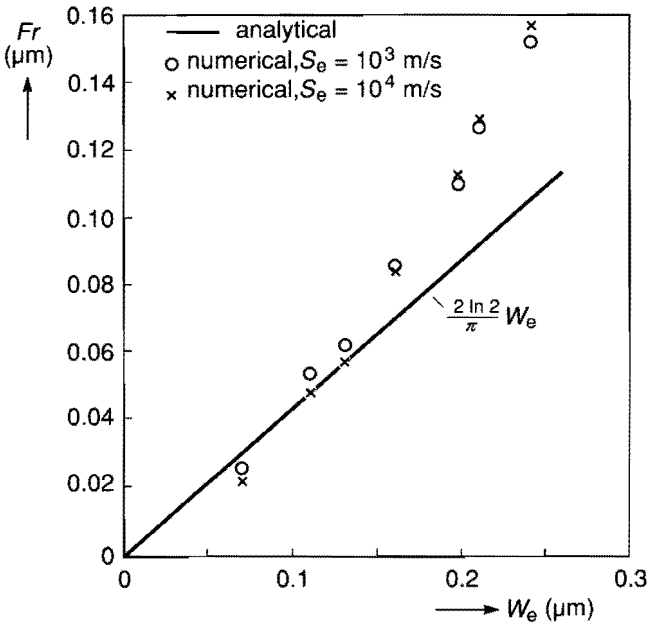
Fig. 2.10. Numerically and analytically calculated fringing effect as a function of the emitter depth W_e.

sidewall junction plays no role and the hole injection is confined to the region under the contact. In that case (2.15) reduces to

$$I_{be} = J_{p1}\left[H_c + \frac{4ln2}{\pi} W_e \right],$$ (2.16)

where H_c is the contact width ($H_c = H_e - 2t_e$). The second term on the right-hand side of (2.16) accounts for the fringing effect near the edge of the contact. By means of numerical simulations it is found that for $W_e < 0.2 \, \mu m$ the above equation also gives a good indication for realistic emitter doping profiles and for a finite surface recombination velocity. This is shown in fig. 2.10 where the second term on the right-hand side of (2.16) (i.e. the fringing effect $Fr = \frac{2 \ln 2}{\pi} W_e$) is plotted, together with the corresponding numerical results.

2.2.2 Recombination at the base contact

The recombination current at the base contact is also calculated analytically by means of conformal mapping. To this end we use the simplified, rectangular base geometry as sketched in fig. 2.2. If a constant base doping level is assumed and bulk recombination in the base is neglected, the differential equation which governs the electron injection into the base is the Laplace equation. If we take an infinite surface recombination velocity at the base contact, the boundary condition for both the base contact and the reverse-biased collector-base junction is $n = 0$. Instead of imposing the boundary condition $\partial n/\partial x = 0$ at $x = 0$ we have taken a single vertex at $x = -\infty$ (see fig. 2.11a). As shown below, this enables us to derive a simple expression for the sidewall effect, while for $H_e > 2W_e$ the above-mentioned boundary condition is automatically satisfied to a high degree of accuracy. Using the boundary conditions denoted in fig. 2.11a, the Laplace equation is solved by conformal mapping, using a Schwarz-Christoffel transformation. By means of this transformation the base geometry in the complex z-plane (fig. 2.11a) is mapped onto the upper half w-plane (fig. 2.11b). In appendix 2C it is shown that for $t_b > W_b$ the recombination current at the base contact I_{bb} (per unit length in the third dimension) is given by
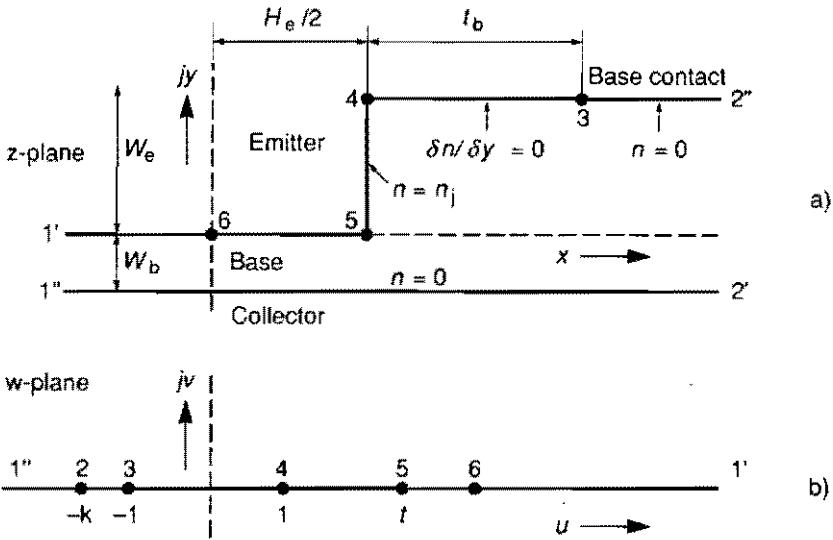
Fig. 2.11. Contours of the base region in the complex z-plane (a) and in the complex w-plane (b).

$$I_{bb} = qD_n \, n_j \, \frac{4}{\pi} \, (1 - \alpha^2)^{-1/2} \left(\frac{1 - \alpha}{1 + \alpha} \right)^{\alpha/2} \exp \left[\frac{-\pi t_b}{2 \, (W_e + W_b)} \right], \quad (2.17)$$

with $\alpha = W_b/(W_b + W_e)$. In (2.17) D_n is the diffusion constant for electrons. From (2.17) it is observed that this current depends exponentially on the spacing t_b with a characteristic length $t_0 = \frac{2}{\pi} (W_b + W_e)$. Physically speaking we can interpret this result as follows: Both the base contact and the reverse-biased collector-base junction, which can also be considered as a contact with an infinite surface recombination velocity, act as a sink for electrons. When the base contact is much farther away from the emitter sidewall than the collector-base junction is from the silicon surface, i.e. $t_b >> (W_e + W_b)$, the electron flow in the external base is almost entirely determined by the collector-base junc-
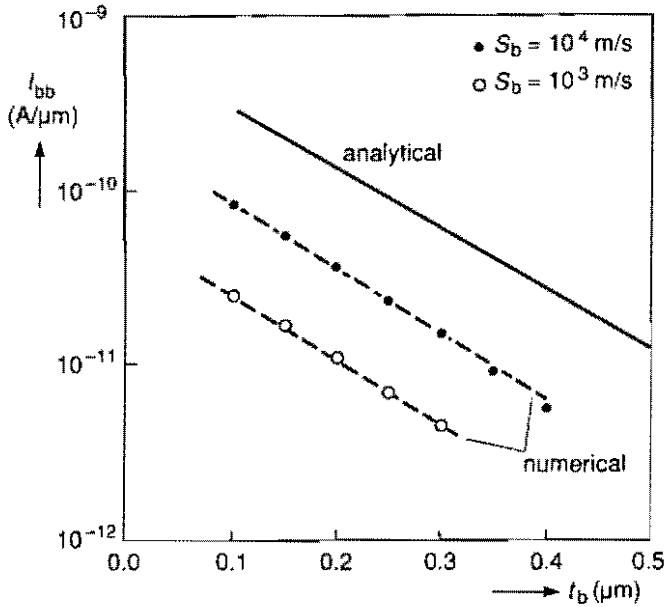
Fig. 2.12. Analytically and numerically calculated values of I_{bb} as a function of the spacing between emitter sidewall and base contact. For this device $W_e = 0.1$ μm and the quasi-neutral base width $W_b = 0.1$ μm. The numerically obtained value of n_j at the top dope in the extrinsic base is taken for the analytical calculations.

tion and few electrons will recombine at the base contact. In the opposite case, $t_b <$ $(W_e + W_b)$, a large amount of the electrons will recombine at the base contact.

As an example, analytically and numerically calculated values of I_{bb} are plotted as a function of the spacing t_b in fig. 2.12. Note that the numerically calculated values of I_{bb} are much smaller than the analytically calculated values. This is due to both the presence of an additional base dopant under the contact and a finite surface recombination velocity at the base contact in the numerical case. Notice that for both

types of calculations an exponential relation between I_{bb} and t_b is found, though with a slightly different slope. We will come back to this in section 2.4.

2.2.3 Recombination at the silicon-oxide interface

To model recombination at the oxide-silicon interface we use a linear relationship between the current and the excess minority concentration at the interface (i.e. $J_n = qS_{ox} (n - n_0)$ in the base region). In the above expression S_{ox} is an effective surface recombination velocity. Although there are more elaborate models for describing recombination at the oxide-silicon interface [2.13], we use this well-known expression [2.6] in order to arrive at a simple analytical result. The analytical model for this sidewall effect is based on the following three considerations:

- Because the doping level in the emitter region is much higher than in the base, the injected minority concentration at the silicon-oxide interface is much lower in the emitter than in the base. Recombination at the silicon-oxide interface in the emitter is therefore neglected.
- From numerical simulations it is observed that for $S_{ox} < 10^3 \, m/s$ the recombination current is proportional to S_{ox} or, in other words, the electron charge in the extrinsic base is not significantly affected by the oxide recombination.
- In practical situations also the presence of the base contact does not significantly affect the electron charge in the base region.

The oxide recombination current (per unit length in the third dimension) can therefore be written as (see fig. 2.2)

$$I_{box} = qS_{ox} \int_{He\,/2}^{\infty} n\,(x,W_e)\,dx, \qquad (2.18)$$

and the above integral can be evaluated using the original boundary conditions as given in fig. 2.2, i.e. for $n\,(x,y)$ we use the solution found for $S_{ox} = 0$. This results in (see appendix 2D)
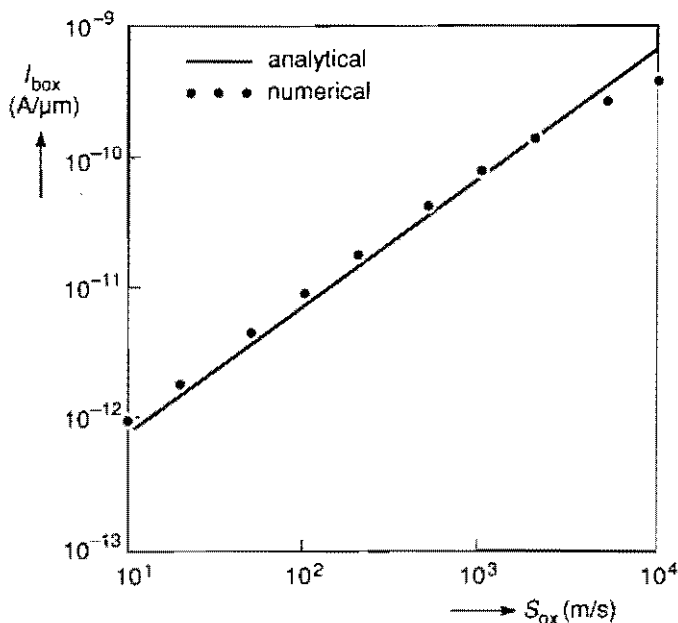
Fig. 2.13. Analytically and numerically calculated values of I_{box} as a function of the oxide recombination velocity. For this device $W_e = 0.1 \ \mu m$ and the quasi-neutral base width $W_h = 0.1 \ \mu m$. The numerically obtained value of n_j at the top dope in the extrinsic base is taken for the analytical calculations.

$$I_{box} = qS_{ox} \frac{8}{\pi^2} n_j G W_h \sqrt{\frac{1 - \alpha^2}{\alpha^2}}, \qquad (2.19)$$

where G is the Catalan constant ($G = 0.91596...$). As an example, fig. 2.13 shows the linear dependence of the numerically calculated oxide recombination current on the oxide recombination velocity up to values of $S_{ox} \simeq 10^3 m/s$, together with the analytical results obtained from (2.19).

2.3 THE COLLECTOR CURRENT

To calculate the sidewall effect on the collector current we use the simplified structure given in fig. 2.11a. Since in practical cases the recombination current at the base contact is much less than the collector current we can derive the expression for the collector current in the absence of the base contact ($t_b \to \infty$). In that case the quantity k (see fig. 2.11) equals unity, as can be observed from (2C.8), where $\gamma = k - 1$. In appendix 2E it is shown that the collector current (per unit length in the third dimension) is given by

$$I_c = J_{n1} \left\{ H_e + \frac{2W_b}{\pi} \left[\left(2 + \frac{W_e}{W_b} \right) \ln \left(2 + \frac{W_e}{W_b} \right) - \frac{W_e}{W_b} \ln \frac{W_e}{W_b} \right] \right\},$$

$$(2.20)$$

where the one-dimensional current density is given by

$$J_{n1} = \frac{q \, D_n \, n_j}{W_b} . \qquad (2.21)$$

The second term in the right-hand side of (2.20) can be considered as the sidewall effect on the collector current. As expected, this effect increases with increasing value of the ratio W_e/W_b. This can be seen from fig. 2.14 where the normalized collector current $I_c/(J_{n1}H_e)$ is plotted as a function of H_e/W_b for three values of W_e/W_b. Note that for $W_e = 0$ solving the Laplace equation for electrons in the base region is the same problem as solving the Laplace equation for holes in the emitter region in the case where $t_e \to \infty$ (fig. 2.9). Therefore, under these circumstances we have the same mathematical problem for the electron current in the base as for the hole current in the emitter. As can be seen from (2.20), in that case the expression for the collector current reduces to an expression similar to (2.16), as it should do.

Fig. 2.15 shows the normalized electron current density J_n/J_{n1} along the base-emitter junction. The parameters are equal to those of

Sidewall effects on the d.c. base and collector currents   *37*

fig. 2.3a, while $W_e/W_b = 2$. The electron current density appears to diverge at the sharp emitter corner. At this point it is interesting to note a striking analogy: If we replace the concentration n_j by the electrostatic potential and consider the emitter and collector regions as electrical conductors we have obtained an electrostatic capacitor-like problem. In this analogy the current density is replaced by the electric field. Problems of this kind have been intensively dealt with analytically in the first half of this century in connection with electrostatic breakdown phenomena in electrical machines and transformers [2.7]. From these analyses it appears that the singularity at the corner disappears when this corner is rounded off. In ref. [2.7] it is shown that the maximum electric field (current density) depends on the ratio r/W_b, where r is the
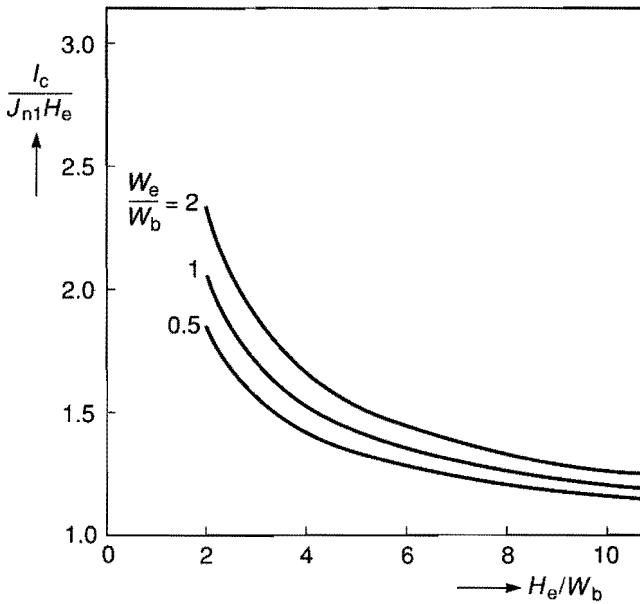
Fig. 2.14. Normalized collector current $I_c/(J_{n1}H_e)$ plotted as a function of H_e/W_b for $W_e/W_b = 0.5$, 1 and 2, respectively.
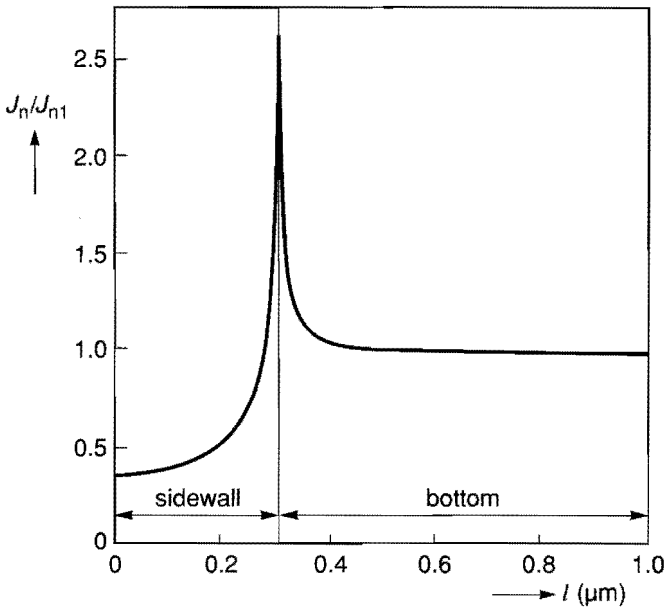
Fig. 2.15. The normalized electron current density along the emitter-base junction. In this case $W_e/W_b = 2$.

radius of the rounded corner. This effect, of course, also occurs in realistic transistors, as shown below.

When we compare the numerically and analytically calculated electron current distributions along the emitter-base junction (see fig. 2.16) we can observe two effects which tend to reduce the sidewall effect on the collector current, as compared with the analytical model. First, as already mentioned before, we have a rounded corner ($r/W_b \simeq 1$) instead of a sharp corner. The maximum of the current density is about 15% higher than J_{n1} and the location of this maximum is shifted slightly towards the centre of the device. This is in accordance with the observations in ref. [2.7]. The second deviation from the analytical solution is that the numerically calculated electron current density through the sidewall is considerably lower. This is mainly caused by the fact that in

Sidewall effects on the d.c. base and collector currents   *39*

a realistic structure the base doping increases towards the surface and, therefore, the injected minority carrier density near the junction n_j decreases.

2.4 GEOMETRICAL SCALING RULES

In practice the total base current can be written as follows:

$$I_b \equiv J_{p1}[A_e + P_e Y_b] , \qquad (2.22)$$

where Y_b is defined as

$$Y_b \equiv Y_{be} + Y_{bb} + Y_{box} . \qquad (2.23)$$
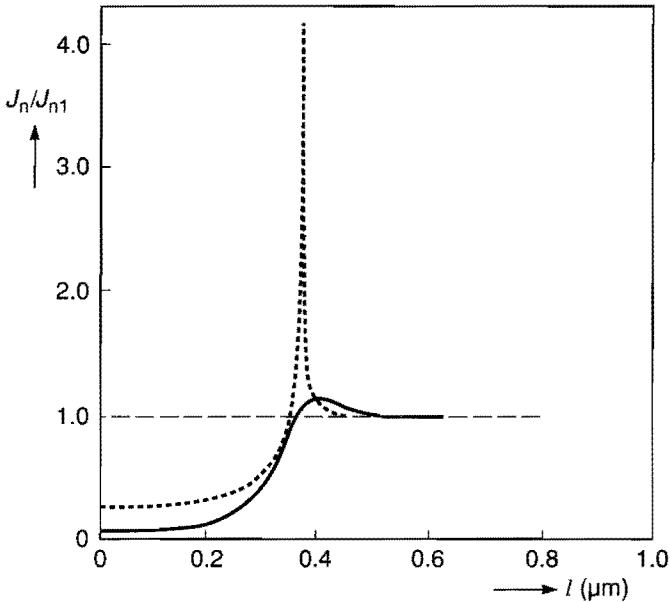
Fig. 2.16. Numerically calculated electron current density along the emitter-base junction for the same device as that for which the results are presented in fig. 2.6. The dotted line gives the corresponding analytical result.

The quantities Y_{be}, Y_{bb} and Y_{box} account for hole injection into the emitter, recombination at the base contact and recombination at the silicon-oxide interface, respectively (see also section 2.1). A_e is the bottom area of the emitter ($A_e = H_e L_e$, where L_e is the lateral emitter dimension in the third direction) and P_e is the emitter perimeter ($P_e = 2 (H_e + L_e)$). J_{p1} is equal to the bottom base current density for an infinitely large emitter area. The product $P_e Y_b$ can be considered as an effective emitter sidewall area, which accounts for the sidewall effects on the base current. This way of describing the sidewall effect has the advantage that its relative importance can immediately be seen from a comparison of Y_b with the emitter dimensions. The above expression is valid in cases where the sidewall effects are equal for the four emitter sides or where the emitter width is much smaller than its length ($H_e < < L_e$). By means of numerical simulations it has been verified that in practical cases the three contributions are indeed additive, so that (2.23) is valid. It should be noted here that in section 2.2.1 it is shown that when the emitter width H_e is in the range of $2L_p \tanh (W_e/L_p)$ or less, the bottom current I_{beb} is much less than $H_e J_{p1}$, which is the value of I_{beb} resulting from a simple one-dimensional analysis. In that case the bottom base current does not scale linearly with H_e and, hence, a linear scaling rule for the base current of the form of (2.22) is no longer valid.

Similarly, the collector current can be written as

$$I_c \equiv J_{n1}[A_e + P_e Y_c] , \qquad (2.24)$$

where J_{n1} is equal to the bottom collector current density for an infinite emitter area. Y_c accounts for the sidewall effect on the collector current. From figs. 2.15 and 2.16 it can be seen that a linear scaling rule for the collector current is valid down to very narrow emitter widths.

In the rest of this section we present simple expressions for Y_{be}, Y_{bb}, Y_{box} and Y_c which may serve as a starting point for a compact model description of these sidewall effects.

First we consider hole injection into the emitter. For a washed emitter, where t_e is determined only by the lateral diffusion of the emitter dopant, the ratio t_e / W_e can be assumed to be constant. It can

be seen from (2.15) that in this case Y_{be} is proportional to W_e. So we write

$$Y_{be} = K_{be} W_e. \tag{2.25}$$

The value of K_{be} depends on the detailed lateral doping profile. For $t_e/W_e = 0.5$, which is a realistic value for washed emitters, and error-type lateral profiles, a good indication for Y_{be} is obtained with the values of K_{be} as given in table 2.I. For $t_e/W_e = 0.5$, it follows from (2.15) that the proportionality constant is 0.09, while from the numerical calculations this constant is found to be 0.2 for a metal contact. The difference is caused by the fact that in the numerical calculations the lateral diffusion causes the doping concentration near the contact edge to be lower than under the contact far from the edge. The injected minority concentration at the contact edge is therefore higher than in the analytical case. Note the resulting negative value of Y_{be} for the case of a polysilicon emitter contact. This is due to the fact that the relative height of the peak current density at the surface is strongly reduced, as compared with the case of a metal contact, while the shape of the bottom current density is more or less unaffected. This is illustrated in fig. 2.4. As discussed in section 2.2.1, Y_{be} consists of two contributions, i.e. a negative contribution from the dip in the bottom current density near the emitter edge and a positive contribution from the hole injection through the sidewall. The net result is a fairly small effect, which turns out to be slightly positive in the case of a metal contact and slightly negative for a poly contact.

For non-washed emitters, for which t_e is much larger than in the case of a washed emitter, the sidewall effect becomes negative because the hole injection is confined to the region under the contact. Using (2.16), we obtain

$$Y_{be} = \frac{2 \ln 2}{\pi} W_e - t_e. \tag{2.26}$$

As shown in section 2.2.1 the above expression also gives reasonable results for inhomogeneously-doped emitters, provided that $W_e < 0.2 \mu m$.

Table 2.I. Values for the prefactors in the expressions for Y_{be}, Y_{bb}, Y_{box} and Y_c for different values of the emitter recombination velocity S_e, base contact recombination velocity S_b and for the top dope of the base under the contact p^{++}.

	K_{bb}			K_{box}	K_{be}	K_c
	p^{++} (m^{-3})					
	10^{25}	5.10^{25}	10^{26}			
$S_b = 10^4$ m/s (metal)	0.14	0.08	0.06			
$S_b = 10^3$ m/s (poly)	0.04	0.025	0.02	0.5		0.2
$S_e = 10^4$ m/s (metal)					0.2	
$S_e = 10^3$ m/s (poly)					-0.2	

To derive the contribution of the recombination at the base contact to the sidewall effect Y_{bb}, I_{bb} is written in the desired form $J_{p1} Y_{bb}$ by using $\beta_\infty = J_{n1}/J_{p1}$, which gives $J_{p1} = q n_j D_n /(\beta_\infty W_b)$. The resulting expression for Y_{bb} can be obtained from (2.17). This gives

$$Y_{bb} = K(\alpha) \, \beta_\infty \, W_b \, \exp\left(- t_b/t_0\right),\qquad(2.27)$$

with

$$K(\alpha) = \frac{4}{\pi} \, (1 - \alpha^2)^{-1/2} \left(\frac{1 - \alpha}{1 + \alpha}\right)^{\alpha/2}.\qquad(2.28)$$

In order to make the above equation applicable to practical situations with an inhomogeneously-doped base, a reduced emitter depth $\overline{W}_e$ is defined by the introduction of a quantity r, according to $\overline{W}_e = r \, W_e$. The value of r, which is defined as the ratio of the injected minority

concentration at the peak dopant concentration in the extrinsic base to that in the intrinsic base, i.e. $r = (n_{ie}^2/N)_{ext}/(n_{ie}^2/N)_{int}$, lies between 0.5 and 0.8 in practical situations. The resulting expression for Y_{bb} is:

$$Y_{bb} = K_{bb}\, \beta_\infty\, r\, W_b\, \exp\, (-t_b/\bar{t}_0),\qquad\qquad (2.29)$$

with

$$\bar{t}_0 = \frac{2}{\pi}\, (W_b + r\, W_e).\qquad\qquad (2.30)$$
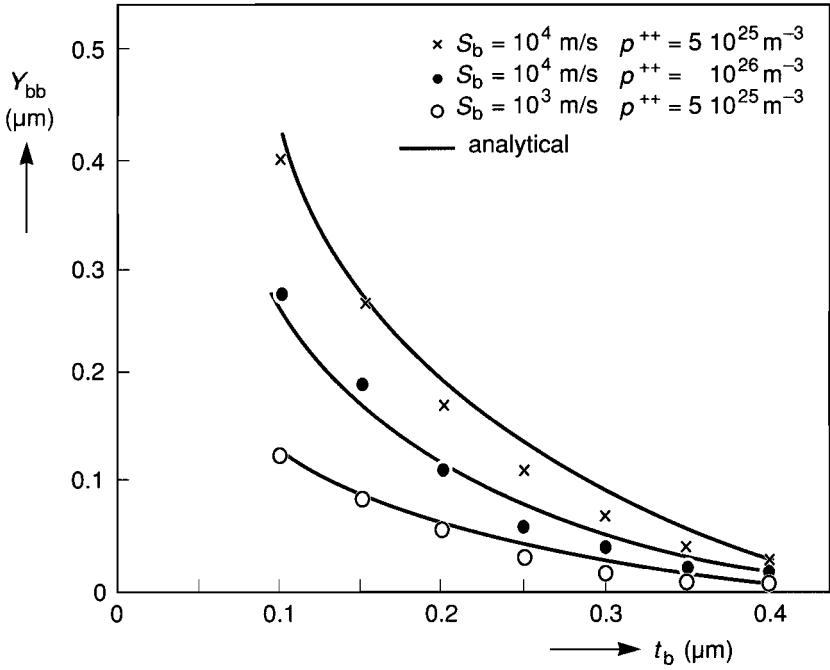
Fig. 2.17. Calculated values of Y_{bb} as a function of the distance between the emitter sidewall and the base contact t_b. For this device $W_e = 0.1\ \mu m$ and the quasi-neutral base width $W_b = 0.1\ \mu m$. $\beta_\infty = 130$.

In (2.29) the term $K_{bb}r$ replaces $K(\alpha)$. The weak dependence of $K(\alpha)$ on the ratio W_e/W_b (through α) is neglected, which is sufficient for our purposes. The prefactor K_{bb} is determined from numerical results and is given in table I for two values of the base surface recombination velocity and for three values of the top of the base dope under the contact. Note that the values of K_{bb} are much smaller than $K(\alpha)$. As already pointed out in section 2.2.2, this is due to the fact that both the presence of an additional base dopant under the contact and a finite surface recombination velocity at the base contact decreases the value of I_{bb}, as compared with the analytical solution. As can be seen from (2.30) the value of the characteristic length $\bar{l}_0$ is slightly reduced due to the presence of the inhomogeneous base dope. This can also be seen from fig. 2.12 where the numerical calculations give a smaller characteristic length than the analytical calculations. As an example, in fig. 2.17 the results, both numerical and analytical, are given for a structure as sketched in fig. 2.1. Expressions (2.29) and (2.30) are used for the analytical calculations, together with the values of K_{bb} from table 2.I.

Following the procedure used to obtain (2.27) with subsequent transformation to the situation with an inhomogeneously doped base, we obtain from (2.19) for Y_{box} :

$$Y_{box} = K_{box}\,\beta_\infty\,r^2\,W_b\,W_e\,\frac{S_{ox}}{D_n}, \qquad (2.31)$$

where the factor $\sqrt{(1-\alpha^2)/\alpha^2}$ has been approximated by $\overline{W}_e/W_b$. The constant K_{box} has been determined numerically and is given in table 2.I.

Initially it might seem surprising that Y_{bb} and Y_{box} are proportional to β_∞. This is due to the fact that these parts of the sidewall current are determined by the electron injection into the base, whereas the bottom base current is determined by the hole injection into the emitter. Hence, a large value of β_∞ means a relatively low value of the bottom base current and, therefore, a relatively large sidewall effect.

Numerical calculations with realistic doping profiles show that the sidewall effect on the collector current is fairly small and depends on the detailed doping profiles of the intrinsic and extrinsic base. An indication of the magnitude of this effect can be obtained from:

$$Y_c = K_c \, r \, W_b. \tag{2.32}$$

In the above formula the term of (2.20) between brackets has been replaced by a constant, which is sufficient for our purposes. This is justified by the consideration that for $W_e \simeq W_b$ the major part of that term accounts for the singularity in the electron current density which occurs at the sharp edge of the emitter-base junction of the model given in fig. 2.2. In practical situations this singularity is largely removed. From the value of K_c in table 2.I it can be seen that this sidewall effect is small in practical situations.

2.5 MEASUREMENTS

2.5.1 Ways of measuring the sidewall effects

A problem which confronts us when we want to measure the sidewall effects is that the actual emitter dimensions are usually not accurately known. The only data normally available are mask dimensions or, with the use of SEM photographs, estimates of the emitter dimensions. Unfortunately, the uncertainties in the emitter dimensions are generally not much smaller than the values of Y_b and Y_c, which in most cases are in the order of a few tenths of a micrometre. The interpretation of the measurement results therefore requires careful consideration. In this respect we look at two ways of measuring the sidewall effects. The first one is to measure the individual currents as a function of the emitter dimensions in order to determine Y_b and Y_c [2.8,2.9]. However, the apparent sidewall effects, i.e. the measured values of Y_b and Y_c, strongly depend on the estimated values for P_e and A_e. As shown in appendix 2F, the base current can be written as

$$\frac{I_b}{\overline{A}_e} = J_{p1} \left[1 + \frac{\overline{P}_e}{\overline{A}_e} \left(Y_b + \frac{\Delta H \overline{L}_e + \Delta L \overline{H}_e}{2 \, (\overline{L}_e + \overline{H}_e)} \right) \right], \tag{2.33}$$

where $\overline{L}_e, \overline{H}_e, \overline{A}_e$ and $\overline{P}_e$ are the estimated values of L_e, H_e, A_e and P_e respectively, and ΔL and ΔH are the errors in the estimated emitter

dimensions (see appendix F). As can be seen from (2.33), an underestimation of the emitter dimensions (viz. ΔL and ΔH are positive) will result in a large apparent sidewall effect, while an overestimation of the emitter dimensions will yield a small or even negative sidewall effect. Unless the uncertainties in the emitter dimensions are significantly less than $|Y_b|$ and $|Y_c|$, the above method will therefore be less suitable for the characterization of the sidewall effects.

The second method is to consider the current gain β ($\beta \equiv I_c/I_b$) as a function of the emitter dimensions [2.8,2.10]. As shown in appendix 2F, by means of a first-order expansion the reciprocal current gain can be written as

$$\beta^{-1} = \beta_\infty^{-1} \left[1 + \frac{\overline{P}_e}{\overline{A}_e} (Y_b - Y_c) \right], \tag{2.34}$$

with $\beta_\infty = J_{n1}/J_{p1}$. In appendix 2F it is shown that the above expression is valid in cases where $|Y_b|$, $|Y_c|$ and the errors in the assumed emitter dimensions are much smaller than the emitter dimensions. This method therefore has the advantage that the results depend only on process parameters because errors in the assumed emitter dimensions are approximately cancelled out. It is thus more suitable for process characterization and will be used here. However, the drawback of measuring β or its inverse as a function of $\overline{P}_e/\overline{A}_e$ is that it only yields the difference in sidewall effects $(Y_b - Y_c)$, as can be seen from (2.34).

At this point we introduce the characteristic length

$$H_\beta \equiv 2 (Y_b - Y_c), \tag{2.35}$$

which quantifies the current gain variation with lateral emitter scaling. The physical meaning of H_β is that for a relatively long emitter ($H_e < < L_e$), H_β is a measure for that emitter width at which $\beta/\beta_\infty = 1/2$ in the case where the current gain decreases with decreasing emitter dimensions. Now (2.34) becomes
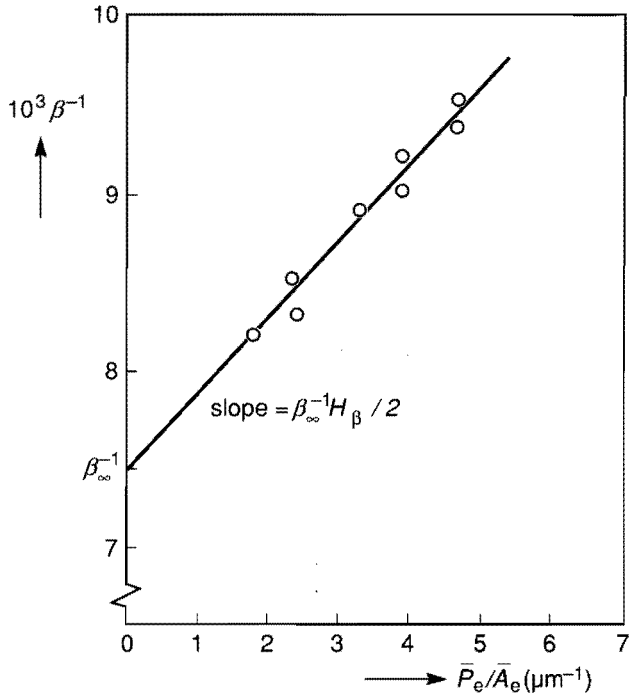
Fig. 2.18. An example of the reciprocal current gain measured as a function of $\overline{P}_e/\overline{A}_e$ (process E, see table 2.II).

$$\beta^{-1} = \beta_\infty^{-1}\left[1 + \frac{\overline{P}_e}{2\overline{A}_e}H_\beta\right]. \tag{2.36}$$

The value of H_β can be determined directly from measuring the slope of β^{-1} versus $\overline{P}_e / \overline{A}_e$, while β_∞ can be measured from the intersection of this line with the vertical axis. The above procedure works quite well in practice. This is shown in fig. 2.18, where the measured reciprocal current gain is plotted versus $\overline{P}_e/\overline{A}_e$ for process E. Since H_β is

approximately independent of the emitter dimensions, this quantity can be used as a figure of merit for the characterization of the lateral scaling properties of the current gain for a certain process. A positive value of H_β means a decrease in the current gain with decreasing emitter dimensions, while a negative value of H_β means an increase in the current gain with decreasing dimensions. The larger the absolute value of H_β, the larger the current gain variation with lateral scaling.

2.5.2 Results

The measurements were performed on five different processes. The relevant parameters are listed in table 2.II. Processes A-D are in-house processes and E is taken from the literature [2.11]. Process A is a modern double-poly process (see fig. 1.2a). B is a fairly conventional process with about the same emitter profile as A, but with a much thicker base. This structure has two versions, i.e. with n-poly on the implanted emitter and with metal on top of the emitter. In the latter case we have investigated the effect of a non-washed emitter. To this end an oxide spacer is applied after the emitter is implanted, so that t_e is larger than in the washed-emitter case. By means of SEM photographs the value of t_e is measured to be approximately 0.22 μm for case B2 (see fig. 2.19). Process C is a self-aligned high-speed process with a silicide base contact (see fig. 1.2c) Three values of t_b are investigated. Process D is a conventional one with rather deep junctions (see fig. 1.1a). The narrow emitters of process E are fabricated by using electron-beam lithography.

  The measured values of β_∞ and H_β (from similar plots as given in fig. 2.18) as well as the calculated values of H_β are listed in table 2.III. For the calculations we have used the measured values of β_∞. The values of the parameters are $r = 0.7$ and $K_{bb} = 0.025$ for device A and 0.08 for devices C (see table 2.I). In fig. 2.20 both the measured and the calculated values of β/β_∞ are plotted as a function of A_e/P_e. So, the value of 0.5 μm on the horizontal axis holds not only for a very long and 1 μm wide emitter but also, for instance, for a 2 μm square emitter. From table 2.III and fig. 2.20 we can observe the following:
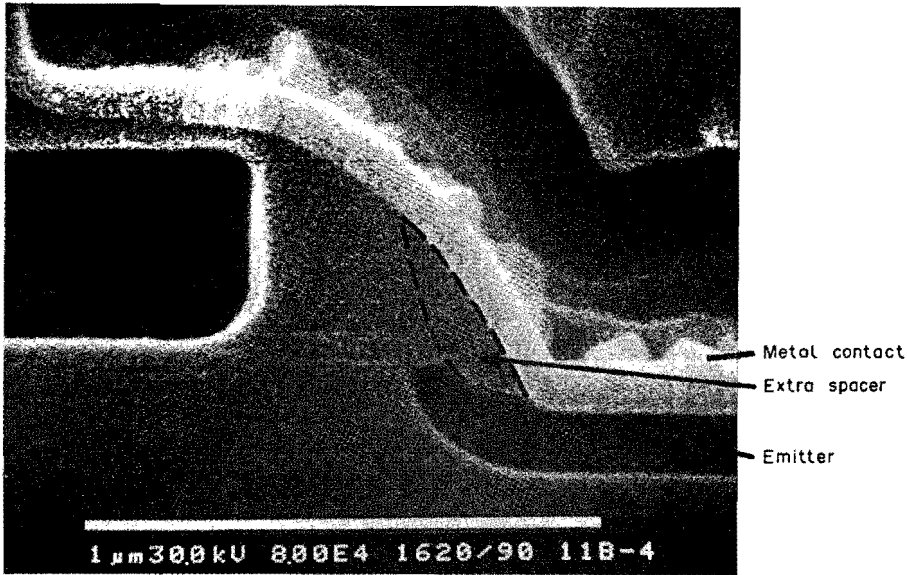
Fig. 2.19. SEM picture of the extra spacer in process B2.

- In the absence of both recombination at the base contact and oxide recombination the current gain increases slightly with decreasing emitter dimensions for a device with a poly emitter contact (process B1) and decreases slightly for a metal emitter contact (process E).
- For the self-aligned processes (A and C) the current gain decreases with decreasing emitter dimensions due to recombination at the base contact. This effect is greater for process C, which has a silicide base contact, than for the double-poly process A. For C the calculated values of H_β are considerably lower than the measured ones (see following discussion). As expected, the sidewall effect increases with decreasing spacing between the emitter sidewall and the base contact.
- A substantial increase in the current gain with decreasing emitter dimensions can be obtained by using an extra spacer to reduce the sidewall hole current (process B2).
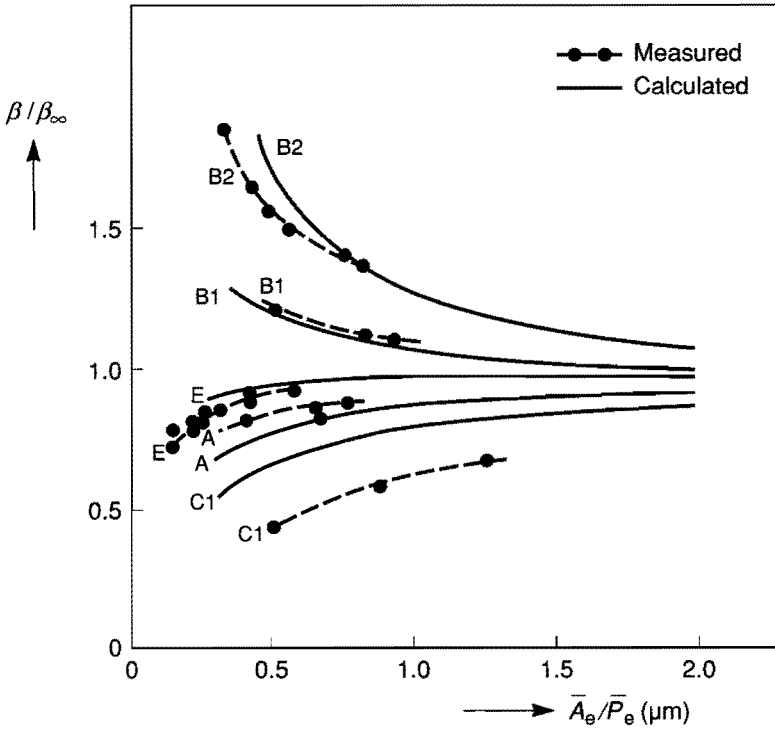
Fig. 2.20. Measured and calculated values of β/β_∞ versus the area-to-perimeter ratio of the emitter.

- For process D the measured value $H_\beta = 0.8\ \mu m$ is much higher than the calculated one. This difference is most likely due to recombination at the Si_3N_4 passivated interface between SiO_2 and $< 1\ 1\ 1 >$ oriented silicon. This will be discussed at a later stage.

Sidewall effects on the d.c. base and collector currents   *51*

Table 2.II. Different processes and their parameters (see also fig. 2.1).

process	W_b (μm)	W_e (μm)	t_e (μm)	t_b (μm)	contact	
					emit.	base
A	0.15	0.15	w	0.15	p	p
B1	0.35	0.15	w	> 1.0	p	m
B2	0.35	0.10	0.22	> 1.0	m	m
C1	0.15	0.10	w	0.15	p	m
C2	0.15	0.10	w	0.25	p	m
C3	0.15	0.10	w	0.35	p	m
D	0.30	0.35	w	> 1.0	m	m
E	0.15	0.20	w	> 1.0	m	m
F	0.30	0.35	w	> 1.0	m	m

w: washed emitter   p: poly contact   m: metal contact

Table 2.III. Results

process	measured		calculated
	β_∞	H_β (μm)	H_β (μm)
A	270	0.15	0.25
B1	100	-0.2	-0.15
B2	25	-0.3	-0.4
C1	165	1.3	0.5
C2	160	0.7	0.15
C3	160	0.4	0.0
D	90	0.8	0.15
E	130	0.1	0.05
F	90	0.5	0.15

2.6 DISCUSSION AND CONCLUSIONS

From the analysis in sections 2.2 and 2.3 it follows that, in principle, a sidewall effect exists on both the base and collector current. It is shown that not only do the sidewalls contribute to I_b and I_c simply by the current flowing through the sidewall junction but, more generally, that both currents are affected by the two-dimensional nature of the carrier injection in the regions near the sidewall. For the sidewall effect on the base current we can distinguish three possible contributions, i.e. hole injection into the emitter, recombination at the base contact and recombination at the $Si - SiO_2$ interface.

It is shown that for shallow emitters ($W_e < L_p$) the hole current through the emitter sidewall is mainly determined by surface recombination near the contact edge and therefore strongly depends on the surface recombination velocity and the distance between the contact edge and sidewall junction. In fact, a part is played not only by this distance but also by the exact shape of the two-dimensional doping profile. Since this two-dimensional doping profile is not well-known, it is impossible to determine this sidewall effect accurately. From both the analytical and numerical calculations it follows that the net sidewall effect on the hole injection into the emitter is proportional to the emitter depth but, rather surprisingly, it is small. From numerical calculations with $t_e / W_e = 0.5$ and error-type lateral profiles it is found that in the absence of both recombination at the base contact and oxide recombination the current gain increases slightly with decreasing emitter dimensions for a device with a poly emitter contact and decreases slightly for a metal emitter contact. This is confirmed by the measurements on devices B1 and E, where the other possible contributions to the sidewall effect on the current gain are expected to be small.

For self-aligned processes the current gain decreases with decreasing emitter dimensions due to recombination at the base contact. From calculations it follows that this effect is greater for a silicide base contact than for a polysilicon base contact because of the higher surface recombination velocity of a metal contact. This is in agreement with the measurement results on processes A (double -poly) and C (silicide base contact). For process C the calculated values of H_β are consider-

ably lower than the measured ones. This may be due to lateral encroachment of the base-link implantation into the intrinsic base region. This base-link implantation, which connects the extrinsic base with the intrinsic base, may cause an increased base dopant concentration at the edge of the intrinsic base and, hence, a negative sidewall effect on the collector current [2.12]. Another possibility might be that the surface recombination velocity at the silicide contact is higher than the assumed value of $10^4 m/s$. This is used because reported measured values of the surface recombination velocity at a metal contact usually lie around this value. However, no measurements on the surface recombination velocity at silicide contacts are known. It has been found theoretically that the recombination current at the base contact depends exponentially on the spacing between the emitter sidewall and the base contact.

The numerical analysis of the recombination at the silicon-oxide interface shows that this effect is proportional to the recombination velocity at this interface and is unimportant for the ideal currents if $S_{ox} << 100 \ m/s$. The main problem, however, is that the value of the interface recombination velocity is usually unknown and, moreover, may depend not only on the interface-state density but also on the detailed potential profile near the interface [2.13]. This potential profile, for instance due to oxide charge, can vary as a function of time (e.g. as a result of heat treatments and stressing). For device D the difference between the measured and calculated sidewall effect is most likely due to recombination at the Si_3N_4 passivated interface between SiO_2 and $< 1 \ 1 \ 1 >$ oriented silicon. Such an interface has a high interface state density [2.2,2.14]. If we attribute the difference between the measured and calculated values of H_β to interface recombination and use the expression for Y_{box} as given by (2.31), we arrive at a value of $S_{ox} = 2.10^2 \ m/s$. This value is quite high, though not unlikely for a measured interface state density of about $10^{16} m^{-2}$ for $< 1 \ 0 \ 0 >$ oriented $Si_3N_4 -$ passivated interfaces [2.15]. In fact, the measured value $H_\beta = 0.8 \ \mu m$ agrees quite well with measurements on similar devices as given in [2.2]. From the reported A_e/P_e - dependence of β, a value of $H_\beta = 0.6 - 0.8 \ \mu m$ can be deduced for an emitter depth of 0.2 μm. A strong change of H_β during accelerated life tests is also reported in [2.2].

The collector current mainly flows through the bottom part of the emitter-base junction and, unlike the base current, the sidewall effect on I_c in our analytical model is determined only by the geometry of the device. In more realistic structures this sidewall effect is reduced, when compared with the analytical model, because of the non-uniform doping of the extrinsic base. As indicated above, sidewall effects on the collector current can also be due to a base-link implantation which connects the extrinsic base with the intrinsic base. This implantation may cause an increased base dopant concentration at the edge of the intrinsic base and, hence, a negative sidewall effect on the collector current [2.12].

For very narrow emitters (viz. $H_e < 2L_p \tanh (W_e/L_p)$) an interesting phenomenon occurs, i.e. the hole current through the bottom of the emitter decreases rapidly with decreasing emitter width, while the hole current through the sidewall remains much more constant. In this case the base current does not scale linearly with the emitter width. In such a situation it is also possible that the major part of the base current will flow through the sidewall or recombines in the extrinsic base. This can have two implications for device design considerations. The first one is that reduction of the emitter width in order to reduce the d.c. base resistance parameter is not very effective in that case. The second one is related to the coupling between base resistance and collector current. In conventional devices with a relatively large emitter area the d.c. base resistance parameter is strongly coupled with the intrinsic base thickness and doping profile (more generally: with the base Gummel number) and is therefore coupled with the current gain and base transit time. The reason for this is that the major part of both the base and collector current flows through the intrinsic base. However, from the above analysis it follows that at very small emitter widths the base resistance parameter will mainly be determined by the doping profile of the extrinsic base, while the collector current is mainly determined by the thickness and doping profile of the intrinsic base. This implies that the base resistance parameter and the collector current become increasingly decoupled when the emitter width decreases.

With respect to the measurement of these sidewall effects we have seen that, unless the emitter dimensions are known to a high accuracy,

it is almost impossible to measure the sidewall effects on I_b and I_c separately. When Y_b and Y_c are measured from the geometry dependence of the corresponding current, the result largely depends on the error in the assumed emitter dimensions. When the sidewall effect on the currents is determined from the geometry dependence of the current gain, only the difference between Y_b and Y_c is measured. The advantage of the latter method, however, is that the result is approximately independent of the errors in the assumed emitter dimensions. A quantity H_β is introduced as a figure of merit for the current gain variation with lateral emitter scaling. This quantity can be measured from the dependence of the current gain on the lateral emitter dimensions.

REFERENCES

2.1   M.H. El-Diwany, M.P. Brassington and P. Tuntasood, IEEE Electron Device Lett., EDL-9, 247 (1988)

2.2   N. Shiono, Jap. J. Appl. Phys., 18, 1097 (1979)

2.3   H. Sagan, Boundary and Eigenvalue Problems in Mathematical Physics, New York: Wiley, p. 105 (1961)

2.4   M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, New York: Dover Publications, p. 260 (1970)

2.5   G.A.M. Hurkx, IEEE Trans. Electron Devices, ED-34, 1939 (1987)

2.6   A.S. Grove, Physics and Technology of Semiconductor Devices, New York: Wiley, p. 136 (1967)

2.7   L. Dreyfus, Archiv fur Elektrotechnik, XIII Band, p. 123 (1924)

2.8   H.M. Rein, Solid-St. Electron., 27, p. 625 (1984)

2.9   D.P. Verret and J. E. Brighton, IEEE Trans. Electron Devices, ED-34, p. 2297 (1987)

2.10 A. Gover and A. Gaash, Solid-St. Elecron., 19, p.125 (1976)

2.11 Y. Tamaki, F. Murai, Y. Kawamoto, K. Uehara, A. Hayasaki and A. Anzai, IEEE Electron Device Lett., EDL-7, 425 (1986)

2.12 G.P. Li, C.T. Chuang, T.C. Chen and T.H. Ning, IEDM Tech. Dig., p.175 (1987)

2.13 M.Y. Ghannam, R.P. Mertens, R.F. de Keersmaecker and R.J. van Overstraeten, IEEE Trans. Electron Devices, ED-32, p. 1264 (1985)

2.14 B.E. Deal, E.L. MacKenna and P.L. Castro, J. Electrochem. Soc., 116, p. 997 (1969)

2.15 A. Linssen, private communication

2.16 P.F. Byrd and M.D. Friedman, Handbook of Elliptic Integrals for Engeneers and Physicists, Berlin: Springer, p. 11 (1954)

2.17 E. Weber, Electromagnetic Fields, New York: Wiley, p.301 (1950)

2.18 W. Groebner and N. Hofreiter, Integraltafel, Erster Teil, Wien: Springer, p. 31 (1961)

2.19 W. Groebner and N. Hofreiter, Integraltafel, Zweiter Teil, Wien: Springer, p. 123 (1961)

APPENDIX 2A

Equation (2.2), together with the boundary conditions as denoted in fig. 2.2 and eq. (2.3), can be solved as follows: Consider only the right half of the structure in fig. 2.2. The additional boundary condition is then $\partial \bar{p}/\partial x = 0$ at $x = 0$. Let us first assume that the excess concentration at the surface $(y = W_e)$ is given by the function $\bar{p}_j \gamma(x)$. This function will be determined later on. The problem can now be solved by a standard separation technique. The result is:

$$\bar{p}(x, y) = \bar{p}_j \left\{ \sum_{n=0}^{\infty} \left[\frac{A_n \sinh(w [W_e - y]) + B_n \sinh (wy)}{\sinh (wW_e)} \right] \cos (mx) \right.$$

$$\left. + \frac{\cosh \left(\frac{x}{L_p} \right)}{\cosh \left(\frac{H_e}{2L_p} \right)} \right\}, \tag{2A.1}$$

with

$$m = (n + \frac{1}{2}) \frac{2\pi}{H_e}, \tag{2A.2}$$

$$w = (L_p^{-2} + m^2)^{1/2}. \tag{2A.3}$$

The coefficients A_n and B_n are:

$$A_n = \frac{4}{H_e} \int_0^{He/2} \left[1 - \frac{\cosh \left(\frac{x}{L_p} \right)}{\cosh \left(\frac{H_e}{2L_p} \right)} \right] \cos (mx) \, dx$$

$$= \frac{4}{H_e} \frac{(-1)^n}{L_p^2 \, m \, (L_p^{-2} + m^2)}, \qquad (2A.4)$$

$$B_n = \frac{4}{H_e} \int_0^{He/2} \left[\gamma(x) - \frac{\cosh \left(\dfrac{x}{L_p}\right)}{\cosh \left(\dfrac{H_e}{2L_p}\right)} \right] \cos \, (mx) \, dx. \qquad (2A.5)$$

Using (2A.1) with $y = W_e$, the function $\gamma(x)$ can be written as

$$\gamma(x) = \sum_{n=0}^{\infty} B_n \cos \, (mx) + \frac{\cosh \left(\dfrac{x}{L_p}\right)}{\cosh \left(\dfrac{H_e}{2L_p}\right)}. \qquad (2A.6)$$

To determine B_n consider the dimensionless surface recombination σ to be a (step-) function of x, which equals zero for $H_c/2 < x < H_e/2$. The product $\gamma(x) \, \sigma(x)$ can be expanded in

$$\gamma(x) \, \sigma(x) = \sum_{n=0}^{\infty} E_n \cos \, (mx). \qquad (2A.7)$$

Substitution of (2A.6) in (2A.7) gives

$$E_n = \frac{4\sigma}{H_e} \int_0^{Hc/2} \left[\frac{\cosh \left(\dfrac{x}{L_p}\right)}{\cosh \left(\dfrac{H_e}{2L_p}\right)} + \sum_{j=0}^{\infty} B_j \cos \, (vx) \right] \cos \, (mx) \, dx, \qquad (2A.8)$$

where

Sidewall effects on the d.c. base and collector currents 59

$$v = \left(j + \frac{1}{2} \right) \frac{2\pi}{H_e} . \qquad (2A.9)$$

Substitution of (2A.7) and (2A.8) into the right-hand side of (2.3) and of the derivative of (2A.1) into the left-hand side of (2.3) leads, after some manipulation, to the following set of linear equations for B_n:

$$\sum_{j=0}^{\infty} M_{nj} B_j = Z_n, \qquad (2A.10)$$

with

$$M_{nj} = \frac{w W_e}{\tanh(w W_e)} + \frac{\sigma W_e}{H_e L_p} \left\{ \frac{\sin (m H_c)}{m} + H_c \right\}, \quad n = j$$

$$= \frac{2\sigma W_e}{H_e L_p} \left\{ \frac{\sin \left[\frac{1}{2} H_c (m + v) \right]}{m + v} + \frac{\sin \left[\frac{1}{2} H_c (m - v) \right]}{m - v} \right\}, \quad n \neq j$$

$$(2A.11)$$

$$Z_n = \frac{4 w W_e}{H_e \sinh (w W_e)} \frac{(-1)^n}{L_p^2 m (L_p^{-2} + m^2)}$$

$$- \left\{ \frac{4\sigma W_e}{L_p H_e \cosh \left(\frac{H_e}{2 L_p} \right) (L_p^{-2} + m^2)} \right.$$

$$\times \left[m \sin \left(\frac{m H_c}{2} \right) \cosh \left(\frac{H_c}{2 L_p} \right) + L_p^{-1} \cos \left(\frac{m H_c}{2} \right) \sinh \left(\frac{H_c}{2 L_p} \right) \right] \right\}.$$

$$(2A.12)$$

This set of equations can be solved by a standard technique, taking only a finite number of equations into account.

APPENDIX 2B

By using the following Schwarz-Christoffel transformation

$$\frac{dz}{dw} = \frac{C}{\sqrt{(w^2 - 1)(w^2 - k^{-2})}}, \qquad (2B.1)$$

the emitter in fig. 2.9a is mapped onto the upper half plane in fig. 2.9b. By integration of the above expression in the w − plane along the u − axis from 0 to 1 we obtain

$$H_e = -2CkK(k), \qquad (2B.2)$$

where $K(k)$ is the complete elliptic integral of the first kind. Similarly, by integration from $u = 1$ to $u = 1/k$ and using (2B.2), we obtain

$$\frac{H_e}{2W_e} = \frac{K(k)}{K(k')}. \qquad (2B.3)$$

Finally, integration from $u = 0$ to $u = v$ yields

$$v = \mathrm{sn}\left(\frac{H_c}{H_e} K(k), k\right), \qquad (2B.4)$$

where $\mathrm{sn}(\zeta, k)$ is a Jacobian elliptic function [2.16].

Now we consider the geometry in the complex t − plane as given in fig. 2B.1. With the use of the transformation

$$\frac{dt}{dw} = \frac{D}{\sqrt{(w^2 - 1)(w^2 - v^2)}}, \qquad (2B.5)$$

the geometry of fig. 2B.1 is mapped onto the upper half plane in fig. 2.9b. By integrating the above expression in the $t -$ plane axis from $t = 1 + jc$ to $t = jc$ we obtain

$$D = \frac{-1}{K(v)}.$$ (2B.6)

Integration in the $t -$ plane from $t = -1 + jc$ to $t = -1$ yields

$$c = \frac{K(v')}{K(v)},$$ (2B.7)

with $v' = \sqrt{1 - v^2}$. Now we have, indirectly, mapped the original geometry of fig. 2.9a onto the geometry as given in fig. 2B.1. It is a general property of conformal transformations that the solution of the Laplace equation in one plane is also a solution of this equation in the other plane [2.17]. Because we have obtained a quasi one-dimensional problem in the $t -$ plane the solution can easily be found to be

$$P = \frac{\bar{p}_j}{c} (1 + jt),$$ (2B.8)

where $P(r,s) \equiv \bar{p}(r,s) + j\,\theta(r,s)$ is the complex hole concentration.

The hole current I_{be} can be determined as follows: The current density in the $z -$ plane is :

$$J = J_x + jJ_y = - qD_p \left(\frac{\partial \bar{p}}{\partial x} + j \frac{\partial \bar{p}}{\partial y} \right).$$ (2B.9)

Using the Cauchy-Riemann equations it can be shown that [2.17]

$$J = - qD_p \left(\frac{dP}{dz} \right)^{cc} = - qD_p \left(\frac{dP\,dt}{dt\,dz} \right)^{cc},$$ (2B.10)

where the superscript cc denotes the complex conjugate of the corresponding number. The hole current is
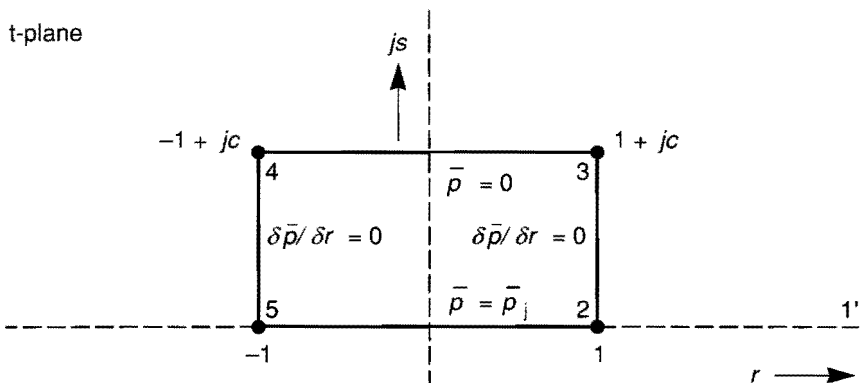
Fig. 2B.1. Contours of the emitter region in the complex t -plane.

$$I_{be} = \int_{-Hc/2}^{Hc/2} J_y \, dx = \int_{x=-Hc/2}^{x=Hc/2} |J| \, dz. \qquad (2B.11)$$

By substitution of (2B.10) in (2B.11) we obtain

$$I_{be} = qD_p \int_{t=-1+jc}^{t=1+jc} \left| \left(\frac{dP}{dt} \right)^{cc} \right| dt = \frac{2qD_p \bar{p}_j}{c}. \qquad (2B.12)$$

When $H_e >> W_e$, both $v \to 1$ and $k \to 1$. In [2.16] it can be found that in this case

$$\frac{K(v)}{K(v')} \simeq \frac{1}{\pi} \ln \left(\frac{16}{1 - v^2} \right). \qquad (2B.13)$$

Substitution of (2B.13) in (2B.12), using (2B.4) and $sn^2(\zeta, k) = 1 - cn^2(\zeta, k)$ gives

$$I_{be} = \frac{2qD_p p_j}{\pi} \ln\left(\frac{16}{cn^2(\zeta,k)} \right), \qquad (2B.14)$$

with

$$\zeta = \frac{H_c}{H_e} K(k). \qquad (2B.15)$$

Using

$$cn(\zeta,k) \simeq \frac{1}{\cosh \zeta} - \frac{k'^2}{4} \sinh \zeta, \qquad (2B.16)$$

$$K(k') \simeq 2 \ln 2 - \ln k', \qquad (2B.17)$$

and using (2B.13) for $\dfrac{K(k)}{K(k')}$ gives

$$cn(\zeta,k) \simeq 4 \, \exp\left(\frac{-\pi H_e}{4W_e} \right) \sinh\left(\frac{\pi t_e}{2W_e} \right) , \qquad (2B.18)$$

for $k \to 1$. After some manipulation (2.15) is obtained.

APPENDIX 2C

To solve the Laplace equation together with the boundary conditions denoted in fig. 2.11a, we use the following Schwarz-Christoffel transformation:

$$\frac{dz}{dw} = -\frac{W_b}{\pi(w+k)} \sqrt{\frac{w-t}{w-1}}. \qquad (2C.1)$$

By this transformation the geometry in the complex z-plane (fig. 2.11a) is mapped onto the upper half w-plane (fig. 2.11b). Integrating (2C.1) from $u = -1$ to $u = 1$ gives [2.18]

$$t_b = -\frac{W_b}{\pi}\left[\ln\left(\frac{s_0+1}{s_0-1}\right) - \alpha^{-1}\ln\left(\frac{v_0+1}{v_0-1}\right)\right], \qquad (2C.2)$$

with

$$s_0 = \sqrt{\frac{t+1}{2}} \qquad (2C.3)$$

and

$$v_0 = \alpha s_0. \qquad (2C.4)$$

A second relation for determining the unknowns t and k is obtained from the distance between 2' and 2'' in the z-plane [2.17]:

$$\frac{t+k}{1+k} = \alpha^{-2}. \qquad (2C.5)$$

We now put $k = 1 + \gamma$. For $t_b > (W_e + W_b)$ it is found that $\gamma \ll 1$. In that case v_0 becomes

$$v_0 \simeq 1 + \frac{\gamma}{4}\left(\frac{t-1}{t+1}\right) \qquad (2C.6)$$

and

$$s_0 \simeq \alpha^{-1}. \qquad (2C.7)$$

Substitution of (2C.6) and (2C.7) in (2C.2) finally gives

Sidewall effects on the d.c. base and collector currents   *65*

$$\gamma \simeq 8 \left(\frac{t+1}{t-1} \right) \exp \left[\frac{-\pi t_b}{W_e + W_b} - \alpha \ln \left(\frac{1+\alpha}{1-\alpha} \right) \right]. \quad (2C.8)$$

The complex concentration in the w − plane can be found in standard textbooks [2.17] as

$$N = n_j \left(\frac{j}{\pi} \text{arccosh} (w) + 1 \right). \quad (2C.9)$$

By using a procedure similar to that described in appendix 2B, differentiation and subsequent integration of (2C.9) gives for the current at the base contact

$$I_{bb} = \frac{q D_n n_j}{\pi} \text{arccosh} (1 + \gamma). \quad (2C.10)$$

Using

$$\text{arccosh} (1 + \gamma) = \ln \left[1 + \gamma + \sqrt{2\gamma + \gamma^2} \right] \simeq \sqrt{2\gamma} \quad (2C.11)$$

and

$$\frac{t+1}{t-1} \simeq \frac{1}{1-\alpha^2} \quad (2C.12)$$

for $\gamma << 1$, we obtain (2.17).

APPENDIX 2D

To evaluate the integral of (2.18) we assume no base contact, so $k = 1$. The total electron concentration at the surface can now be written as

$$n_{tot} = \int_{He/2}^{\infty} n (x, W_e) \, dx. \quad (2D.1)$$

By the substitution of (2C.1) and (2C.9) in (2D.1) we obtain

$$n_{tot} = \frac{n_j \, W_b}{\pi} \int_1^{-1} \left(1 - \frac{1}{\pi} \arccos{(u)} \right) \sqrt{\frac{t - u}{1 + u}} \quad \frac{du}{\sqrt{1 - u^2}} \, . \qquad (2D.2)$$

If we put $u = \cos{(\xi)}$ then

$$n_{tot} = \frac{n_j \, W_b}{\pi} \int_0^{\pi} \left(1 - \frac{\xi}{\pi} \right) \sqrt{\frac{t - \cos{(\xi)}}{1 + \cos{(\xi)}}} \, d\xi. \qquad (2D.3)$$

By using the identities $\cos{(\xi)} = 1 - 2\sin^2{(\xi/2)}$ and $\cos{(\xi)} = -\cos{(\pi - \xi)}$ we find

$$n_{tot} = \frac{\sqrt{2} \, n_j \, W_b}{\pi^2} \int_0^{\pi} \sqrt{t - \cos{(\xi)}} \, \frac{(\pi - \xi)/2}{\sin{(\pi - \xi/2)}} \, d\xi. \qquad (2D.4)$$

If we replace $\sqrt{t - \cos{(\xi)}}$ by $\sqrt{t - 1}$, which is justified by the fact that $t >> 1$, and put $(\pi - \xi)/2 = \theta$ we obtain

$$n_{tot} \simeq \frac{2\sqrt{2t - 2} \, n_j \, W_b}{\pi^2} \int_0^{\pi/2} \frac{\theta}{\sin{(\theta)}} \, d\theta. \qquad (2D.5)$$

In [2.19] it can be found that

$$\int_0^{\pi/2} \frac{\theta}{\sin{(\theta)}} \, d\theta = 2G. \qquad (2D.6)$$

Substitution of this result in (2D.5) gives (2.19).

APPENDIX 2E

Differentiation and subsequent integration of (2C.9) gives for the collector current (per unit length in the third dimension):

$$I_c = \frac{q \, D_n \, n_j}{\pi} \, \text{arccosh} \, (u_m),$$ (2E.1)

where the value of u_m is given by the relation :

$$\frac{H_e}{2} = \frac{W_b}{\pi} \left\{ \ln \left[\frac{\sqrt{u_m - t} + \sqrt{u_m - 1}}{\sqrt{u_m - 1} - \sqrt{u_m - t}} \right] \right.$$

$$\left. - \frac{1}{\alpha} \ln \left[\frac{\alpha\sqrt{u_m - t} + \sqrt{u_m - 1}}{\sqrt{u_m - 1} - \alpha\sqrt{u_m - t}} \right] \right\},$$ (2E.2)

with

$$\alpha = \frac{W_b}{W_b + W_e}.$$ (2E.3)

The above relation is obtained by integrating (2C.1) along the $u -$ axis from $u = t$ to $u = u_m$. For $H_e > 2W_b$, u_m turns out to be much larger than t, so (2E.2) can be significantly simplified by using the following series expansions:

$$\sqrt{1 - \frac{t}{u_m}} \simeq 1 - \frac{t}{2u_m}$$ (2E.4)

and

$$\sqrt{1 - \frac{1}{u_m}} \simeq 1 - \frac{1}{2u_m}.$$ (2E.5)

After substitution of the resulting expression for u_m in (2E.1) we obtain (2.20).

APPENDIX 2F

In principle, the sidewall effects Y_b and Y_c can be obtained by using of equations (2.22) and (2.24). However, the problem is that the actual values of A_e ($A_e = H_e L_e$) and P_e ($P_e = 2 (H_e + L_e)$) are not known. All we know are (at a certain bias condition) I_b and I_c (and therefore β) for a number of transistors with different emitter dimensions. We denote the estimated values of L_e, H_e, A_e and P_e by $\overline{L}_e$, $\overline{H}_e$, $\overline{A}_e$ and $\overline{P}_e$, respectively. Now we consider two ways of measuring the sidewall effects, as given in section 2.5.1:

1) Measuring the individual values of Y_b and Y_c . We denote the measured values of Y_b by $\overline{Y}_b$. The errors in the estimated emitter dimensions are given by:

$$\Delta L = L_e - \overline{L}_e, \qquad (2F.1)$$

$$\Delta H = H_e - \overline{H}_e. \qquad (2F.2)$$

Eq. (2.22) now becomes

$$I_b = J_{p1} [\overline{H}_e\overline{L}_e + \Delta L \overline{H}_e + \Delta H \overline{L}_e + \Delta L \Delta H + 2 (\overline{L}_e + \overline{H}_e + \Delta L + \Delta H)\, Y_b],$$

$$(2F.3)$$

or, to first order (this means that products such as $\Delta L \Delta H$ and $\Delta L Y_b$ are neglected)

$$I_b = J_{p1} [\overline{H}_e\overline{L}_e + \Delta L \overline{H}_e + \Delta H \overline{L}_e + 2 (\overline{L}_e + \overline{H}_e)\, Y_b]. \qquad (2F.4)$$

Using $\overline{A}_e = \overline{H}_e \overline{L}_e$ and $\overline{P}_e = 2 (\overline{H}_e + \overline{L}_e)$, (2F.4) can be written as

$$\frac{I_b}{\overline{A}_e} = J_{p1} \left[1 + \frac{\overline{P}_e}{\overline{A}_e} \left(Y_b + \frac{\Delta H \overline{L}_e + \Delta L \overline{H}_e}{2\,(\overline{L}_e + \overline{H}_e)} \right) \right] \qquad (2F.5)$$

From (2F.5) it can be seen that when there is a significant spreading of ΔH and ΔL between the different emitter geometries, a straight line is generally not obtained when $I_b/\overline{A}_e$ is plotted vs. the estimated ratio $\overline{P}_e/\overline{A}_e$. When ΔH and ΔL are approximately independent of the emitter geometry, the measured value of Y_b can be obtained from the slope of the line. This gives

$$\overline{Y}_b = Y_b + \left(\frac{\Delta H \overline{L}_e + \Delta L \overline{H}_e}{2\,(\overline{L}_e + \overline{H}_e)} \right)_{averaged} \qquad (2F.6)$$

So, the measured sidewall effect on I_b is strongly dependent on the errors in the estimated emitter dimensions. For $\overline{Y}_c$ an analogous expression can be derived.

2) When we consider the current gain, substitution of (2F.3) and the equivalent expression for I_c in $\beta^{-1} = I_b/I_c$ gives, after a first-order expansion of I_c^{-1},

$$\beta^{-1} = \beta_\infty^{-1} \left[1 + \frac{2\,(\overline{L}_e + \overline{H}_e + \Delta L + \Delta H)}{\overline{L}_e \overline{H}_e + \overline{L}_e \Delta H + \overline{H}_e \Delta L + \Delta L \Delta H}\,(Y_b - Y_c) \right]. \qquad (2F.7)$$

To first order (same approximation as in (2F.4)) this becomes

$$\beta^{-1} = \beta_\infty^{-1} \left[1 + \frac{\overline{P}_e}{\overline{A}_e}\,(Y_b - Y_c) \right]. \qquad (2F.8)$$

So, when plotting β^{-1} versus $\overline{P}_e/\overline{A}_e$ we obtain a straight line with a slope of $\beta_\infty^{-1}(Y_b - Y_c)$. This means that the measured value of H_β (see eq. (2.35)), i.e. $\overline{H}_\beta$, is to first order equal to $2(Y_b - Y_c)$, which is the actual value of H_β. This shows that the measured value of H_β is indeed much

less sensitive to errors in the assumed emitter dimensions than when Y_b and Y_c are measured separately.

Chapter 3

TUNNELLING EFFECTS IN P-N JUNCTIONS

3.1 INTRODUCTION

In this chapter we investigate various aspects of the influence of tunnelling processes on the reverse and non-ideal forward characteristics of a heavily-doped diode. The processes considered here are usually referred to as tunnelling, because unlike the ordinary generation-recombination mechanisms such as the Shockley-Read-Hall and Auger processes, these processes involve transitions of electrons and holes between spatially separated states in the conduction and valence bands. Two different mechanisms can be distinguished, i.e. band-to-band tunnelling and tunnelling via impurity states in the gap. The former mechanism gives an I-V curve according to the solid line in fig. 3.1 and is important in the reverse-bias regime and, at extremely high dopant concentrations, in the low forward-bias regime [3.1]. The second mechanism gives an I-V curve according to the dashed line in fig. 3.1 and is most important in the low and medium forward-bias regime. This current is called excess current because it is in excess of the normal diode current [3.2] . As pointed out in chapter one, in bipolar transistors these effects influence the non-ideal forward and reverse characteristics of the emitter-base junction. Due to the proximity of the base contact to the edge of the emitter-base junction these effects are especially important in modern self-aligned structures such as those sketched in fig. 1.2. Band-to-band tunnelling determines the emitter-base leakage current, while the excess current contributes to the non-ideal forward base current and, therefore, adversely affects the current gain at low and medium bias conditions. The primary goal of the work described in this chapter is to obtain model descriptions of these tunnelling effects which are suitable for implementation in a numerical device simulator and which can serve as a basis for a compact model description.

In section 3.2 we investigate the feasibility of the well-known expression for band-to-band tunnelling (e.g. see [3.1]) as a model for
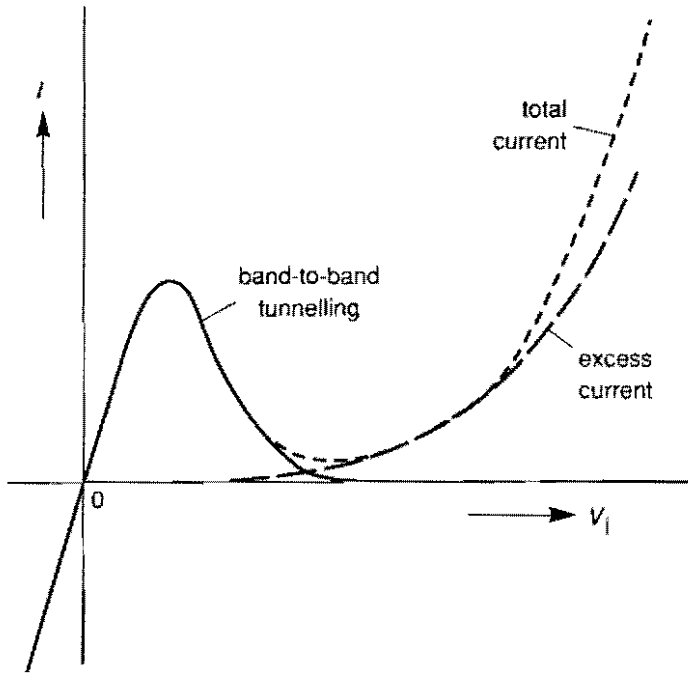
Fig. 3.1. I-V characteristic of a heavily-doped diode. The total current, including the normal ideal current, is denoted by the dotted line.

band-to-band tunnelling in reverse-biased junctions. Section 3.3 describes a model for the recombination rate, taking account of trap-assisted tunnelling. This model is based on a simple quantum-mechanical description of a p-n junction using the so-called envelope-function approach. The physical basis for a compact model describing the I-V characteristics of reverse-biased junctions is given in section 3.4. Besides band-to-band tunnelling, this model also includes avalanche generation and tunnelling-enhanced SRH generation.

We now qualitatively discuss the tunnelling processes at zero temperature, using the simplified band structures of a p-n junction as sketched in fig. 3.2. The semiconductor material is assumed to be very

highly doped so that the quasi-Fermi levels lie within the bands. To understand the solid I-V curve as given in fig. 3.1 we consider the tunnelling current per unit energy as a product of three quantities, i.e.

- The density of occupied states $D_i(E)$ on the side from which an electron tunnels.
- The density of unoccupied states $D_f(E)$ on the side to which an electron tunnels.
- The probability per unit time of an electron with that energy tunnelling through the gap.

At zero bias voltage (fig. 3.2a) the tunnelling current obviously equals zero because the product $D_i(E)D_f(E)$ is zero for all energies. Fig. 3.2b shows an electron tunnelling from the valence band at the p-side into the conduction band at the n-side when a reverse bias is applied. The increase in the tunnelling current with increased reversed bias voltage can be understood in terms of a tunnelling distance. This tunnelling distance, which is equal to E_g/qF, decreases with increasing reverse bias due to the increase in the electric field F. When a small forward bias is applied (fig. 3.2c) the tunnelling probability decreases monotonically because the electric field decreases monotonically. The peak in the I-V characteristic, however, occurs because the energy interval in which both unoccupied states in the valence band at the p-side and occupied states in the conduction band at the n-side are present, shows a maximum with increasing forward bias. In other words, the product $D_i(E)D_f(E)$ integrated over all energies shows a maximum. Notice that this peak can only occur in very highly doped samples where the quasi-Fermi levels lie within the bands. The tunnelling process can be either direct or phonon-assisted. Direct tunnelling involves purely elastic transitions and dominates in direct semiconductors such as GaAs and InP. Phonon-assisted tunnelling occurs in indirect semiconductors such as Si and Ge where the valence-band maximum and conduction-band minimum do not occur at the same point in $\underline{k}$-space.

The excess current, which is experimentally observed at high doping levels or at low temperatures, occurs in a situation where elastic tunnelling is impossible. This phenomenon is thought to be due to tunnelling via impurity states in the gap [3.2]. It is therefore also called trap-assisted tunnelling. The basic process is a tunnelling transition of
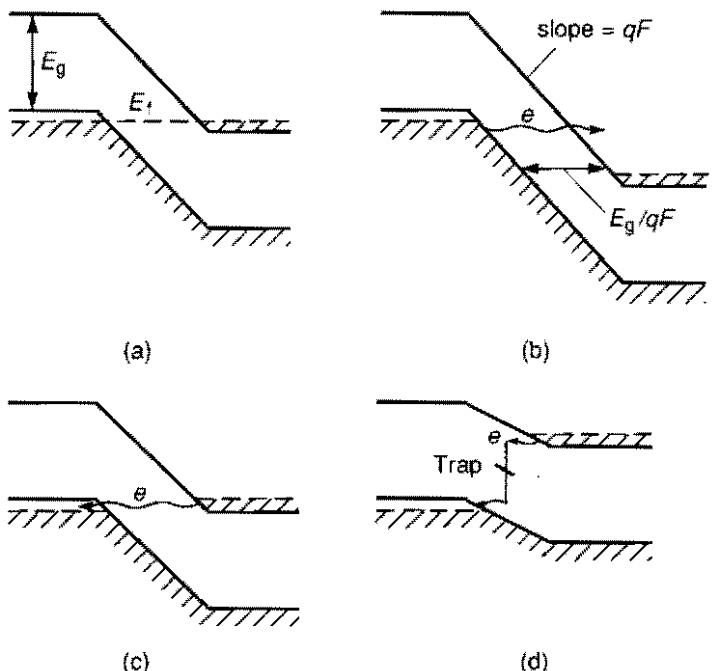
Fig. 3.2. Simplified energy-band diagrams of a heavily-doped diode at equilibrium (a), at reverse-biased condition (b), forward-biased such that the peak current is obtained (c) and forward-biased such that elastic tunnelling is impossible (d).

an electron from the conduction band to an empty impurity state, followed by a transition from an occupied impurity state to an empty state in the valence band (see fig. 3.2d). For holes the reverse process occurs. This process can be considered equivalent to the ordinary Shockley-Read-Hall process via deep states. The difference is the character of the states from which an electron makes a transition to a trap. In the ordinary SRH process the initial state is a non-localized (Bloch) state in the conduction band, while for the tunnelling process the initial state is a decaying state in the gap.

3.2 BAND-TO-BAND TUNNELLING IN REVERSE-BIASED JUNCTIONS

3.2.1 Introduction

To outline the theoretical concepts given in the literature, we start with the expression for the net tunnelling current density in the energy range $[E, E + dE]$, allowing tunnelling in either direction. This expression, as found in standard textbooks, is [3.1]

$$dJ_{bbt} \sim (f_n(E) - f_p(E)) T_t N_c(E) N_v(E) dE, \qquad (3.1)$$

where $f_n(E)$ and $f_p(E)$ are the Fermi-Dirac electron occupancy factors at the n- and p-sides respectively (see fig. 3.3), while $N_c(E)$ and $N_v(E)$ are the density of states in the conduction band and valence band, respectively. In the above expression we can recognize the three components of the tunnelling current density as given in section 3.1, i.e. the tunnelling probability T_t per unit time and the density of (un)occupied states on either side of the junction. The limits for E are the band edges.

The calculations of the tunnelling probability, as given in the literature, are based either on the application of the WKB approximation to tunnelling through a potential barrier or, more elaborately, on a direct solution of the Schrödinger equation. In both cases the electric field around the junction is assumed to be constant, i.e. a corresponding potential which depends linearly on the position x (see fig. 3.3). In the WKB approximation the tunnelling probability is given by [3.3]:

$$T_t \sim \exp\left[-2 \int_{x_1}^{x_2} |\kappa(x)| \, dx \right], \qquad (3.2)$$

where x_1 and x_2 are the turning points (see fig. 3.3). The imaginary wave vector of the electron in the gap $\kappa(x)$ depends on the shape of the potential barrier according to:

$$\kappa(x) = \sqrt{\frac{2m^{\times}}{\hbar^2} [U(x) - (E_{per})]},$$ (3.3)

where $U(x)$ is the potential barrier, E_{per} is the energy associated with the motion perpendicular to the junction interface and $m^{\times}$ is the effective mass of the electron. However, it is not quite clear what form the potential barrier should take in the case of an electron tunnelling through the gap. Since $\kappa(x)$ must be zero at the band edges, it is supposed that $\kappa(x)$ varies in a continuous way through the gap and has a maximum somewhere near the middle of the gap [3.3]. When T_t is calculated by means of a direct solution of the Schrödinger equation, a term $q\underline{F}.\underline{r}$ is added to the Hamiltonian of the system. For the direct case these calculations have been carried out by Kane [3.4] and for the indirect case by Keldysh [3.5]. In [3.3 − 3.5] it can be found that all the above approaches lead to an expression for T_t of the following form:
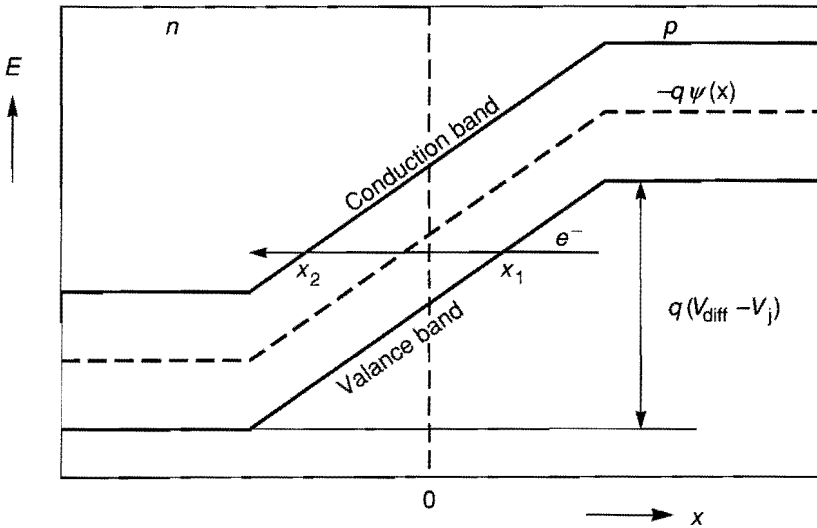

Fig. 3.3. A sketch of the assumed potential energy around a p-n junction. The metallurgical junction is at $x = 0$.

$$T_t \sim \exp\,(-F_0/|F|), \tag{3.4}$$

where F is the electric field. The parameter F_0 is proportional to $\sqrt{m^\times}$ and $E_g^{3/2}$, where E_g is the bandgap. Due mainly to different values of the effective mass, different theoretical values of F_0 have been reported, ranging from $2.10^9\ V/m$ to $3.10^9\ V/m$ [3.6 − 3.8]. It is shown in [3.9] that expression (3.1) for the current density per unit energy can now be written in the following form:

$$dJ_{bbt} = c\,|F|^{\sigma}\,D\,(F,E,V_j)\,\exp\,(-F_0/|F|)\,dE, \tag{3.5}$$

where σ is a numerical constant ($\sigma = 1$ for direct tunnelling and $\sigma = 3/2$ for phonon-assisted tunnelling). The function $D\,(F,E,V_j)$, which contains the occupancy factors and a term associated with the motion parallel to the junction interface, is practically unity for energy values more than a few times kT away from the quasi-Fermi levels. V_j is the junction voltage, which is taken to be negative in reverse bias. At a sufficiently high reverse-bias the term $D\,(F,E,V_j)$ can be set equal to one in our case.

An expression which is often used to describe band-to-band tunnelling in reverse-biased p-n junctions [3.1,3.6 − 3.8] takes the following form

$$J_{bbt} = q\,c_{bbt}\,V_j\,|F|^{\sigma}\,\exp\,(-F_0/|F|). \tag{3.6}$$

Eq. (3.6) is obtained from (3.5) by assuming that all electrons with an energy between the quasi-Fermi levels have the same tunnelling probability, i.e. F is constant. In that case $c_{bbt} = c$. In practice the electric field around a junction is far from constant (see fig. 3.4) and the question arises as to which value of F to use in the above expression. Among the few papers which discuss tunnelling in reverse-biased silicon p-n junctions in quantitative terms this question has received little attention, but it is of particular importance when considering using (3.6) as a model description of tunnelling for device simulations and compact modelling. Sometimes it is tacitly assumed that the maximum field
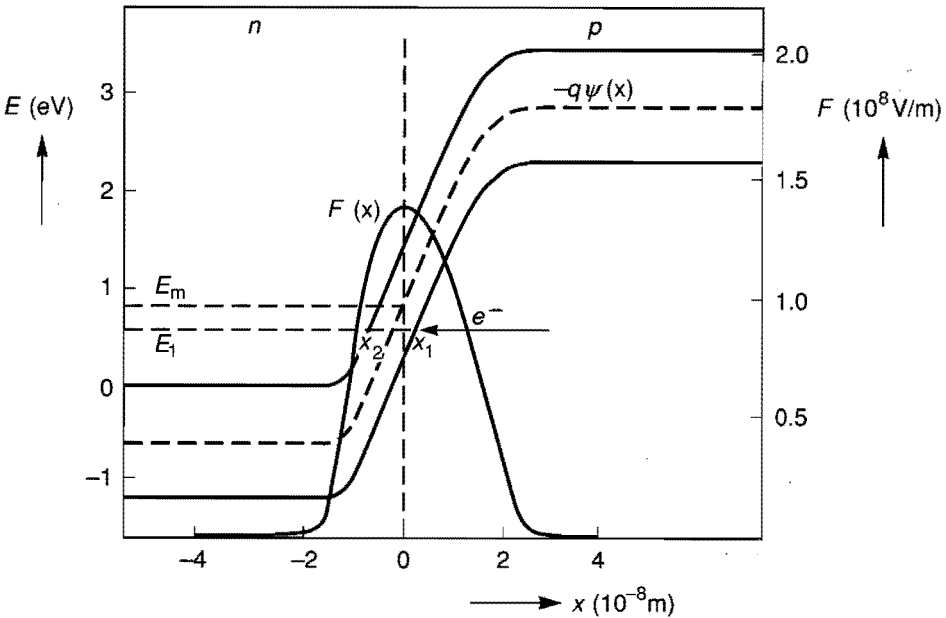
Fig. 3.4. A sketch of the actual potential and electric-field distribution around a p-n junction. The electrical junction is at $x = 0$.

F_{max} has to be used [3.6,3.7], but other suggestions have also been made, such as the average field $F_{av} = (V_{diff} - V_j)/W$ [3.3,3.9], where W is the depletion layer width and V_{diff} is the diffusion voltage. In this section we concentrate on the feasibility of expression (3.6) as a model for tunnelling in reverse-biased p-n junctions and, in particular, on what value should be used for F in (3.6), F_{max} or F_{av}.

3.2.2 Measurements

To test the expression (3.6) F_{max} and F_{av} must be determined experimentally. By the application of Gauss' law, the value of F_{max} can be
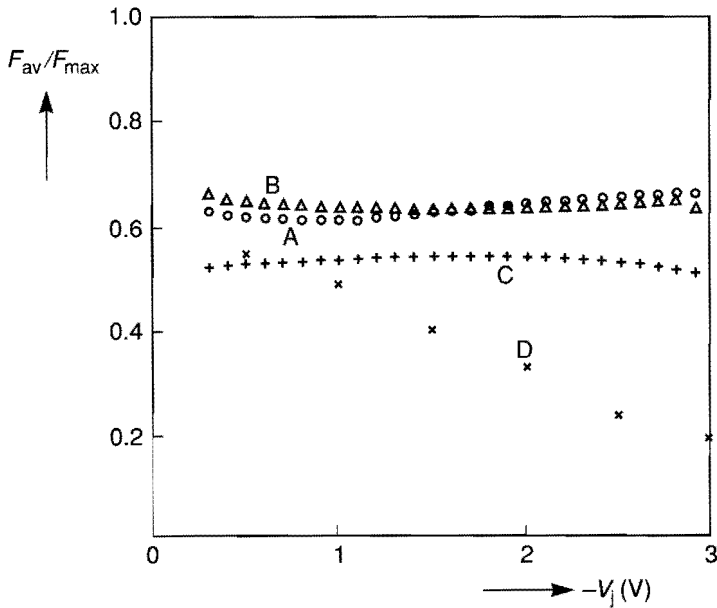
Fig. 3.5. Experimentally determined ratio of the average field and the maximum field as a function of the applied voltage for four devices.

determined by a numerical integration of the measured depletion capacitance $C(V_j)$, as described in [3.10] (see also section 3.4.3). The average electric field F_{av}, as defined above, can also be determined from the measured depletion capacitance by using the relation $W(V_j) = \varepsilon_0 \varepsilon_r A_j / C(V_j)$, where A_j is the junction area. Fig. 3.5 shows the measured ratio of the average field and the maximum field for four different devices. Devices A and B are Zener diodes, C and D are the bottom part of the emitter-base junction of a bipolar transistor (D is from the literature [3.10]). This ratio, which is around 2/3 for A and B and around 1/2 for C, indicates that A and B have linearly graded junctions and C can be considered to have an abrupt junction. For device D this ratio shows a different behaviour which is due to the fact
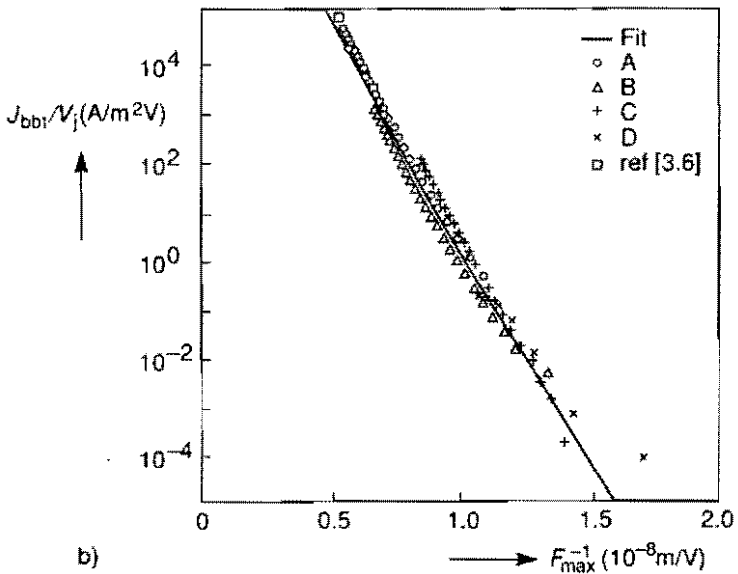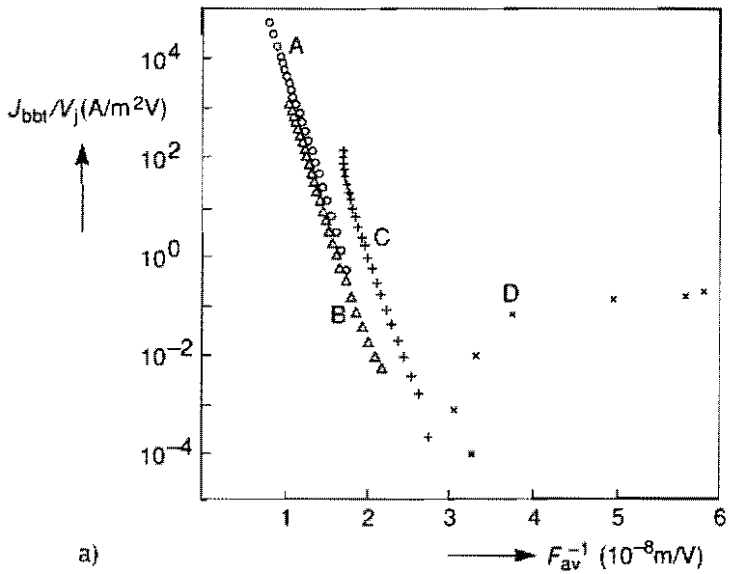
Fig. 3.6. The ratio of the tunnelling current density and the applied voltage as a function of the reciprocal electric field for two cases, (a) the average field and (b) the maximum field.

that the boron distribution in the base falls off rapidly, resulting in a strong decrease in the depletion capacitance with increasing reverse bias, as explained in [3.10].

In order to test expression (3.6), the measured values of the ratio of the current density and the applied voltage are plotted logarithmically against the reciprocal measured electric field for two cases, i.e. the average field and the maximum field, as shown in fig. 3.6a and fig. 3.6b respectively. If the model is correct for either F_{max} or F_{av} the corresponding measured points, when plotted as described above, should all follow a single curve, which is approximately a straight line with a slope of $-F_0 \log e$ (the actual slope is somewhat larger than this value and not exactly constant because of the field-dependent prefactor). As can be clearly seen from fig. 3.6, this is true for the case where F_{max} is used and not for the case where F_{av} is used. The solid line in fig. 3.6b represents eq. (3.6) with $F_0 = 1.9\ 10^9 V/m$, $\sigma = 3/2$ and $c_{bbt} = 1.6\ 10^{15}$ $s^{-1}\ m^{-1/2}\ V^{-5/2}$. Fig. 3.6b also gives a plot of the measured tunnelling currents as presented in [3.6]. The value of F_0 is in reasonable agreement with the theoretical values as given in the preceding section. The measured value of the prefactor c_{bbt} is about three decades lower than the calculated value for direct tunnelling [3.9]. The difference can be

Table 3.1. Measured and calculated temperature dependence of the tunnelling currents.

	$d\,(\ln |J_{bbt}|\,)/dT\ \ (10^{-3}/K)$	
	Measured	Calculated
A	5.5	3.2
B	5.5	3.9
C	7.4	7.1
D	10.1	6.0
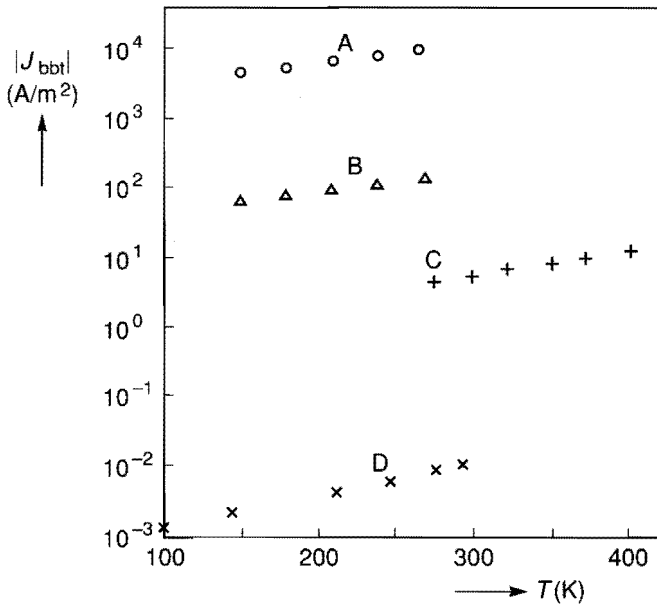
Fig. 3.7. Measured temperature dependence of the tunnelling currents. The applied voltage is $2V$ for devices A and B and $1.8V$ for devices C and D.

understood by realizing that there is an additional factor involved in the indirect case, namely the probability of an electron being scattered by phonons [3.3]. In fact, this measured value of c_{bbt} agrees with the rough theoretical estimation for the indirect case as given by Kane [3.9]. It is important to notice that in the case where all devices would have the same constant ratio between the average field and the maximum field it would have been impossible to determine the most appropriate choice for the electric field in (3.6). In that case the measured points would also have followed a single curve when plotted against F_{av}^{-1}, but with another slope and, correspondingly, a smaller value of F_0.

The tunnelling nature of the measured currents has been verified by measuring the temperature dependence of the currents. The observed small positive dependence of the current on the temperature (see fig. 3.7) is assumed to be mainly due to the temperature dependence of F_0 (through the temperature dependence of the bandgap) [3.6]. With this assumption it can easily be shown that

$$\frac{d\,(\ln |J_{bbt}|\,)}{dT} = -\frac{3}{2E_g}\frac{dE_g}{dT}\frac{F_0}{|F|}. \tag{3.7}$$

Table 3.I presents the measured and the calculated values of $d\,(\ln |J_{bbt}|\,)/dT$, where the almost linear dependence of the bandgap on the temperature, as given in [3.6], has been used. The reasonable agreement between the measured and calculated temperature dependence virtually excludes the possibility of other mechanisms (avalanche multiplication and thermally activated processes such as those summarized in [3.1]) being responsible for the I-V characteristics observed.

3.2.3 Interpretation of measurement results

The experiments clearly show that the results can satisfactorily be described by the conventional expression for Zener tunnelling, provided that F_{max} is used in that expression. In other words, in actual situations where the electric field is far from constant, the tunnelling probability is effectively determined by the maximum field and not, as one might intuitively expect, by an average field. In this section we show theoretically that an evaluation of (3.5), taking account of the actual electric field distribution, does indeed lead to a voltage dependence of the tunnelling current which can effectively be described by (3.6), provided that F_{max} is used in that equation. To this end we consider the tunnelling current density per unit energy as given by eq. (3.5). In practice, the tunnelling probability has a sharp maximum at a certain energy level. In order to find the actual relation between F and E, let us consider an electron with a certain energy, say E_1, entering the gap at $x = x_1$ and leaving it at $x = x_2$ (see fig. 3.4). For this electron we take as a value for F in (3.5) the electric field in the middle of the gap, i.e.

$F = F (x_1/2 + x_2/2)$. The reason for choosing the midgap electric field is that from a WKB analysis it follows that the largest contribution to the tunnelling probability comes from the region near the middle of the gap [3.3], where the potential can be approximated by $\psi \simeq - x F (x_1/2 + x_2/2)$.

In practical cases the value of F_0 is much larger than the maximum electric field. This implies that the function dJ_{bbt}/dE has a very sharp maximum around the energy value $E = E_m$ at which the midgap electric field is maximum or, equivalently, where the tunnelling distance $|x_2 - x_1|$ is minimum. So, in effect only electrons with an energy near E_m contribute to the tunnelling current. We make use of this fact in order to integrate (3.5) analytically. In appendix 3A it is shown that this leads to the following expression for the tunnelling current density

$$J_{bbt} = - cq \sqrt{\frac{2 \pi \varepsilon_0 \varepsilon_r}{q F_0 a}} \, |F_{max}|^{\sigma+2} (1 - \frac{\sigma + 1}{2} \frac{|F_{max}|}{F_0}) \exp \left(\frac{- F_0}{|F_{max}|} \right)$$

(3.8)

In the above equation $qa = (d\rho/dx)_{x=0}$, i.e. the gradient of the space charge density at the electrical junction which, for reverse-biased junctions, is virtually equal to the gradient of the dope at the electrical junction. The accuracy of the analysis leading from (3.5) to (3.8) is tested by numerical integration of (3.5), using the relation between F and E as obtained from a 1D device simulator. A comparison of the numerically obtained value of the current density and the analytically obtained value (expression (3.8)) gives a maximum difference of 4% for practical doping profiles and for values of F_{max} up to $2 \ 10^8 \ V/m$. Apart from the low voltage regime, where both models become inaccurate because the term $D (F,E,V_j)$ is set equal to one, the models show similar J-V curves. A comparison between (3.8) and (3.6) confirms the experimental results, i.e. the difference between the case of a constant field and the actual case lies in a different prefactor, which can be interpreted effectively as a smaller number of electrons involved in the tunnelling process, rather than in a different tunnelling probability, which would be the case if F_{av} had to be used.

3.3 TUNNELLING IN FORWARD-BIASED JUNCTIONS

3.3.1 Introduction

In forward bias where the bands are "uncrossed" (i.e. beyond the point where the conduction band minimum at the n-side is at the same energy level as the valence band maximum at the p-side, see fig. 3.2d) elastic tunnelling is no longer possible. Despite this, at high doping levels or at low temperatures the measured forward diode current is considerably in excess of the normal diode current [3.2,3.11 − 3.13]. This diode current is believed to be primarily caused by a tunnelling process since its dependence on temperature and pressure parallels the dependence of the band-to-band tunnelling current on these quantities [3.2]. The possible mechanisms leading to the excess current can be divided into two classes, i.e. that which requires states in the forbidden gap and that which does not. In the latter case the electrons can lose energy by emitting photons, phonons or in Auger processes. Kane [3.9] has considered several of these hypotheses theoretically and has concluded that they are unlikely to account for the excess current. Moreover, the phonon emission hypothesis has been tested experimentally but no evidence for this mechanism could be found [3.2]. The former case, i.e. tunnelling via deep states in the gap, is believed to be the most likely (see fig. 3.8). This mechanism was first suggested by Yajima and Esaki [3.11] and, subsequently, several authors have obtained strong confirmation of it by showing that the magnitude of the excess current can be altered by changing the trap density, either by suitable doping or by radiation damage [3.2,3.12,3.15]. It is interesting to note the striking analogy between the trap-assisted tunnelling processes and the ordinary Shockley-Read-Hall recombination processes. Not only do the two classes of transitions (band-to-band and via deep states) occur in both processes, but in the case of the SRH process the transitions via deep impurity states are also most likely [3.14].

Several authors [3.2,3.13] give a semi-empirical relation between the trap-assisted tunnelling current density and a certain function $h(V_j)$ of the applied voltage V_j of the following form
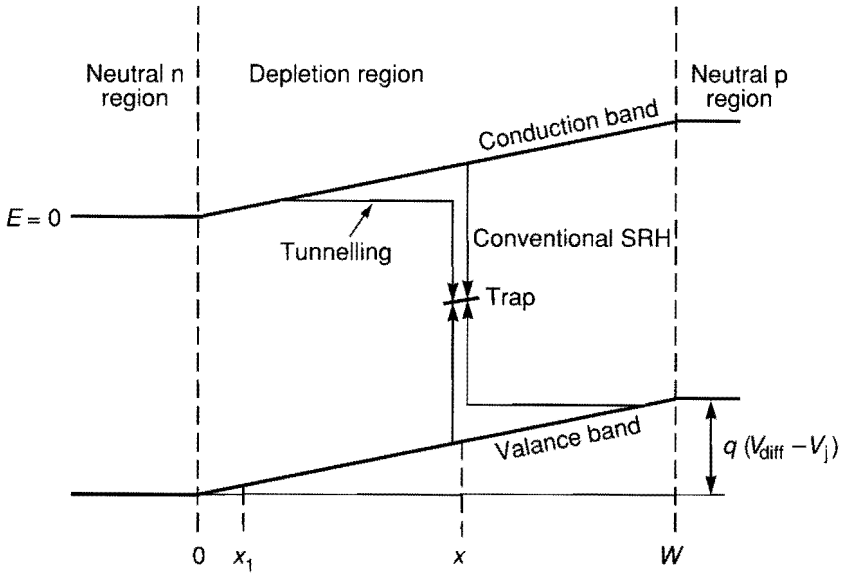
Fig. 3.8. Schematic energy-band diagram of a forward-biased junction. Both the conventional SRH process and the trap-assisted tunnelling process are denoted.

$$J_{tat} = c_{tat} \, e^{h(V_j)}. \tag{3.9}$$

The idea behind these models is that the current is proportional to a tunnelling probability T_t, given by

$$T_t \sim e^{h(V_j)}. \tag{3.10}$$

In [3.2], $h(V_j)$ is adapted from Franz [3.16] to be

$$h(V_j) = -\frac{4\sqrt{2m^\times}\,\Delta E^{3/2}}{3qh|F|}. \tag{3.11}$$

Fig. 3.9. Potential barriers as assumed to obtain expressions (3.11) (a) and (3.12) (b).

In [3.16] the above expression is obtained from the derivation of the probability that an electron will make an elastic transition from an impurity state to the conduction band. In this application ΔE is the height of the tunnelling barrier for an electron that tunnels from one side of the depletion layer to the ot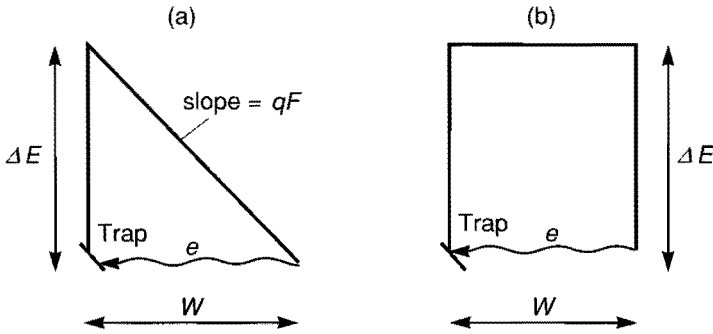her (see fig. 3.9a). This barrier height is assumed to be $\Delta E = q \ (V_{diff} - V_j)$. In [3.13], $h \ (V_j)$ is obtained from the tunnelling probability through a rectangular barrier of width W and height $\Delta E = q \ (V_{diff} - V_j)$ (see fig. 3.9b). In that case

$$h \ (V_j) = - \frac{2W}{h} \sqrt{2m^{\times}\Delta E} \, . \tag{3.12}$$

It can easily be shown that, although the physical approach to the derivation of $h \ (V_j)$ is different, the models exhibit equal dependence of J_{tat} on V_j.

However, when we consider formulation (3.9) as a basis for a physical model in numerical device simulations, it suffers from the following drawbacks:

- The magnitude of the proportionality constant c_{tat} cannot be obtained from the model but must be determined by fitting the experimental I-V curves.
- A model that describes the tunnelling effect in the form of a current density is less suitable for incorporation into a numerical device simulator. To this end the tunnelling contribution must be given in the form of a recombination term in the continuity equations.

Furthermore, for these models it is assumed that only electrons at $x = 0$ (and similarly holes at $x = W$) contribute to the tunnelling current and they therefore neglect the possibility of injected electrons in the depletion layer (e.g. at x_1 in fig. 3.8) tunnelling to a trap. This might be particularly significant in a transition region where both conventional SRH recombination and trap-assisted tunnelling are important.

In the next section we present a rather heuristic recombination model which includes both SRH recombination and trap-assisted tunnelling. As will be shown, this model does not suffer from the above-mentioned drawbacks.

3.3.2 A recombination model including trap-assisted tunnelling

Basically, SRH recombination and trap-assisted tunnelling are similar mechanisms, i.e. recombination via impurity states. The difference lies in the character of the state from which the electron makes a transition to a trap state. In the conventional SRH process the initial state is a non-localized (Bloch) state in the conduction band, and in the tunnelling process the initial state is a tunnelling state in the gap. In our model the electron and hole concentrations in the depletion layer are considered to consist of two contributions, i.e.

$$n = n_{sc}(x) + n_{tun}(x), \quad (3.13a)$$

$$p = p_{sc}(x) + p_{tun}(x), \quad (3.13b)$$

where n_{sc} and p_{sc} are the conventional, semi-classical electron and hole concentrations (i.e. from states in the bands) while n_{tun} and p_{tun} are the concentrations due to tunnelling. If we make no distinction in the

capture processes between carriers in the band and tunnelling carriers, the net recombination rate is given by an SRH-like expression

$$R = \frac{p\,n - p_1 n_1}{\tau_p\,(n + n_1) + \tau_n\,(p + p_1)}. \tag{3.14}$$

In the above expression τ_p and τ_n are the recombination lifetimes, while n_1 and p_1 are the carrier concentrations at equilibrium.

The expressions for the carrier concentrations are obtained by the application of the so-called envelope-function formalism, first developed by Slater [3.17] and frequently used, for instance, in solving heterojunction problems and for obtaining impurity energy levels [3.18]. In the unperturbed situation the bottom of the conduction band everywhere in the device is at $E = 0$. The perturbation potential is due to both the built-in potential and the applied voltage. This potential equals zero for $x < 0$ and is equal to $q\,(V_{diff} - V_j)$ for $x > W$. The total electron concentration in the depletion layer $n\,(x)$ is given by

$$n\,(x) = \int_0^\infty N(E)\,f\,(E)\,\phi_E^2\,(x)\,dE, \tag{3.15}$$

instead of the semi-classical expression

$$n_{sc}\,(x) = \int_{E_c\,(x)}^\infty N(E)\,f\,(E)\,dE. \tag{3.16}$$

In the above expressions $N(E)$ is taken to be the unperturbed density of states in the conduction band and $f\,(E)$ is the Fermi-Dirac occupancy factor. $\phi_E\,(x)$ is the envelope wave function which is the eigenfunction of the effective-mass Schrödinger equation

$$\left[-\frac{\hbar^2}{2m^\times}\frac{d^2}{dx^2} + U(x) \right]\phi_E\,(x) = E\,\phi_E\,(x). \tag{3.17}$$

For a linearly varying potential in the depletion layer as sketched in fig. 3.8 ($U(x) = qFx$), (3.17) can be written in the following form

$$\left[-\frac{d^2}{d\xi^2} + \xi \right] \phi_E (\xi) = \frac{2m^\times E}{\hbar^2 \gamma^2} \phi_E (\xi), \tag{3.18}$$

with $\xi = \gamma x$ and

$$\gamma = (2q|F|m^\times \hbar^{-2})^{1/3}. \tag{3.19}$$

The solutions of equations of the type of (3.18) are the Airy functions Ai and Bi [3.19]. Since for $(V_{diff} - V_j) >> kT/q$ we can assume that only the decaying function Ai is allowed, the solution of (3.17) is

$$\phi_E (x) \sim Ai[\gamma(x - x_1)], \tag{3.20}$$

with $x_1 = E/(qF)$. As can be seen from fig. 3.10, for $x > x_1$ the Airy function decays and represents the tunnelling behaviour, while for $x < x_1$ the Airy function oscillates and represents the non-localized behaviour. The exact value of the effective mass $m^\times$ to be used is not clear and, moreover, depends on the crystal orientation. Experiments and theoretical treatments on Zener tunnelling suggest a value between 0.1 and 0.2 [3.6,3.7,3.20].

In order to make the result suitable for incorporation into a conventional numerical device simulator we must express $n(x)$ as a function of the conventional electron concentration $n_{sc}(x)$. This can be accomplished by putting $N(E) f(E) \Delta E \simeq - (dn_{sc}/dx_1)\Delta x_1$ (see appendix 3B). We obtain

$$n(x) = Ai^{-2}(0) \int_0^\infty \left(-\frac{d n_{sc}(x)}{dx} \right)_{x=x_1} Ai^2 [\gamma(x - x_1)] dx_1. \tag{3.21}$$

Note that the upper limit of the above integral has been put equal to ∞, which is a good approximation only when $(V_{diff} - V_j) >> kT/q$. The contribution of the integration interval $[0,x]$ is the tunnelling contrib-

ution, while the contribution of the interval $[x, \infty]$ corresponds to the semi-classical injected electron density. In order to be consistent with expression (3.16) for $n_{sc}(x)$, (3.21) is rewritten as

$$n(x) = n_{sc}(x) + \int_0^x \left(-\frac{d\, n_{sc}(x)}{dx} \right)_{x=x_1} \frac{Ai^2\,[\gamma(x - x_1)]}{Ai^2(0)}\, dx_1. \quad (3.22)$$

The second term on the right-hand side of the above expression is the tunnelling contribution. The physical meaning of the term $Ai^2[\gamma(x - x_1)]/Ai^2(0)$ is the probability that an electron at x_1 will tunnel to a trap at x. It should be noted that, since only electrons at an energy level above the trap level contribute to the recombination, in the
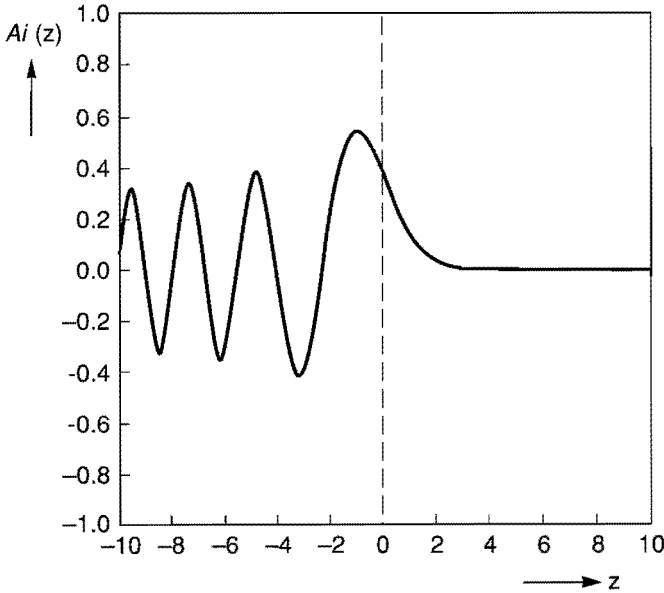

Fig. 3.10. A plot of the Airy function $Ai\,(z)$.
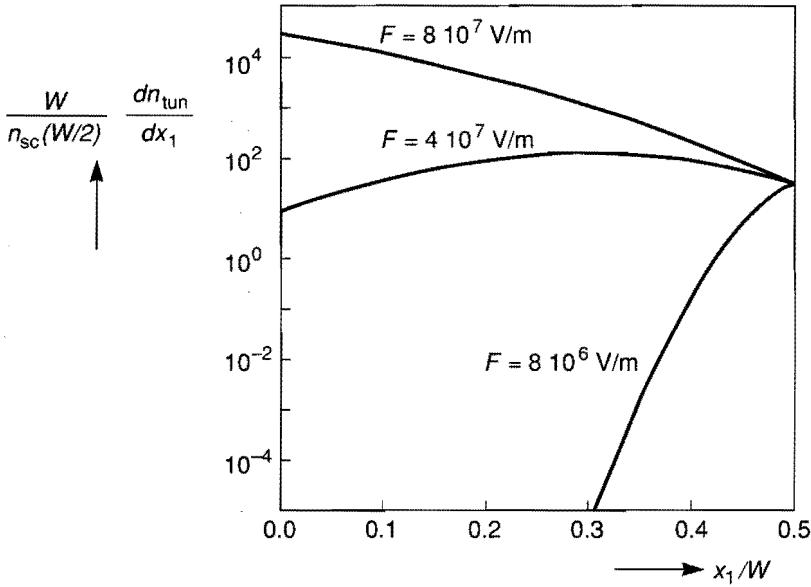
$$\frac{W}{n_{sc}(W/2)}\frac{dn_{tun}}{dx_1}$$

Fig. 3.11. The normalized contribution from x_1 to the tunnelling electron concentration at $x = W/2$ for three values of the electric field. $T = 300K$ and $V_{diff} - V_j = 0.5V$.

case of shallow impurities, i.e. $E_T < q|F|x$, the lower bound of the integral in (3.22) must be replaced by $x - E_T/(q|F|)$, where E_T is the trap level measured from the conduction band. In fig. 3.11 the normalized contribution from x_1 to the tunnelling electron concentration in the middle of the depletion layer (i.e. the integrand of (3.22) for $x = W/2$, divided by $n_{sc}(W/2)/W$) is plotted as a function of x_1/W for three values of the electric field. This plot shows where the maximum contribution to the tunnelling electron density at $x = W/2$ originates. We see that for high fields the largest contribution comes from $x_1/W = 0$, i.e. from the edge of the depletion layer. For low fields the largest contribution comes from the region close to x_1, while for intermediate fields the
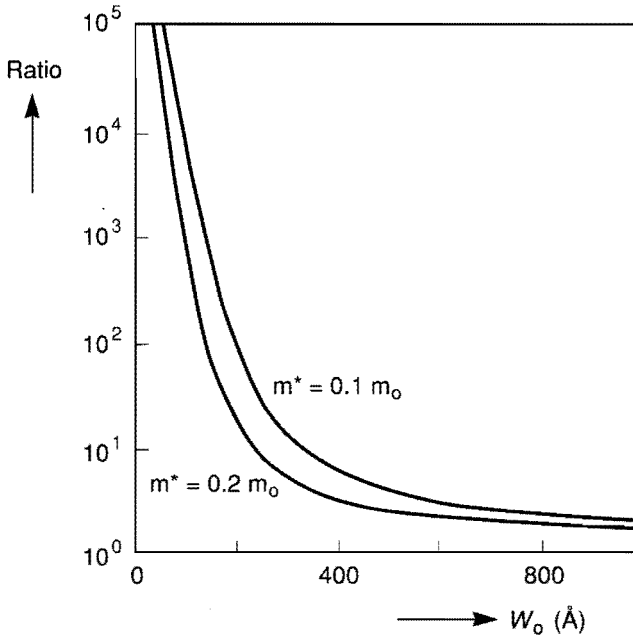
Fig. 3.12. Ratio of the calculated current obtained using the proposed model and that obtained with the conventional SRH model as a function of the zero-bias depletion layer width. The junction voltage is $0.3V$ and $T = 300K$.

maximum contribution comes from the region somewhere between $x_1/W = 0$ and $x_1/W = 0.5$.

For holes a similar procedure leads to

$$p(x) = p_{sc}(x) + \int_W^x \left(-\frac{d\,p_{sc}(x)}{dx} \right)_{x=x_1} \frac{Ai^2\left[\gamma(x_1 - x)\right]}{Ai^2(0)} \, dx_1. \text{(3.23)}$$

It is interesting to note that when only tunnelling from one side of the depletion layer to the other is considered, i.e. $(x - x_1) = W$, the asymptotic expansion of the Airy function yields

$$Ai^2 (\gamma W) \sim \exp \left(- \frac{4\sqrt{2m^\times} \left[q \left(V_{diff} - V_j \right) \right]^{3/2}}{3q\hbar |F|} \right). \qquad (3.24)$$

This expression is identical to the tunnelling probability as used by Chynoweth et al. (see (3.11)) to obtain an expression for the current density. Furthermore, for a reverse-biased junction, where the tunnelling distance is $E_g/(q|F|)$, the asymptotic expansion of the Airy function yields the well-known field dependence of elastic tunnelling, viz.

$$Ai^2 (\gamma E_g/q|F|) \sim \exp \left(- \frac{4\sqrt{2m^\times} E_g^{3/2}}{3q\hbar |F|} \right). \qquad (3.25)$$

3.3.3 Results and comparison with experiments

Because (3.22) and (3.23) apply to a linear potential we have approximated the potential in the depletion layer by a straight line. We have used the average electric field and not the field at the electrical junction because, unlike the conventional SRH process, in the case of tunnelling the whole depletion region contributes to the values of n and p at the electrical junction. A standard model for the recombination lifetime is used [3.21], together with the Slotboom-De Graaff model for band-gap-narrowing [3.22]. The current density is obtained by numerically solving the stationary continuity equation (1.2) for either holes or electrons. This yields

$$J = -q \int_0^W R(x) \, dx. \qquad (3.26)$$

From the calculations it follows that the ratio of the current obtained using the proposed model and that obtained with the conventional SRH model, depends strongly on the zero-bias depletion width W_0 or, equivalently, on the electric field. This is shown in fig. 3.12 where this
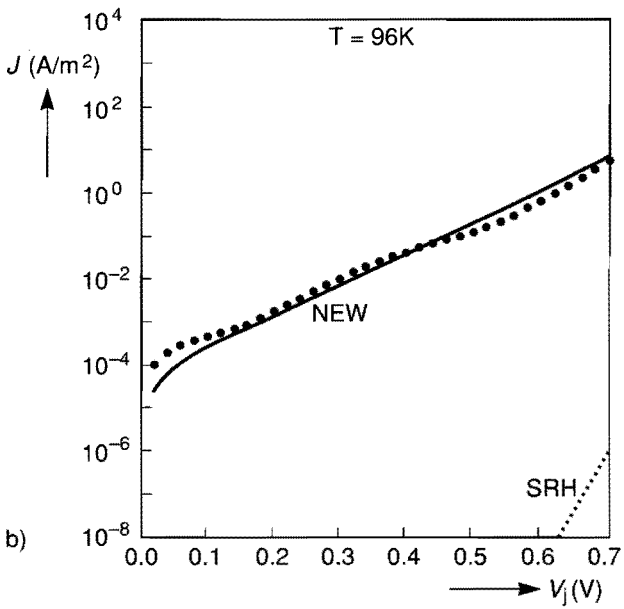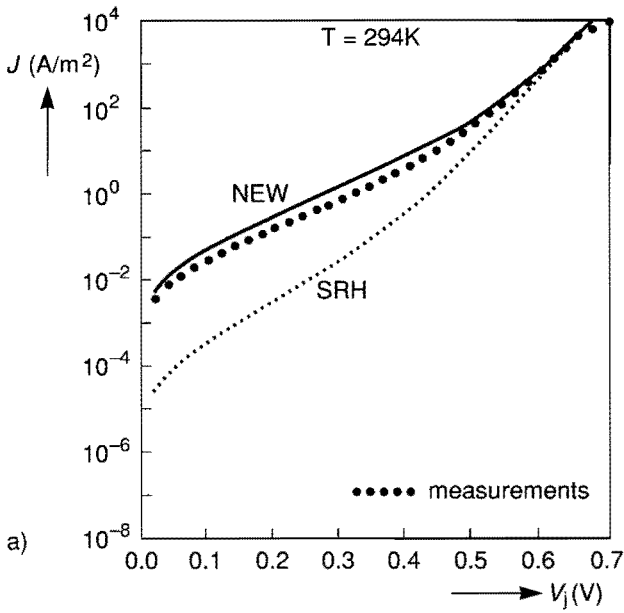
Fig. 3.13. Measured (dots) and calculated I-V characteristic for a diode with a linearly-graded junction at $T = 294K$ (a) and $T = 96K$ (b).

Tunnelling effects in p-n junctions

97

ratio is plotted for two values of the effective mass. Furthermore, from this figure it is apparent that irrespective of the exact value of the effective mass, the tunnelling effect increases strongly when the zero-bias depletion width is less than about 300Å. In the case of a two-sided abrupt junction this value corresponds to a doping level of $3 \ 10^{24} m^{-3}$. This is in rough agreement with experiments [3.13]. For these calculations we have used the dependence of W on V_j for a linearly graded junction. The results are not essentially different for an abrupt junction.

In order to test the model, we have compared calculated results with measurements on diodes having a linearly graded junction and used the dependence of W on V_j as obtained from capacitance meas-
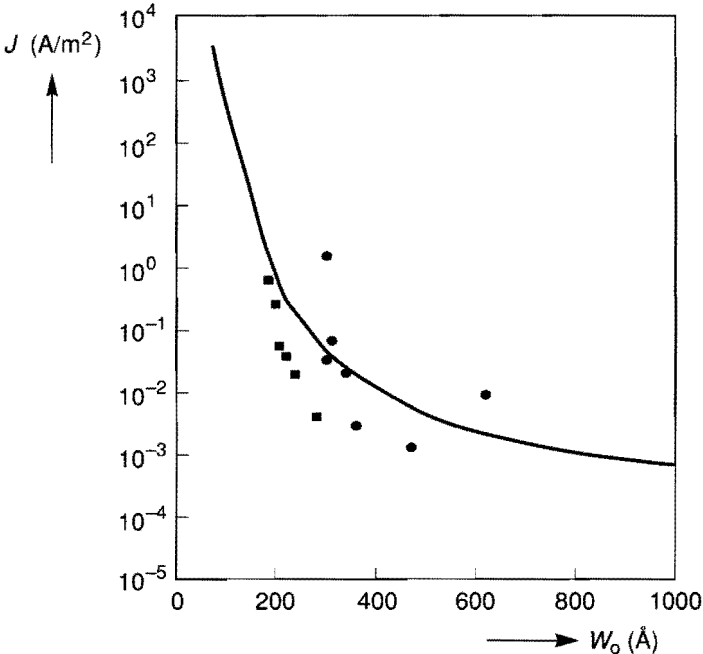

Fig. 3.14. Measured and calculated (including tunnelling) non-ideal diode current density as a function of the zero-bias depletion width at $0.3V$ junction voltage and at room temperature. The squares are own measurements and the dots are from [3.13].
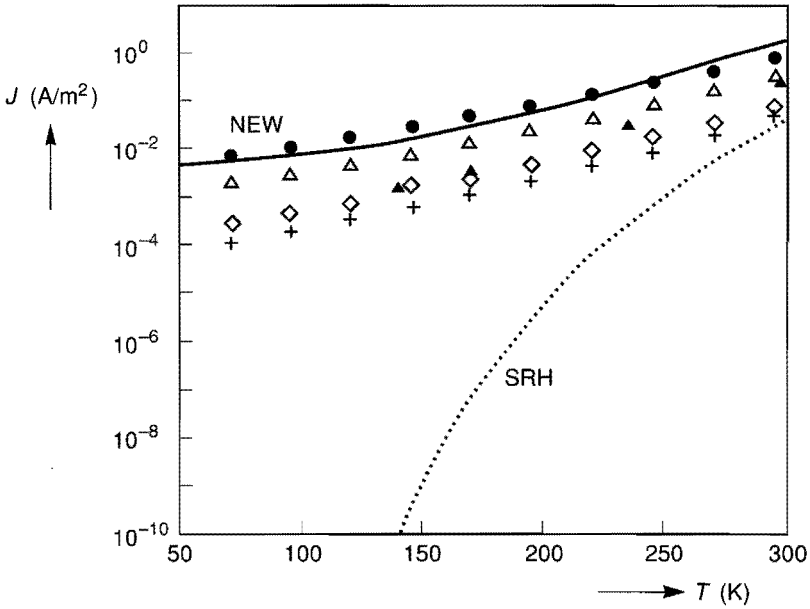
Fig. 3.15. Measured and calculated non-ideal diode currents as a function of temperature at $0.3V$ junction voltage. Different symbols denote different diodes while the calculations are performed for the diode denoted by the solid dots.

urements. In these calculations an effective mass of $0.15m_0$ is used. In fig. 3.13 the measured and calculated J-V characteristics are plotted for a diode with a 185Å zero-bias depletion layer width at two temperatures. Note that the difference between, on the one hand, the conventionally calculated current and, on the other hand, that obtained by the proposed model and the measurements is more than one decade at low bias at room temperature. Although this difference is most striking, it is not the most fundamental one because it might be due to a wrong estimation of the lifetimes τ_n and τ_p, resulting in a wrong absolute value of the current density. These lifetimes are inversely proportional to the trap density. Although the models for the lifetimes

Tunnelling effects in p-n junctions
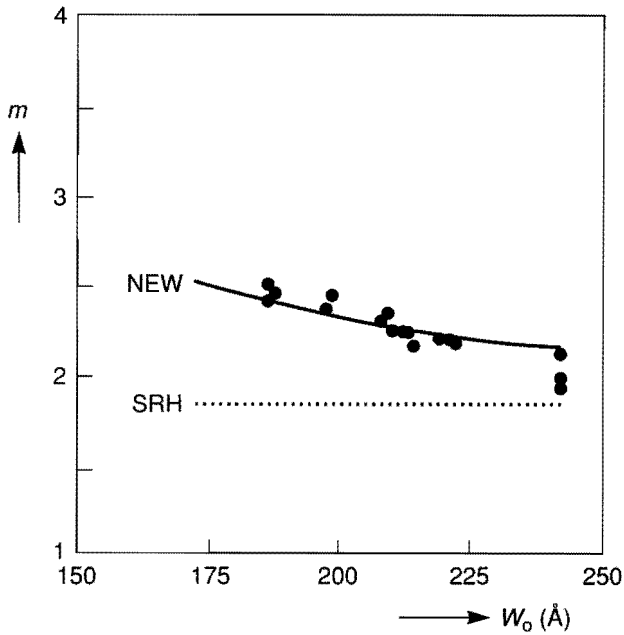
Fig. 3.16. Measured and calculated non-ideality factors for a number of diodes as a function of the zero-bias depletion layer width. $T = 294K$.

assume a direct correspondence of these lifetimes with the doping density, the trap density and therefore the lifetimes are believed to be very process-dependent. This is illustrated in fig. 3.14 where measured non-ideal diode currents are plotted as a function of the zero-bias depletion width for diodes having slightly different doping profiles. Note the large spread between the various measurements. In this figure the corresponding calculations are also denoted.

Figs. 3.15-3.17 show two characteristic features associated with trap-assisted tunnelling which are satisfactorily described by the proposed model and cannot be explained by a possible erroneous choice of the recombination lifetimes, viz.
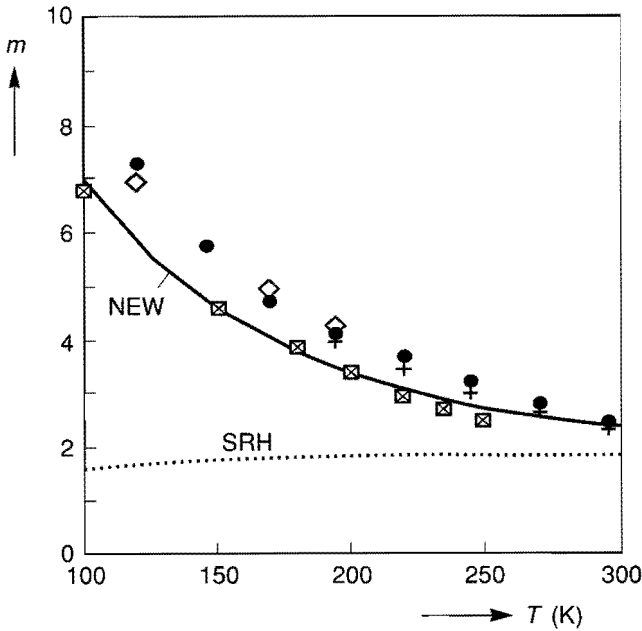
Fig. 3.17. Measured and calculated non-ideality factors as a function of temperature.

- The temperature dependence of the currents is much weaker than that predicted by the conventional SRH model (fig. 3.15). In fig. 3.15 the calculations are performed for the diode from fig. 3.13 represented by the solid dots. The other symbols denote different diodes with a slightly different depletion width. The solid triangles are from [3.13]. As can be observed from fig. 3.15, the difference between the current obtained with the proposed model and that obtained with the conventional SRH model may be as much as six decades at 150K. This is due to the weak temperature dependence of the tunnelling contribution to the carrier density.
- If J is expressed as $J \sim \exp{(qV_j/mkT)}$ the non-ideality factor m is larger than two, whereas the conventional SRH model predicts a va-

lue less than two. In fig. 3.16 this non-ideality factor is plotted as a function of the measured zero-bias depletion layer width while in fig. 3.17 m is plotted as a function of temperature. In both figures the calculations are performed for the diode in fig. 3.13 represented by the solid dots. Especially at low temperatures the difference becomes more pronounced. In fact, for these devices the tunnelling term predominates and the slope of the I-V characteristics is therefore virtually independent of temperature, resulting in a value of m which is almost inversely proportional to T.

3.4 A COMPACT DIODE MODEL INCLUDING TUNNELLING

3.4.1 Qualitative discussion

In this section the physical basis for a compact model describing reverse-biased diode currents is presented. Present diode models such as those implemented in the circuit-simulation packages PHILPAC and SPICE give poor results in the reverse bias regime [3.23,3.24]. The model in PHILPAC gives a constant low leakage current at each reverse bias point, while the model in SPICE gives a constant leakage current up to the breakdown voltage and beyond breakdown the negative forward characteristic is taken. To illustrate the problems in modelling the reverse characteristic, fig. 3.18 shows both the measured forward and reverse I-V characteristics of three different Zener diodes. From this figure it is apparent that while the forward characteristics are virtually equal, the reverse characteristics are quite different. Not only do the values of the current at a certain voltage differ by several orders of magnitude, but the shape of the characteristics is also different. According to standard textbooks (e.g. [3.1]) tunnelling predominates when the breakdown voltage V_{br} is less than $4E_g/q$. For junctions with a breakdown voltage in excess of $6E_g/q$, Shockley-Read-Hall generation together with avalanche multiplication predominates, while for intermediate values of V_{br} all three mechanisms may be important. A common way to discriminate between the three mechanisms is to investigate the temperature dependence of the characteristics. The SRH generation mechanism has a strong (exponential) positive temperature
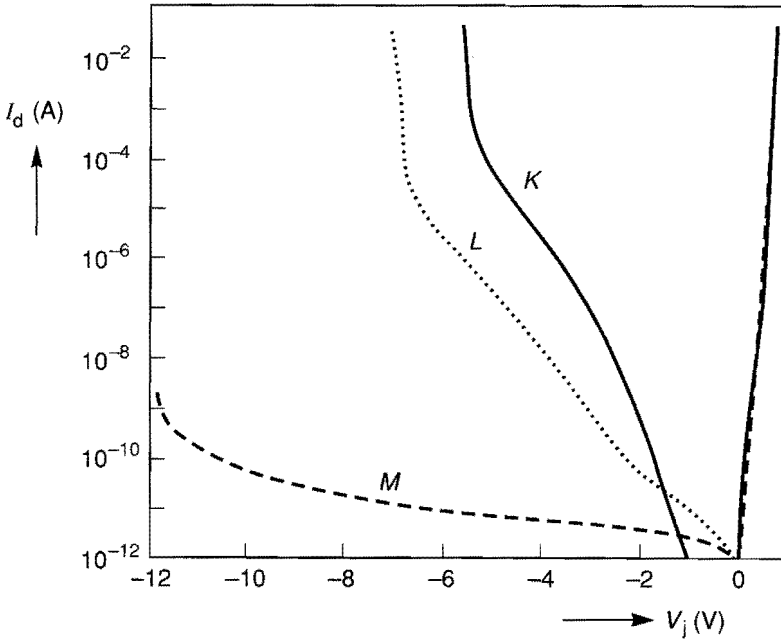
Fig. 3.18. Forward and reverse I-V characteristics of three different types of Zener diodes. The absolute value of the current is plotted on the vertical axis.

dependence, the band-to-band tunnelling mechanism has a very weak positive temperature dependence and the avalanche mechanism has a negative temperature dependence [3.1]. In figs. 3.19-3.21 the I-V characteristics of the above diodes at various temperatures are given. These measurements support the above-mentioned rule which relates the different underlying physical mechanisms to the breakdown voltage, i.e. in diode K tunnelling predominates, while in diode M SRH generation is predominant. In the case of diode L we can clearly distinguish two regions. In the low-bias regime a strong temperature dependence is observed, indicating that SRH generation is predominant, while in the high bias regime tunnelling predominates.
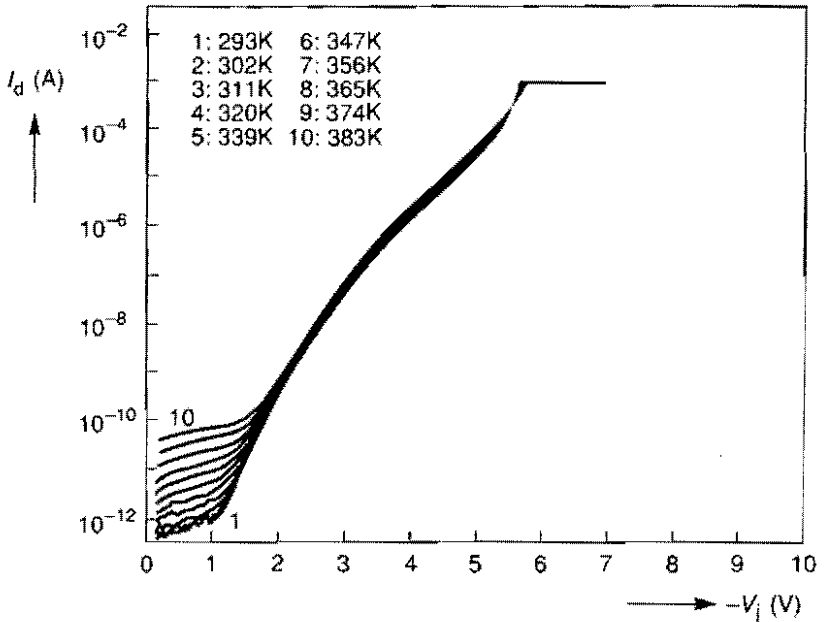
Fig. 3.19. Reverse I-V characteristics of diode K at various temperatures.

Based on these qualitative observations we infer that (at least) three generation mechanisms must be taken into account, viz.
- SRH generation in the depletion layer

$$R_{SRH} = \frac{(p\,n - n_i^2)}{\tau_p\,(n + n_i) + \tau_n\,(p + n_i)} \, , \qquad (3.27)$$

- avalanche multiplication

$$R_{av} = \frac{-1}{q}\,(\alpha_n|J_n| + \alpha_p|J_p|), \qquad (3.28)$$

- band-to-band tunnelling
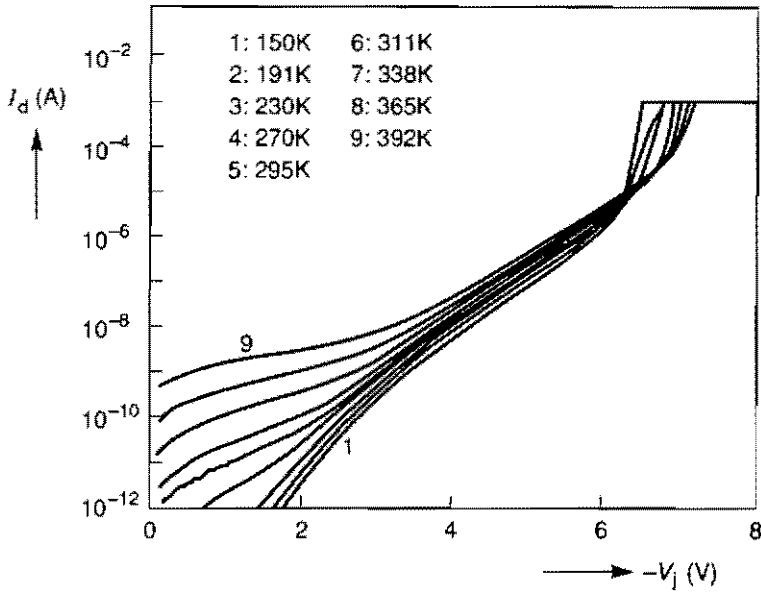
Fig. 3.20. Reverse I-V characteristics of diode L at various temperatures.

$$R_{bbt} = -\frac{|J_{bbt}|}{q} \, \delta(x). \qquad (3.29)$$

Expressions (3.27) and (3.28) can be found in standard textbooks while J_{bbt} is the tunnelling current density as given by (3.6) and $\delta(x)$ is the Dirac delta function. The origin $x = 0$ is at the location of the maximum electric field. The tunnelling mechanism is taken into account as a generation term for the following reason: basically, band-to-band tunnelling is the generation of electron-hole pairs by tunnelling of electrons from the valence band to the conduction band, leaving a hole behind. These electron-hole pairs can subsequently be accelerated to cause avalanche multiplication. This is taken into account by the above

description of tunnelling as a delta-function generation of electron-hole pairs at the location of the maximum electric field.

3.4.2 Model derivation

We consider a one-dimensional diode as sketched in fig. 3.22. Taking account of the three generation mechanisms given in the previous section, the hole and electron current densities behave as sketched in this figure. In order to derive the diode current J_d we have to solve the continuity equation for either holes or electrons. The stationary hole continuity equation (1.2b) for this structure reads

$$\frac{dJ_p}{dx} = -qR_{SRH}(x) + \alpha_n(x) J_n(x) + \alpha_p(x) J_p(x) + J_{bbt}\,\delta(x), \quad (3.30)$$
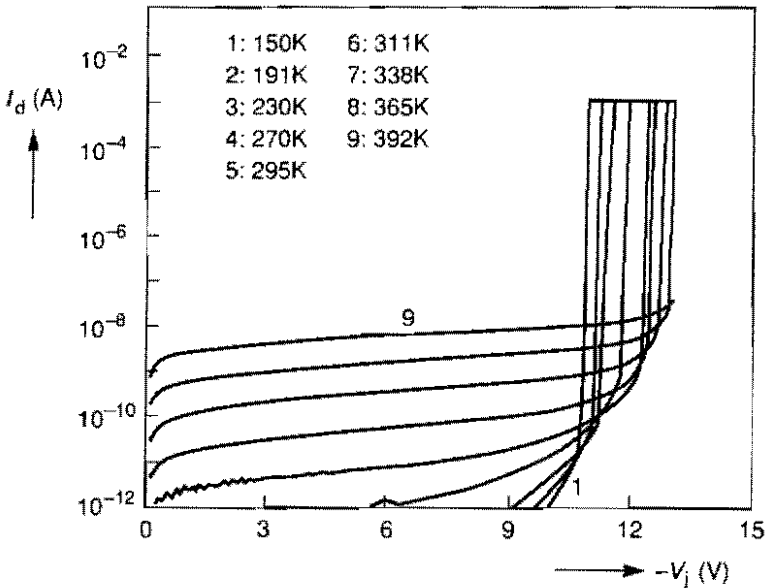
Fig. 3.21. Reverse I-V characteristics of diode M at various temperatures.
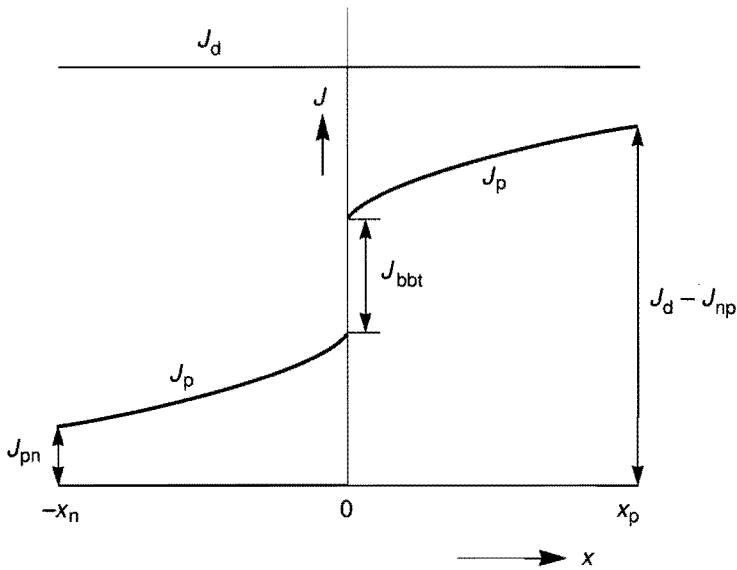
Fig. 3.22. Distribution of the hole current density around a one-dimensional p-n junction. Notice the discontinuity of the current density due to tunnelling generation.

where J_{bbt} is taken to be positive. The boundary conditions are

$$J_p\left(-x_n\right) = J_{pn} \tag{3.31}$$

and

$$J_p\left(x_p\right) = J_d - J_{np}. \tag{3.32}$$

The current densities J_{pn} and J_{np} are caused by recombination in the regions outside the depletion layer at the n and p side of the diode, respectively. The sum of these term J_s is the saturated ideal current density. Using the above boundary conditions, together with the relation

$$J_d = J_n + J_p, \qquad (3.33)$$

the above first-order differential equation can be solved and the following expression is obtained

$$J_d = \frac{J_{bbt}\, \phi(0) + J_{pn}\, \phi(-x_n) + J_{np} - q\displaystyle\int_{-x_n}^{x_p} R_{SRH}\,(x)\phi(x)\, dx}{1 - \displaystyle\int_{-x_n}^{x_p} \alpha_n\, \phi(x)\, dx}, \qquad (3.34)$$

with

$$\phi(x) = \exp\left(-\int_{x}^{x_p}(\alpha_n - \alpha_p)\, dx' \right). \qquad (3.35)$$

Breakdown occurs when the denominator equals zero, which is equal to the breakdown condition as given, for instance, in [3.1]. From the denominator in (3.34) it can be seen that this condition is determined only by the ionization integrals of α_n and α_p, hence by the electric field distribution in the depletion layer. This means that the breakdown voltage is entirely determined by the avalanche effect while the other effects (SRH generation and tunnelling) only serve as "source" terms for avalanche multiplication. In other words, in principle all three terms can contribute to the magnitude of J_d whereas V_{br} is determined by the avalanche effect alone.

It is obvious that expression (3.34) for the diode current is far too complicated in a circuit simulation environment and that essential simplifications have to be made. These fairly rigorous simplifications must be such that, on the one hand, the result is a simple analytical expression which can be used without much computational effort, while on the other hand the physical basis of the model is maintained. These simplifications are as follows:

1) A symmetrical step junction is assumed, so $R_{SRH}(x)$ is constant in the depletion layer, having a value of

$$R_{SRH}(x) = -\frac{n_i}{2\bar{\tau}}(1 - e^{qV_j/kT}) \equiv -G_{SRH}, \qquad (3.36)$$

except within a distance of $W_0/2$ from the depletion layer boundaries, where $R_{SRH} = 0$ [3.25]. Expression (3.36) is obtained from (3.27) by neglecting n and p in the denominator, which is justified by the fact that in reverse bias these concentrations are much less than n_i. In the numerator the relation $pn = n_i^2 \exp(qV_j/kT)$ is used in order to maintain zero generation at zero bias. Furthermore τ_p and τ_n are replaced by an effective lifetime $\bar{\tau}$. For the saturated ideal current it holds that $J_{pn} = J_{np} = J_s/2$.
2) The dependence of α_n on the electric field is given by the commonly used expression [3.26]

$$\alpha_n = \alpha_{n\infty} e^{-b_n/|F(x)|}, \qquad (3.37)$$

and α_p is taken to be proportional to α_n, i.e. $\alpha_p = \alpha_n/2$.
After a number of additional mathematical simplifications (see appendix 3C) we arrive at the following expression

$$J_d = \frac{J_{bbt} e^{-\mu_{av}} + (J_{SRH} + J_s)(1 + e^{-2\mu_{av}})/2}{1 - 2\mu_{av}(1 + e^{-2\mu_{av}})}, \qquad (3.38)$$

with

$$J_{SRH} = qG_{SRH}(W - W_0), \qquad (3.39)$$

and

$$\mu_{av} = 0.5\, d_{av}\, \alpha_{n\infty} \exp(-b_n/|F_{max}|), \qquad (3.40)$$

where d_{av} is an effective length for avalanche multiplication (see (3C.3)). W is the depletion layer width and W_0 is the depletion layer width at zero bias. Both (3.34) and (3.38) reflect the physical mechanism that
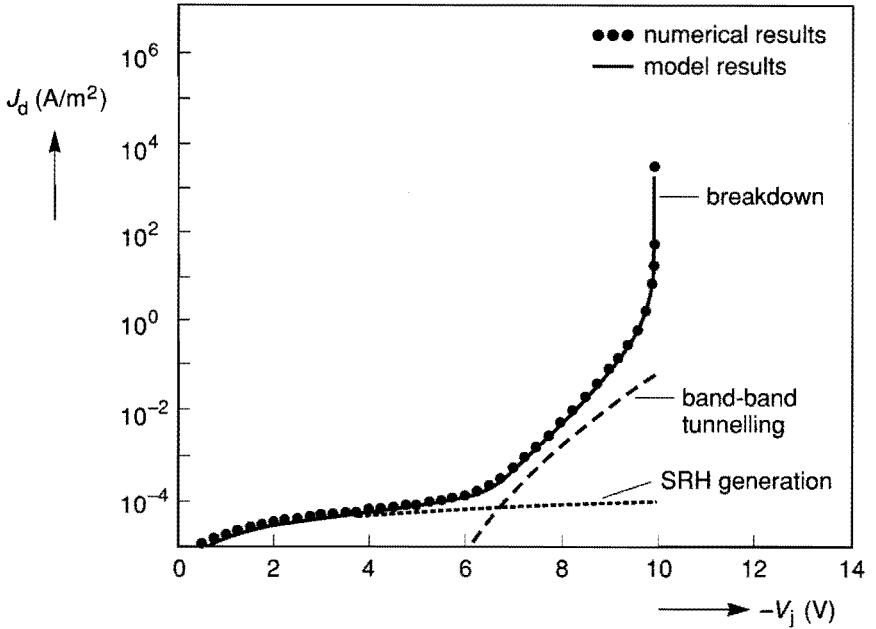
Fig. 3.23. A comparison of model results and numerical calculations. The contributions of the SRH generation and band-to-band tunnelling, as given by the compact model, are also indicated.

the two "source" terms, i.e. tunnelling and SRH generation, are subject to avalanche multiplication. The multiplication factor for the two terms is slightly different due to the fact that G_{SRH} is a homogeneously distributed generation term while the tunnelling generation term is localized. Breakdown occurs when the denominator of (3.38) becomes zero, so $2\mu_{av}(1 + e^{-2\mu_{av}}) = 1$. This gives for μ_{av} at breakdown a value of $\mu_{av,br} \simeq 0.3295$. Substitution of this result into (3.40) gives

$$\mu_{av} = 0.3295 \left(\frac{F_{max}}{F_{max,br}} \right)^2 \exp \left(\frac{|F_{max}| - |F_{max,br}|}{F_{max} F_{max,br}} \right). \quad (3.41)$$

Chapter 3
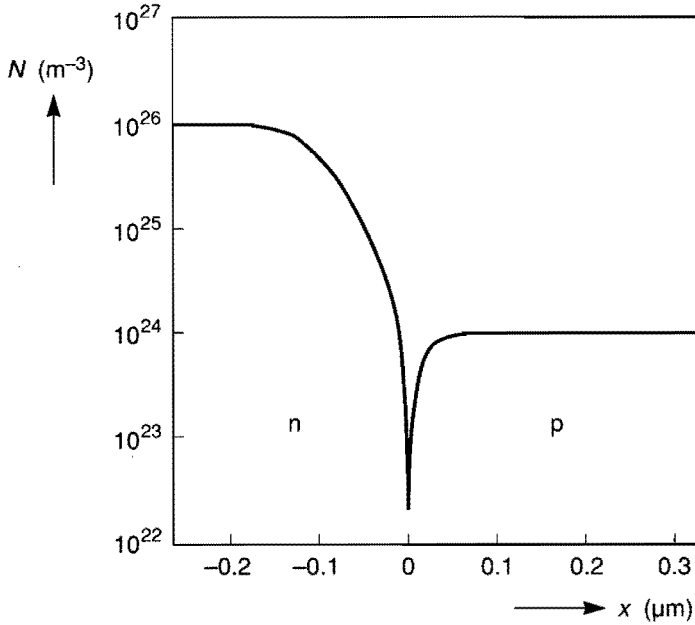
Fig. 3.24. The doping profile of the diode as used in the numerical simulations from fig. 3.23.

This expression has the advantages that the unknown d_{av} is eliminated and that J_d becomes infinity at exactly the measured breakdown voltage.

Equation (3.38) together with (3.6), (3.36), (3.39) and (3.41) form the basis of the model. The strategy for obtaining the I-V characteristic is as follows:

1) Determine $W(V_j)$ and $F_{max}(V_j)$ from capacitance measurements as will be described in section 3.4.3.

2) Determine V_{br} from measurements and subsequently $F_{max,br}$ can be determined.

3) Determine $\mu_{av}(V_j)$ from (3.41).

Tunnelling effects in p-n junctions   *111*

4) Determine J_{bbt} from (3.6).

5) Determine J_s from the forward characteristic. For silicon this term is negligible in reverse bias at around room temperature.

6) The diode current density J_d can now be calculated from (3.38).

In this procedure the unknown constant G_{SRH} is obtained by fitting the calculations to the measurements.

In order to show the validity of the simplifications involved, fig. 3.23 presents a comparison of model calculations with results of 1D numerical simulations. In the numerical simulations, at each bias condition the quantities R_{SRH} and F are calculated at each mesh point in the discretized domain. Subsequently, expression (3.34) is evaluated numerically. The doping profile of this diode is given in fig. 3.24. The numerically calculated values of W and F_m at each bias point are used to obtain the compact model results. The quantities G_{SRH} and V_{br} are also taken from the numerical simulations.

3.4.3 Comparison with measurements

To test the physical significance of the resulting model we compared the model results with the measurements given in section 3.4.1. To this end the depletion layer width $W(V_j)$ and the electric field $F_{max}(V_j)$ have been obtained from measurements following a standard procedure as given, for instance, in [3.27]. The measured depletion capacitance is fitted to the following expression:

$$C(V_j) = \frac{C_0}{(1 - V_j / V_{diff})^p}, \qquad (3.42)$$

where the zero-bias depletion capacitance C_0, the grading coefficient p and the diffusion voltage V_{diff} are fitting parameters. For the three diodes p is measured as $\simeq 0.33$, i.e. linearly-graded junctions, while $V_{diff} = 0.5 - 0.6V$. By using Gauss' law and $dQ/dV = C$, we obtain [3.27]

$$|F_{max}(V_j)| = \frac{1}{A_j \, \varepsilon_0 \varepsilon_r} \int_{V_{diff}}^{V_j} C(V)dV \; + \; |F_{max}(V_{diff})|$$

$$\simeq \frac{C_0 V_{diff}}{A_j \, \varepsilon_0 \varepsilon_r \, (1-p)} \left(1 - \frac{V_j}{V_{diff}} \right)^{1-p}, \qquad (3.43)$$

where it is assumed that

$$F_{max}(V_{diff}) = 0. \qquad (3.44)$$
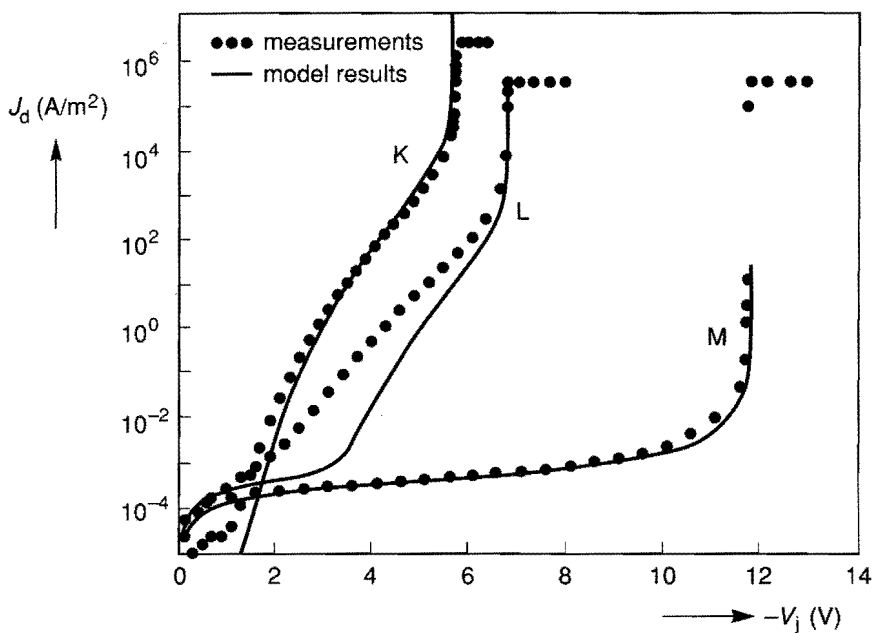
Fig. 3.25. Measured and calculated reverse characteristics at room temperature.

The quantity A_j is the effective junction area, i.e. the area of the region where the electric field is largest. In our case we have large planar junctions with sidewall effects eliminated by guard rings. However, in general the effective junction area will not be precisely known. For instance in the case of an emitter-base junction it is most probably the edges which determine the reverse characteristics and not the bottom part of the junction. It should be emphasized that, strictly speaking, $F_{max}(V_j)$ can be determined in this way only apart from an additional
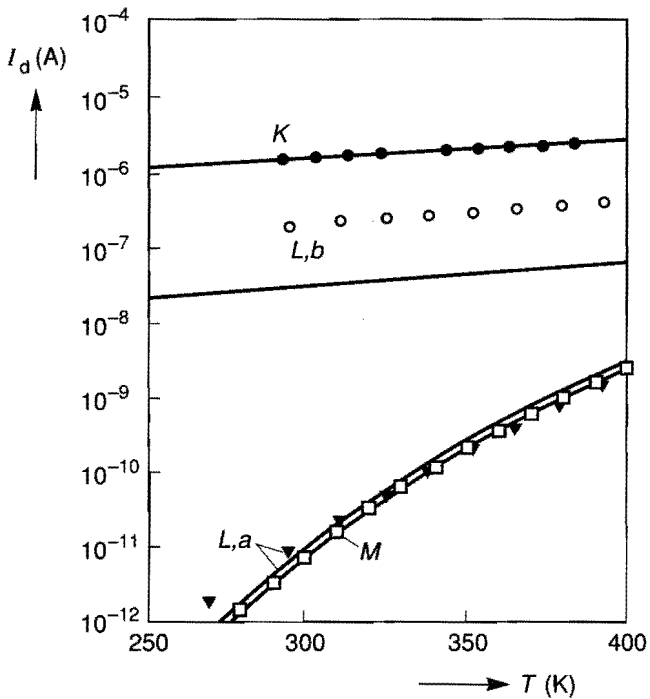
Fig. 3.26. The diode current I_d as a function of the temperature. For the diode K $V_j = 4V$ while for diode M $V_j = 6V$. Curve L,a is in the SRH generation regime ($V_j = 1V$), and curve L,b is in the tunnelling regime ($V_j = 5V$).
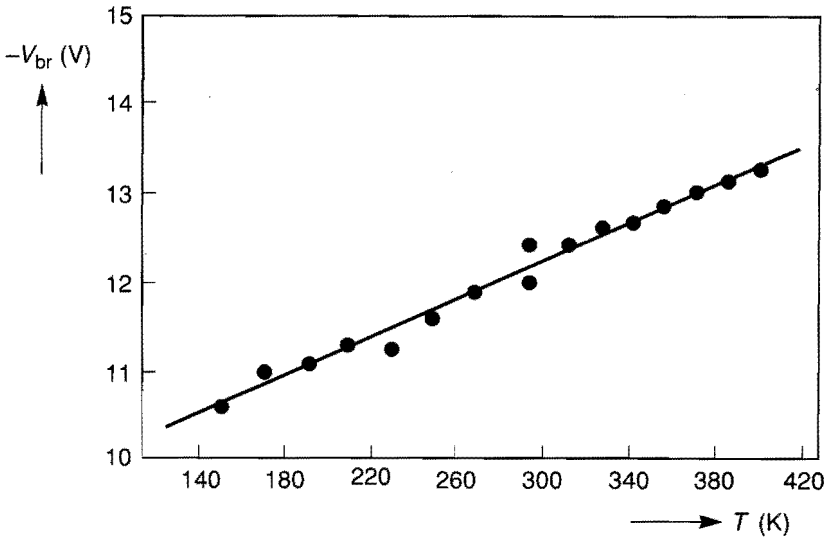
Fig. 3.27. The breakdown voltage of diode M as a function of the temperature. The straight line is a least-square fit.

constant $F_{max}(V_{diff})$, which is assumed to be negligible. The depletion layer width is given by

$$W(V_j) = \frac{A_j \, \varepsilon_0 \varepsilon_r}{C \, (V_j)} . \qquad (3.45)$$

Using this procedure, together with that described at the end of the previous section, we obtain the calculated results as shown in fig. 3.25. For the tunnelling constants we used the values given in section 3.2.2, while $b_n = 1.23 \, 10^8 \, V/m$. From fig. 3.25 we can observe that, despite the fairly large deviations between the model results and experiments at certain bias points, the overall shape of the measured characteristics is rather well reproduced by the model. Even the transition between SRH generation and tunnelling of diode L is reproduced, although the transition in the calculations is much steeper than measured. Possible

Tunnelling effects in p-n junctions

causes for the deviations between the model results and the measurements are discussed in section 3.5.

For the temperature dependence of the model we have assumed:

- The tunnelling current depends on temperature as given in section 3.2.2, i.e. only the bandgap is temperature-dependent.
- The quantity $n_i/\overline{\tau}$ has the temperature dependence of n_i (i.e. it is proportional to $T^{3/2}$ exp $(-E_g/2kT)$ for midgap states [3.1]).

In fig. 3.26 the current in the tunnelling regions of the diodes K and L at a certain voltage is plotted as a function of temperature. The solid lines are the theoretical curves according to the temperature rules as given above. The currents in the thermal generation regions of diodes L and M are also plotted. Note the clear distinction in temperature behaviour between the two regions, which is as it should be. In fig. 3.27 the measured breakdown voltage is plotted as a function of temperature. This breakdown voltage increases with temperature due to the increasing electron-phonon interaction with increasing temperature [3.1]. Note that the temperature dependence of the breakdown voltage can be approximated reasonably well by a straight line.

3.5 DISCUSSION AND CONCLUSIONS

In section 3.2 it is shown experimentally that the conventional expression for Zener tunnelling can be used as a model for tunnelling currents in reverse-biased p-n junctions, provided the maximum electric field at the junction is used in that expression. Theoretically this finding is made plausible by taking the actual behaviour of the electric field near the junction into account in the derivation of the dependence of the tunnelling current on the applied junction voltage. The resulting expression (3.8) shows roughly the same dependence of the current density on the maximum electric field as the conventional model (3.6). However, expression (3.8) is less suitable for implementation in a numerical device simulation package and for compact modelling purposes, because:

- Expression (3.8) involves the determination of one extra parameter, i.e. the doping gradient a.

- Expression (3.8) does not cover the situation of a mathematically abrupt junction ($a = \infty$), which may occur in a simulation.

The values of the parameters c_{bbt} and F_0 obtained experimentally are in rough agreement with the theory and give a reasonable fit for a wide range (more than eight decades) of currents and for different types of junctions.

Expression (3.6) is used in the compact diode model, as described in section 3.4. It is also implemented in a two-dimensional device simulation package [3.28] as a post-processing routine. In this routine the maximum electric field around the metallurgical junction is determined in each mesh of the grid and subsequently J_{bbt} is determined. The total tunnelling current (per unit length in the third dimension) is obtained by the multiplication of the current density by the corresponding length of the segment, followed by a summation over all segments. The above post-processing procedure is justified only when the tunnelling currents are so low that they do not significantly influence the solution of the Poisson problem from which the electric field is calculated. If this is not the case, the band-to-band tunnelling effect must be taken into account as a recombination term in the continuity equations (1.2), which are solved together with Poisson's equation. As pointed out in section 3.4.1, also in cases where avalanche multiplication of electron-hole pairs generated by tunnelling is important, a description in terms of a current density is unsuitable. In these cases, too, the tunnelling effect should be taken into account implicitly by a recombination term in the current continuity equations.

Band-to-band tunnelling can be formulated as a recombination rate as follows: Using the relations $dE = -q\,d\psi$ and $\underline{F} = -\underline{\nabla}\psi$, expression (3.5) for dJ_{bbt}/dE can be transformed into a recombination rate, defined by

$$R_{bbt} \equiv -\frac{1}{q}\,\underline{\nabla}.\underline{J}_{bbt}. \qquad (3.46)$$

This gives

$$R_{bbt} = -\frac{1}{q} \left(\frac{d\underline{J}_{bbt}}{d\psi} \cdot \underline{\nabla}\psi \right) = -\left(\frac{d\underline{J}_{bbt}}{dE} \cdot \underline{F} \right). \qquad (3.47)$$

Substitution of (3.5) into the above expression gives

$$R_{bbt} = -c \, |F|^{\sigma+1} D \,(F,E,\,V_j) \, \exp \,(-F_0/|F|). \qquad (3.48)$$

An expression for $D \,(F,E,V_j)$ which is suitable for implementation in a device simulator can be obtained from [3.8]. This gives

$$D = \frac{1}{\exp[q \,(V_p - \psi)/kT] + 1} - \frac{1}{\exp[q \,(V_n - \psi)/kT] + 1}, (3.49)$$

where V_p and V_n are the applied voltages at the contacts of the p and n regions, respectively $(V_j = V_p - V_n)$. This function approximately equals 1 for $V_j < < -kT/q$ and -1 for $V_j > > kT/q$, while for zero bias it equals zero. In the above transformation from $d\underline{J}_{bbt}$ to R_{bbt} tunnelling of an electron at a certain energy, say E_1 (see fig. 3.4), between x_1 to x_2 is represented by the generation of an electron-hole pair in the middle of the gap $(x_1 + x_2)/2$.

The proposed model for recombination including forward-biased tunnelling, presented in section 3.3, effectively describes the characteristic features associated with the non-ideal diode current of highly-doped devices. These features are (a) a temperature dependence of the current which, on the one hand, is much weaker than that predicted by the conventional SRH recombination model and, on the other hand, is definitely stronger than that of band-to-band tunnelling and (b) a non-ideality factor m which is larger than two and almost inversely proportional to the temperature. As already pointed out in section 3.3 these effects cannot be due to an erroneous assumption for the recombination lifetimes. However, the absolute value of the non-ideal current is not always predicted accurately by the model, as shown in fig. 3.14. From this figure it is also clear that great differences occur between the measurement results from different processes.

The major advantages of the proposed model over the existing models [3.2,3.13] are:

- The proposed model, unlike the existing models, basically contains no unknown fitting parameters such as c_{tat} in (3.9). However, the exact value of the effective mass to be used is not known. Although the overall features of the model do not significantly change within the range of physical meaningful values of $m^{\times}$, the exact value of the current density at high doping concentrations depends fairly strongly on $m^{\times}$. This is shown in fig. 3.12. After more or less fitting the value of $m^{\times}$ to the measurements we arrive at a value of around $0.15m_0$. This is in reasonable agreement with the values obtained experimentally for band-to-band tunnelling [3.6,3.7,3.20] .

- The proposed model comprises both tunnelling effects and conventional SRH recombination, and the gradual transition between the two regimes (around 300Å zero-bias depletion width), both in bias dependence and in temperature dependence, is automatically obtained. The existing models contain only the tunnelling term.

- Since these effects occur in the forward-bias regime, the tunnelling effect must be incorporated into a numerical device simulator in the form of a recombination term in the continuity equations. A model formulated in terms of a current density is not suitable. However, it must be remarked that the proposed model in the form of expressions (3.22) and (3.23) is also not readily suitable for implementation because these expressions not only contain local variables, but also non-local variables such as the average electric field and an integral over the depletion layer.

It is found that irrespective of the exact value of the effective mass, the tunnelling effect increases dramatically when the zero-bias depletion width is less than about 300Å. In the case of a two-sided abrupt junction this value corresponds to a doping level of $3 \, 10^{24} m^{-3}$. This is in reasonable agreement with experiments [3.13].

Considering the compact diode model described in section 3.4, we have seen that, despite the fairly rigorous assumptions, the model results agree quite well with the numerical simulations. The overall shape of the measured characteristics is also reasonably well reproduced by the model. However, especially at low bias there are quite large dis-

parities between the measurements and the calculations. There might be several reasons for this.

The first one is that the electric field depends on the diffusion voltage (see eq. (3.43)). The measured values of V_{diff} are $0.5 - 0.6V$. These values are rather low for such a highly-doped junction. In fact, they are much lower than those obtained from numerical calculations on this type of junction, which yield $0.9V < V_{diff} < 1.0V$ for such highly doped devices. Consequently, the electric field obtained from capacitance measurements might be lower than the actual electric field, which would result in an underestimation of the tunnelling current. Obviously, this effect is most important at low bias. A horizontal shift to the left of around $0.4V$ of curve L in fig. 3.25 would indeed bring the calculations more into agreement with the measurements.

Another possible cause for the deviations at low bias, where SRH generation dominates over band-to-band tunnelling, is the fact that we have neglected electric field-enhanced emission from traps which causes the generation lifetimes to decrease with increasing electric field. The physical origin of this effect is as follows: The emission probability from traps is proportional to $\exp(-E_b/kT)$ where E_b is an energy barrier for electrons or holes to be emitted from the the trap. Due to high electric fields, the effective emission barrier can be lowered [3.29]. This may be either an actual barrier lowering due to the so-called Poole-Frenkel effect for a Coulomb trap, or a virtual barrier lowering due to phonon-assisted tunnelling from a trap. In [3.30] it is shown that these effects can be effectively modelled by an additional trap-assisted tunnelling current density of the form

$$J_{tat} = qG_{tat}W \frac{\gamma}{|F_{max}|} \left[\exp\left((\frac{F_{max}}{\gamma})^2 \right) - \exp\left((\frac{F_{max}W_0}{\gamma W})^2 \right) \right]. \quad (3.50)$$

The quantity G_{tat} depends on the intrinsic carrier concentration and carrier lifetime, while γ depends on the temperature and the effective mass. Similarly to the band-to-band tunnelling effect, in [3.30] the trap-assisted tunnelling effect is taken into account as a $\delta-$ function generation term, because the corresponding electron-hole pairs are generated in a narrow high-field region around the junction. This effect

yields an enhancement of the increase in the SRH generation current with increasing bias and explains the gradual transition measured between the SRH generation regime and the tunnelling regime of diode L.

For practical use of the diode model the problem arises that, strictly speaking, the electric field can be obtained from capacitance measurements only for a one-dimensional problem. For large planar junctions without sidewall effects this is not expected to be a problem. However, for junctions where for some reason the electric field along this junction is not constant, the tunnelling and avalanche currents will mainly flow in those regions where the electric field is highest. The tunnelling current density is an especially sensitive function of the electric field. For instance, at an emitter-base junction in a bipolar transistor the electric field will be highest at the sidewalls, because there the base doping concentration is higher than at the bottom. Moreover, the field will not be constant at the sidewall, but will vary continuously along the junction. In such cases it is impossible to obtain the electric field distribution from capacitance measurements. A common practice then is to treat the most sensitive parameters which depend on the magnitude of the electric field, such as the ratio F_0/F_{max}, as fitting parameters, i.e. the voltage dependence is obtained from capacitance measurements while the magnitude is fitted. This can be accomplished by treating the effective junction area A_j in expression (3.43) for the electric field $F_{max}(V_j)$ as a fitting parameter, i.e. the voltage dependence of the electric field is obtained from capacitance measurements while the magnitude, determined by the zero-bias capacitance per unit area c_0 ($c_0 = C_0/A_j$) in (3.43), is obtained from the I-V curve. Other parameters which must be treated as fitting parameters are G_{SRH} and G_{tat}, both depending on the intrinsic carrier concentration and the recombination lifetime in the depletion layer. In particular, the value of the lifetime which not only depends on the doping concentration but also on process characteristics, is usually unknown. As an example, fig. 3.28 shows the result of such a fitting procedure [3.30]. The trap-assisted tunnelling current has been taken into account as a δ − function generation term and is added to J_{bbt} in expression (3.38) for the diode current. The following parameters are used: the prefactors G_{SRH}, G_{tat}
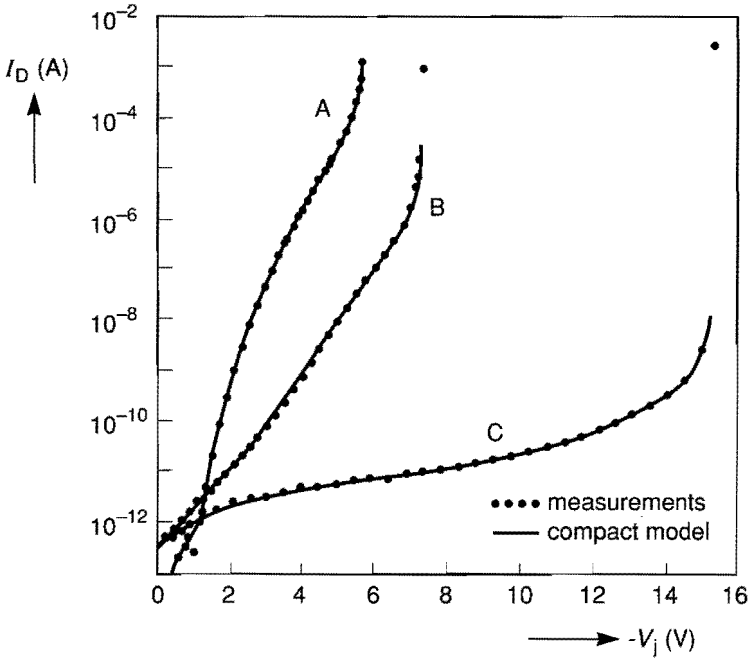
Fig. 3.28. A comparison of model results with measurements on three diodes. In these cases the effective junction areas are treated as fitting parameters.

and c_{bbt}, the capacitance c_0 and V_{br} (all from the reverse characteristics). The saturated ideal current I_s is obtained from the forward characteristic, while the grading coefficient p and the diffusion voltage V_{diff} are obtained from the capacitance measurements. Note that in this case the diode current is fitted and not the current density, so the prefactors also comprise the effective junction area. The resulting values of the parameters G_{SRH} and G_{tat} yield a value of τ of around $5\mu s$, both for diodes B and C. The resulting effective junction areas are $3.53 \ 10^{-8} m^2$, $2.43 \ 10^{-8} m^2$ and $2.72 \ 10^{-8} m^2$ for diodes A, B and C respectively, while the macroscopic junction areas are $4.16 \ 10^{-8} m^2$ for A and $3.24 \ 10^{-8} m^2$ for diodes B and C. The saturated ideal current has proved to be unimportant in reverse bias at room temperature.

REFERENCES

3.1 S.M. Sze, Physics of Semiconductor Devices, New York: Wiley, p. 513 (1981)

3.2 A.G. Chynoweth, W.L. Feldman and R.A. Logan, Phys. Rev., 121, p. 684 (1961)

3.3 J.L. Moll, Physics of Semiconductors, New York: McGraw-Hill, p. 249 (1964)

3.4 E.O. Kane, J. Phys. Chem. Solids, 12, p. 181 (1959)

3.5 L.V. Keldysh, Soviet Phys. JETP, 34, p. 665 (1958)

3.6 R.B. Fair and H.W. Wivell, IEEE Trans. Electron Devices, ED-23, p. 512 (1976)

3.7 M. Singh Tyagi, Solid-St. Electron., 11, p. 99 (1968)

3.8 E.O. Kane, J. Appl. Phys., 32, p. 83 (1961)

3.9 E.O. Kane, Phys. Rev., 131, p. 79 (1963)

3.10 J.M.C. Stork and R.D. Isaac, IEEE Trans. Electron Devices, ED-30, p. 1527 (1983)

3.11 T. Yajima and L. Esaki, J. Phys. Soc. Japan, 13, p. 1281 (1958)

3.12 T.A. Longo and A.G. Chynoweth, Bull. Am. Phys. Soc., 5 , p. 160 (1960)

3.13 J.A. Del Alamo and R.M. Swanson, 18th Conf. on Solid State Devices and Materials, Tokyo, p. 283 (1986)

3.14 M. Singh Tyagi and R. van Overstraeten , Solid-St. Electron., 26, p. 577 (1983)

3.15 R.R. Blair and J.W. Easley, J. Appl. Phys., 31, p. 1772 (1960)

3.16 W. Franz, Handbuch der Physik, Berlin: Springer, 17 p. 155 (1956)

3.17 J.C. Slater, Phys. Rev., 76, p. 1592 (1949)

3.18 R.A. Smith, Wave Mechanics of Crystalline Solids, London: Chapman and Hall Ltd, p. 403 (1969)

3.19 M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, New York: Dover Publications, p.260 (1970)

3.20 R.A. Logan and A.G. Chynoweth, Phys. Rev., 131, p. 89 (1963)

3.21 S. Selberherr , Analysis and Simulation of Semiconductor Devices, Wien: Springer, p. 106 (1984)

3.22 J.W. Slotboom and H.C. de Graaff, Solid-St. Electron., 19, p. 857 (1976)

3.23 PHILPAC User Manual, Philips CFT Automation

3.24 SPICE User Manual, Berkeley University

3.25 P.U. Calzolari and S. Graffi, Solid-St. Electron., 15, p. 1003 (1972)

3.26 R.J. van Overstraeten and H.J. de Man, Solid-St. Electron., 13, p. 583 (1970)

3.27 H.C. de Graaff, Compact Bipolar Transistor Modelling, in W.L. Engl, ed., Process and Device Modeling, Amsterdam: North-Holland, p. 413 (1986)

3.28 CURRY User Manual, Philips CFT Automation

3.29 G. Vincent, A. Chantre and D. Bois, J. Appl. Phys., 50 , p. 5484 (1979)

3.30 G.A.M. Hurkx, H.C. de Graaff, W.J. Kloosterman and M.P.G. Knuvers, to be published in Proc. ESSDERC'90, Nottingham

APPENDIX 3A

In order to integrate (3.5) analytically, we replace dE by $-q\,d\psi$ and subsequently transform it to an integration over F (the electric field is taken to be positive). The desired relation between $d\psi$ and dF can be obtained by expanding $F(x)$ around its maximum:

$$F(x) = F_{max}\left[1 + \frac{x^2}{2\,F_{max}}\left(\frac{d^2F}{dx^2}\right)_{x=0} + \ldots \right]. \qquad (3A.1)$$

We truncate the above series after the second-order term. By differentiation of Poisson's equation and using $F = -d\psi/dx$, we obtain after some manipulation:

$$\frac{d\psi}{dF} = \pm \sqrt{\frac{\varepsilon_0\,\varepsilon_r}{2\,a\,q}\,\frac{F^2}{F_{max} - F}}, \qquad (3A.2)$$

where the $+$ sign holds in the region $x > 0$ and the - sign holds for $x < 0$. Substitution of (3A.2) into (3.5) gives:

$$J_{bbt} = -c\,q\,\sqrt{\frac{2\,\varepsilon_0\,\varepsilon_r}{q\,a}} \int_{F_1}^{F_{max}} \frac{F^{\sigma+1}}{\sqrt{F_{max} - F}}\; e^{-F_0/F}\,dF, \qquad (3A.3)$$

where the lower bound F_1 must be sufficiently far from F_{max}. By the substitution of

$$y \equiv \sqrt{1 - \frac{F}{F_{max}}}, \qquad (3A.4)$$

(3A.3) becomes

$$J_{bbt} = -c\,q\,\sqrt{\frac{8\,\varepsilon_0\,\varepsilon_r}{q\,a}}\; F_{max}^{\sigma+3/2} \int_0^{y_1} (1 - y^2)^{\sigma+1} \exp\left[-\frac{F_0}{F_{max}(1 - y^2)} \right] dy,$$

$$(3A.5)$$

where y_1 must be sufficiently far from 0. Since the largest contribution to the integral comes from the region where $y < < 1$, and because the integrand decreases strongly with increasing y, we can use Taylor series expansions to evaluate (3A.5). By using $(1 - y^2)^{-1} \simeq 1 + y^2$ and $(1 - y^2)^{\sigma+1} \simeq 1 - (\sigma + 1) y^2$ for $y < < 1$ and subsequently putting $y_1 = \infty$, eq. (3A.5) is transformed into a sum of two standard integrals and can be evaluated to yield (3.8).

APPENDIX 3B

To investigate the validity of approximating $N(E) f(E) \Delta E$ by $- (dn_{sc}/dx_1) \Delta x_1$, we write (see fig. 3B.1)

$$n_{sc}(x_1) = \int_E^\infty N(E' - E) f(E') \, dE', \qquad (3B.1)$$

and

$$n_{sc}(x_1 + \Delta x_1) = \int_{E+\Delta E}^\infty N(E' - E - \Delta E) f(E') \, dE'. \qquad (3B.2)$$

From the above expressions it follows that

$$-\frac{dn_{sc}}{dx_1} \Delta x_1 = \int_E^\infty N(E' - E) f(E') \, dE'$$

$$- \int_{E+\Delta E}^\infty N(E' - E - \Delta E) f(E') \, dE'. \qquad (3B.3)$$
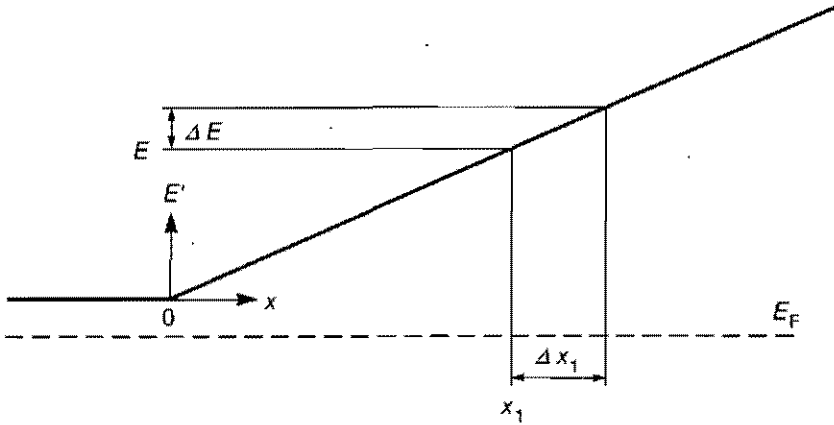
Fig. 3B.1. A graphical representation of the transformation from the integration variable E to the variable x_1.

When we write $\varepsilon = E' - \Delta E$ and expand

$$f(\varepsilon + \Delta E) = f(\varepsilon) + \Delta E \frac{df}{d\varepsilon}, \tag{3B.4}$$

(3B.2) becomes

$$n_{sc}(x_1 + \Delta x_1) = \int_E^\infty N(\varepsilon - E)\left[f(\varepsilon) + \Delta E \frac{df}{d\varepsilon}\right]d\varepsilon. \tag{3B.5}$$

After substitution of this expression into (3B.3) we obtain

Tunnelling effects in p-n junctions
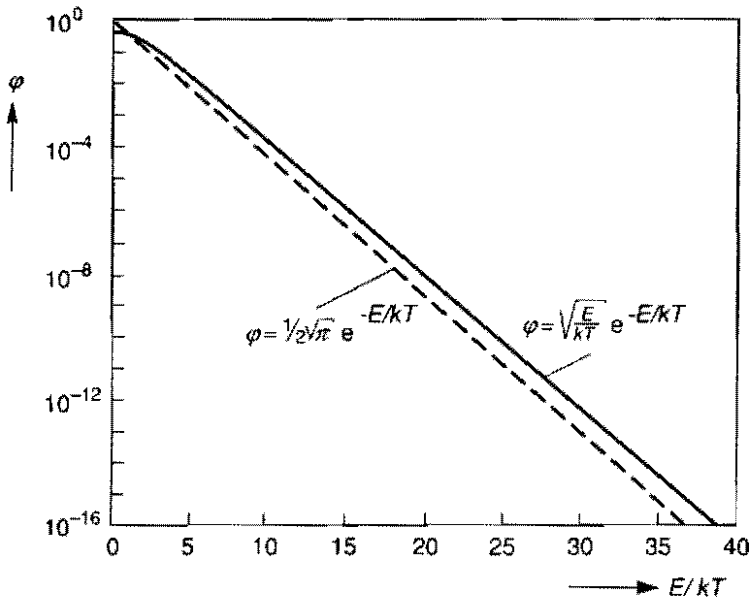
Fig. 3B.2. A comparison of the function values F given by (3B.8) and (3B.9). The solid line gives the function to be approximated and the dashed line gives the approximation. The prefactor $c\,(kT)^{1/2}\exp\,(E_F/kT)\Delta E$ has been put equal to one in both cases.

$$-\frac{dn_{sc}}{dx_1}\,\Delta x_1 = -\,\Delta E\int_E^\infty N(\varepsilon - E)\,\frac{df}{d\varepsilon}\,d\varepsilon. \tag{3B.6}$$

When we write $\varepsilon - E = y$, (3B.6) becomes

$$-\frac{dn_{sc}}{dx_1}\,\Delta x_1 = -\,\Delta E\int_0^\infty N(y)\,\frac{df\,(y + E)}{dy}\,dy. \tag{3B.7}$$

Assuming the electron gas is not degenerated, which is a good assumption in the depletion layer, $f(y + E) = e^{-(y+E-E_F)/kT}$. Using $N(y) = c\sqrt{y}$, (3B.7) becomes

$$-\frac{dn_{sc}}{dx_1}\Delta x_1 = c\sqrt{kT}\, e^{-(E-E_F)/kT}\Delta E \int_0^\infty \sqrt{y/kT}\, e^{-y/kT} d\,(y/kT)$$

$$= c/2\sqrt{\pi kT}\, e^{-(E-E_F)/kT}\Delta E. \tag{3B.8}$$

On the other hand

$$N(E)f(E)\Delta E = c\sqrt{E}\, e^{-(E-E_F)/kT}\Delta E. \tag{3B.9}$$

According to our assumption the right-hand side of (3B.9) is approximated by the right-hand side of (3B.8). From a comparison of these expressions it is obvious that, apart from a different prefactor, the exponential dependence on E/kT is equal. In fig. 3B.2 both functions are plotted versus E/kT. In practical cases the energy interval is from 0 to $q(V_{diff} - V_j)$ and therefore the interval for E/kT at room temperature is from 0 to 30-40. We see that the energy dependence of $N(E)f(E)\Delta E$ is reasonably well described by $-dn_{sc}/dx_1\,\Delta x_1$.

APPENDIX 3C

In order to arrive at (3.38) we rewrite α_n in (3.37) as a function of x. We expand the expression for the electrical field around a symmetrical step junction around its maximum

$$F(x) = F_{max}\left(1 - \left|\frac{qNx}{F_{max}\,\varepsilon_0\varepsilon_r}\right|\right), \tag{3C.1}$$

where N is the dopant concentration. Substitution of (3C.1) in (3.37) gives after a first-order expansion of the exponent

$$\alpha_n \simeq \alpha_{n\,\infty} \exp\left(\frac{-b_n}{F_{max}} - \frac{|x|}{d_{av}} \right), \tag{3C.2}$$

with

$$d_{av} = \frac{F_{max}^2\, \varepsilon_0 \varepsilon_r}{qNb_n}. \tag{3C.3}$$

In (3.35) the following integral occurs

$$I(x) = \int_x^{x_p} (\alpha_n - \alpha_p)\, dx'. \tag{3C.4}$$

When we consider α_p to be proportional to α_n, i.e. $\alpha_p = s\,\alpha_n$, substitution of (3C.2) into (3C.4) gives

$$I(x) = \mu_{av} \left(2 - e^{\,x/d_{av}}\right) \qquad x < 0, \tag{3C.5}$$

$$= \mu_{av}\, e^{-x/d_{av}} \qquad x \geq 0,$$

where μ_{av} is given by

$$\mu_{av} = (1-s)\, d_{av}\, \alpha_{n\,\infty} \exp\left(-b_n/|F_{max}|\right). \tag{3C.6}$$

Because usually $d_{av} \ll W$ (for a two-sided abrupt junction $d_{av}/W = F_{max}/2b_n$), we can approximate

$$I(x) \simeq 2\mu_{av} \qquad x < 0, \tag{3C.7}$$

$$\simeq \mu_{av} \qquad x = 0,$$

$$\simeq 0 \qquad x > 0.$$

When we assume the junction to be symmetrical, substitution of (3C.7) into (3.35) and subsequent integration of (3.34) gives

$$J_d = \frac{J_{bbt}\, e^{-\mu_{av}} + (J_{SRH} + J_s)\,(1 + e^{-2\mu_{av}})/2}{1 - \dfrac{\mu_{av}}{(1 - s)}\,(1 + e^{-2\mu_{av}})}. \qquad (3C.8)$$

Assuming $s = 1/2$, we arrive at (3.38).

Chapter 4

MODELLING OF SMALL-SIGNAL A.C. TRANSISTOR PROPERTIES

4.1 INTRODUCTION

The small-signal a.c. behaviour of bipolar transistors is often described by means of a quasi-static analysis of the currents and charges in the transistor. This type of approach was initiated in 1957 by Beaufoy and Sparkes [4.1], mainly as a new way to describe the large-signal transient behaviour of a bipolar transistor. The essence of this method is the relationship it makes between the terminal currents and the stored d.c. charge in the device. One of the most important results of using this quasi-static approach is that the small-signal a.c. common-emitter current gain can be approximated by

$$h_{FE} = \frac{\tilde{i}_c}{\tilde{i}_b} = \frac{h_{FE0}}{1 + j\omega h_{FE0}\,\tau_{qs}}, \tag{4.1}$$

where the transit time τ_{qs} is given by

$$\tau_{qs} = \left(\frac{dQ}{dI_c}\right)_{V_0}. \tag{4.2}$$

In (4.1), h_{FE0} is the low-frequency current gain ($h_{FE0} = dI_c/dI_b$) and ω is the angular frequency. In (4.2) Q is the total integrated d.c. hole or electron charge in the device at a certain d.c. bias condition denoted by V_0. So, for the a.c. current gain the device can be considered as a first-order system and the characteristic parameters h_{FE0} and τ_{qs} can be obtained from a static (d.c.) analysis. The quasi-static common-emitter cut-off frequency f_{Tqs} is related to τ_{qs} by $f_{Tqs} = (2\pi\tau_{qs})^{-1}$. As can be seen from (4.1), f_{Tqs} is that frequency at which the current gain approaches unity, provided that the d.c. current gain h_{FE0} is much larger than unity,

Modelling of small-signal a.c. transistor properties   *133*

i.e. $f_{T_{qs}}$ is obtained by extrapolation of the relation $\omega |h_{FE}| \tau_{qs} = 1$ to the unity-gain frequency.

This quasi-static result is often used both in numerical device simulations and in compact modelling. Since $f_{T_{qs}}$ is a very important figure of merit used in the characterization of the high-frequency behaviour of bipolar devices, relation (4.2) is often used to calculate $f_{T_{qs}}$ from a static numerical device simulation [4.2, 4.3]. This is done by a post-processing routine which calculates $f_{T_{qs}}$ by determining dQ/dV and dI_c/dV at a certain bias condition. The total transit time and, therefore, the cut-off frequency then follows from (4.2). Moreover, the total charge dQ/dV is often split up into contributions from the different quasi-neutral regions and the space-charge regions in the device, simply by the integration of dn/dV and dp/dV over the corresponding volume [4.5]. Therefore the total transit time can also be decomposed into contributions from different regions (the so-called regional approach). This offers the possibility to investigate numerically the contributions of the different transistor regions to the cut-off frequency and to reveal the relationships between process parameters (junction depths, doping profiles) and the small-signal a.c. behaviour of the device [4.2, 4.3].

In compact modelling we can distinguish between two basically different approaches. In one approach, followed in the compact model Mextram [4.5], the collector current and base minority charge are modelled independently, assuming a certain type of doping distribution in the base. The base transit time follows from the relation $\tau_b = dQ_b/dI_c$, and is not necessarily constant. The other approach, followed for instance in the well-known Gummel-Poon model, is based on the so-called Charge-Control concept which, in turn, is also based on the quasi-static approximation leading to (4.1) and (4.2). However, in the Charge-Control concept an additional assumption is made which finds its origin in the work of Beaufoy and Sparkes [4.1]. They showed that under low injection conditions a linear relation exists between the d.c. collector current and the d.c. base minority charge, viz. $Q_b/I_c = $ constant $\equiv \tau_b$, and they used this relation, together with the quasi-static assumption, to describe the large-signal transient behaviour of a bipolar transistor. The essence of the Charge-Control concept is that, once the transit time in the quasi-neutral regions is obtained (e.g. from transient

or a.c. measurements), the relation between the total minority charge in a quasi-neutral region and the terminal current is at each bias point given by this transit time. In the Gummel-Poon model such a constant transit time is used to calculate the total minority charge in the quasi-neutral regions from the collector current by using $Q_{min} \equiv \tau_f I_c$, where the transit time τ_f is a model parameter.

In this chapter we will only deal with the quasi-static appproach leading to (4.1) and (4.2) and not with the validity of the additional assumption of a linear relation between the charge and the current, as is made in the Charge-Control concept.

Although an outline of the theoretical basis of the quasi-static method has been given in many papers (see for instance [4.4,4.6]), a straightforward derivation of (4.1) and (4.2), together with a clear indication of the assumptions involved, is rarely discussed. Moreover, most treatments of the foundations of this method date from more than a decade ago, when the dimensions of the devices were such that the maximum cut-off frequency was almost entirely determined by the base region. As a consequence, most of these treatments are confined to the base region. Due to downscaling, however, nowadays not only the base region but also other regions contribute significantly to the cut-off frequency [4.7]. In view of this downscaling it therefore seems interesting to revisit the foundations of this quasi-static approach and to analyse the validity of this concept for downscaled devices.

In section 4.2 the quasi-static expressions for the small-signal a.c. base and collector currents are derived from the time-dependent continuity equations. The assumptions are also discussed. A different treatment of the small-signal a.c. currents, based on a perturbation expansion of the time-dependent continuity equations in one dimension, is given in section 4.3 in order to investigate the validity of the approximations made in the conventional quasi-static analysis. As will be shown the results of this analysis can be used to remedy two of the shortcomings of the quasi-static approach. Based on the results of the analysis in sections 4.2 and 4.3, the use of the quasi-static $f_{T_{qs}}$ (from eq. (4.2)) as a figure of merit for the characterization of high-frequency properties of advanced devices is investigated in section 4.4. In addi-

tion, the results are compared with exact numerical a.c. solutions of the continuity equations for some representative downscaled devices.

4.2 DERIVATION OF THE QUASI-STATIC RELATIONS

We consider a vertical n-p-n transistor as sketched in fig. 4.1. The currents at the four contacts are also denoted in this figure. Also indicated is the recombination current at the silicon-oxide interface for which, of course, the hole and electron current densities are equal. The time-dependent continuity equation for the majority carriers in the base of a n-p-n transistor reads (cf. expression (1.2b))

$$\frac{\partial p}{\partial t} + R + \frac{1}{q} \nabla \cdot \underline{J}_p = 0, \tag{4.3}$$

where p is the hole concentration, $\underline{J}_p$ is the hole current density and $R(\underline{r}, t)$ is the net recombination rate. After integration of the above equation over the volume of the device and subsequent application of Gauss' law we obtain

$$\int_{Vol} \frac{\partial p}{\partial t} \, d\underline{r} + \int_{Vol} R \, d\underline{r} + \frac{1}{q} \int_{Surf} \underline{J}_p \cdot d\underline{a} = 0. \tag{4.4}$$

The last term on the left-hand side of the above equation is the total hole flow at the contacts (including recombination at the oxide-silicon interface). When we define $Q_p(t)$ as the total integrated hole charge in the device we obtain from (4.4), after interchanging the time derivative and the volume integral, the following first-order differential equation

$$\sum_m I_{pm} = \frac{dQ_p}{dt} + q \int_{Vol} R \, d\underline{r}. \tag{4.5}$$

In the above equation I_{pm} is the hole current through contact m which is defined positive for an inflowing current. So, by integration of the continuity equation the functional dependence on the spatial coordinates is removed and (4.5) is obtained, giving a relation between the sum of the terminal hole currents and the stored hole charge in the device.

Up to now no approximation has been made and (4.5) is generally valid. However, in order to obtain a useful result from (4.5) the following essential approximation is made:

$$p\,(\underline{r},\,t) = p\,(\underline{r},\,V\,(t)\,),\qquad\qquad(4.6)$$

and similarly for the electron concentration. This means that the concentrations instantaneously follow the applied voltage $V(t)$ without any
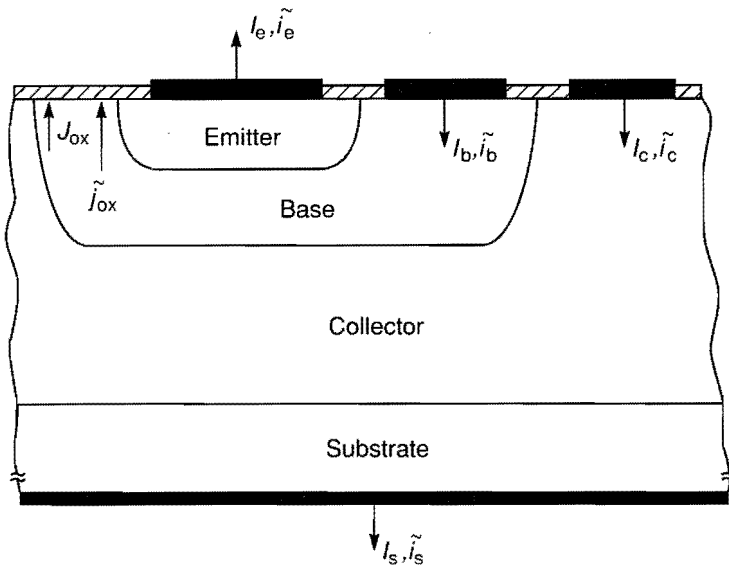

Fig. 4.1. Schematic cross-section of a vertical n-p-n transistor.

time delay or, equivalently, without any phase shift. This is the so-called quasi-static approximation. Since the recombination rate R only varies with time through the electron and hole densities, we can also write $R(\underline{r}, t) = R(\underline{r}, V(t))$. Equation (4.5) now becomes

$$\sum_m I_{pm}(t) = \left(\frac{dQ_p}{dV} \frac{dV}{dt} \right)_{V_0} + q \int_{Vol} R(\underline{r}, V(t)) \, d\underline{r}. \qquad (4.7)$$

In the above equation V_0 represents the d.c. bias condition of the device. Using (4.7), it is possible, in principle, to determine the time-dependent sum of hole currents from a static analysis since both dQ_p/dV and $R(\underline{r}, V(t))$ can be determined from a static simulation. Now we turn to a small-signal a.c. analysis, i.e. the applied voltage varies with time as

$$V(t) = V_0 + \tilde{v} \, e^{j \omega t}, \qquad (4.8)$$

where V_0 is the d.c. bias condition. The a.c. amplitude $\tilde{v}$ is taken to be very small, i.e. $q\tilde{v}/kT << 1$, so we can apply a first-order Taylor expansion:

$$p(\underline{r}, V(t)) = p(\underline{r}, V_0) + \left(\frac{\partial p}{\partial V} \right)_{V_0} \tilde{v} \, e^{j \omega t}, \qquad (4.9)$$

and similarly for $n(\underline{r}, V(t))$ and $R(\underline{r}, V(t))$. Substitution of these expansions into (4.7) and subtracting the d.c. solution ($\omega = 0$) we obtain the following expression for the amplitude of the a.c. hole terminal currents

$$\sum_m \tilde{i}_{p_m} = \tilde{v} \sum_m \left(\frac{dI_{p_m}}{dV} \right)_{V_0} + j \omega \tilde{v} \left(\frac{dQ_p}{dI_c} \frac{dI_c}{dV} \right)_{V_0}. \qquad (4.10)$$

As can be seen from the above equation the amplitude of the a.c. currents consists of both a real and an imaginary part. The real part corresponds to a conductance, while the imaginary part corresponds to a capacitance, i.e. the stored charge in the device.

It is interesting to note that, in fact, the quasi-static assumption leads to an inconsistency, which can be seen as follows: As mentioned above, from (4.10) it can be seen that the sum of the amplitudes of the a.c. hole current through the contacts has an imaginary term, corresponding to the charge storage. On the other hand, we have assumed a quasi-static behaviour of the carriers (eq. (4.6)) from which it follows that the amplitude of the a.c. hole concentration is in phase with the a.c. voltage, i.e. the amplitude of the a.c. hole density has no imaginary component (see (4.9)). From the general drift-diffusion expression for the hole current density (e.g. eq. (1.3b)) it then follows that the hole current density has no out-of-phase term and therefore the sum of the amplitudes of the a.c. currents at the contact cannot have an imaginary part. This is in contradiction with (4.10). The origin of this contradiction lies in the fact that the quasi-static assumption for n and p is made after integration of the hole continuity equation over the device in order to take explicit account of the charge-storage term dQ_p/dt in the sum of the terminal currents, i.e. after eq. (4.5) is obtained.

Starting with the continuity equation for electrons and following a procedure similar to that above, we arrive at

$$\sum_m \tilde{i}_{nm} = \tilde{v} \sum_m \left(\frac{dI_{nm}}{dV} \right)_{V_0} - j\omega\tilde{v} \left(\frac{dQ_n}{dI_c} \frac{dI_c}{dV} \right)_{V_0}, \qquad (4.11)$$

where an inflowing current is again regarded as positive.

Up to now we have made only one essential assumption, viz. the quasi-static assumption. However, in order to arrive at the commonly used expression (4.1) together with (4.2), one more assumption must be made. This assumption is inherently related to the fact that in the quasi-static approach it is impossible to determine the current through a single contact; by the application of Gauss' law only the net hole or electron current through the boundary of a certain region is determined.

This means that we can only calculate the total hole or electron current associated with the stored minority charge in a certain region, but we cannot determine the distribution of the associated inflowing current over the various contacts of that region. Unfortunately, it is the total current through a single contact that we really are interested in, i.e. $I_b = I_{pb} + I_{nb}$ and $I_c = I_{pc} + I_{nc}$. So, in principle it is impossible to calculate the base and collector currents from a quasi-static analysis and we have to make an additional assumption about the supply of the stored charges in the device in order to arrive at (4.1). This assumption is that the stored hole charge in the device is supplied entirely through the base contact, while the stored electron charge in the device is supplied entirely through the emitter contact. So, except for I_{pb} and I_{ne} it is assumed for all currents that

$$\tilde{i} = \tilde{v} \left(\frac{dI}{dV} \right)_{V_0}, \qquad (4.12)$$

i.e. the a.c. current through the contact is in phase with the applied a.c. voltage. In that case, by writing

$$\sum I_{p_m} = (I_b - I_{nb}) + I_{pe} + I_{pc} + I_{ps} + I_{pox}, \qquad (4.13)$$

and similarly for the corresponding a.c. currents, we obtain from (4.10)

$$\tilde{i}_b = \tilde{v} \left(\frac{dI_b}{dV} \right)_{V_0} [1 + j\omega h_{FE0}\tau_{qs}]. \qquad (4.14)$$

In (4.13) I_{nb} is the electron current at the base contact and I_{pox} is the hole current due to recombination at the silicon-oxide interface. I_{pe}, I_{pc} and I_{ps} are the hole currents at the emitter, collector and substrate contacts, respectively.

It is interesting to notice that the condition under which (4.10) reduces to (4.14) is slightly less restrictive than that sometimes mentioned in the literature. For instance, in [4.4] and [4.8] it is stated that (4.10)

reduces to (4.14) only when the current at the base contact consists only of holes and that the currents at all the other contacts consist of electrons only. As we have seen above, it is sufficient that this condition holds only for the capacitive part of these currents. Since we have also assumed that

$$\tilde{i}_c = \tilde{v} \left(\frac{dI_c}{dV} \right)_{V_0}, \tag{4.15}$$

division of $\tilde{i}_c$ by $\tilde{i}_b$ yields the desired relation (4.1).

All the above-mentioned assumptions can be summarized by the following assumptions concerning the a.c. behaviour of the minority charges injected into the regions which are bounded by a forward-biased junction and either a reverse-biased junction (when the base region is considered) or a contact (when the emitter or collector region is considered):

1) The quasi-static assumption, i.e. the minority carrier density is in phase with the applied junction voltage (eq. (4.6)).

2) The stored minority charge is entirely supplied and reclaimed through the forward-biased junction, i.e. no imaginary current through the contact or the reverse-biased junction.

Recently, two proposals have been made in the literature [4.9,4.10] for dealing with the imaginary collector current. It is especially the phase shift between the a.c. junction voltage and the a.c. collector current, i.e. the so-called excess phase shift, which can be important in high-frequency applications. To this end Fossum and Veeraraghavan [4.9] derive the correcting term from a one-dimensional integration of the time-dependent electron current density over the base region of a n-p-n device, while Klose and Wieder [4.10] use a time-dependent version of Gummel's approach. Except for the case of a homogeneously-doped base these methods give different results, as will be shown below. Te Winkel [4.11] also obtained low-frequency results for the imaginary collector current analytically by using a Taylor-series expansion of the exact solution of the continuity equation for an exponentially doped base.

In the next section the validity of the two above-mentioned assumptions is investigated analytically. To this end a perturbation-expansion technique is introduced to obtain a low-frequency solution of the time-dependent continuity equation. This method is not based on the quasi-static assumption or on the integration[7] of the continuity equations and therefore does not involve the previous assumptions 1 and 2. As we will see, this method has the advantage that it has a sound mathematical basis, while the results can be presented in such a way that they are easy to compute from a d.c. solution. Moreover, unlike the conventional quasi-static method, it also yields internal a.c. device quantities such as minority carrier densities and current densities.

4.3 A PERTURBATION ANALYSIS OF THE A.C. CONTINUITY EQUATIONS

4.3.1 Derivation of the basic relations

We start with the following generalized Moll-Ross expression for the minority current density [4.12]

$$J(x) = \pm \frac{q\, n_{i0}^2}{G(x)} \frac{dY}{dx}, \qquad (4.16)$$

where $Y = n(x)p(x)/n_{ie}^2(x)$, q is the elementary charge and $n(x)$ and $p(x)$ are the electron and hole densities respectively. The intrinsic carrier densities of undoped and doped silicon are denoted by n_{i0} and $n_{ie}(x)$, respectively. The plus sign in (4.16) holds for electrons and the negative sign for holes. For the hole current density, $G(x)$ is defined by

$$G_p(x) = \frac{n(x)}{D_p(x)\,(n_{ie}(x)/n_{i0})^2}, \qquad (4.17)$$

while for the electron current density, $G(x)$ is

$$G_n(x) = \frac{p(x)}{D_n(x)\,(n_{ie}(x)/n_{i0})^2}. \tag{4.18}$$

$D(x)$ is the diffusion constant of the minority carriers. Eq. (4.16) is generally valid, provided that the gradient of the majority carrier quasi-Fermi level is much smaller than the gradient of the minority carrier quasi-Fermi level.

Now we turn to a small-signal a.c. analysis. At $x = 0$ an a.c. voltage is applied according to (4.8). The quantity $Y(x,t)$ varies with time as

$$Y(x,t) = Y_0(x) + \tilde{y}(x)e^{j\omega t}, \tag{4.19}$$

where $Y_0(x)$ is the d.c. component of $Y(x,t)$ and $\tilde{y}(x)$ is the complex amplitude of the a.c. component. Combining (4.16) with the corresponding continuity equation the following differential equation is obtained for $\tilde{y}(x)$:

$$\frac{d^2\tilde{y}}{dx^2} - \frac{1}{G(x)}\frac{dG}{dx}\frac{d\tilde{y}}{dx} - \frac{\tilde{y}(x)}{D(x)\,\tau_r(x)} = j\omega\,\frac{\tilde{y}(x)}{D(x)}, \tag{4.20}$$

where $\tau_r(x)$ is the recombination lifetime. The above equation holds for both electrons and holes. The boundaries of the interval in which (4.20) is to be solved are at $x = 0$ and $x = W$. For the base region of a bipolar device $x = W$ corresponds to the edge of the collector-base depletion region and for the emitter region $x = W$ denotes the location of the emitter contact. The boundary $x = 0$ denotes the corresponding edge of the emitter-base depletion layer. The boundary conditions for $\tilde{y}(x)$ are

$$\tilde{y}(0) = \frac{\tilde{v}}{V_T}\,Y_0(0) = \frac{\tilde{v}}{V_T}\,e^{V_j/V_T} \tag{4.21}$$

and

$$\left(\frac{1}{\tilde{y}} \frac{d\tilde{y}}{dx} \right)_{x=W} = \frac{-S}{D(W)} \qquad (4.22)$$

where S is the saturated drift velocity when this is applied to the base region [4.16] and S is the surface recombination velocity if $x = W$ corresponds to the emitter contact. If we consider the term on the right-hand side of the above equation as a perturbation term, (4.20) can be solved by means of a perturbation expansion [4.13]. To this end we first apply the following coordinate transformation:

$$\xi = \frac{1}{W} \int_0^x g(x') \, dx', \qquad (4.23)$$

with $g(x) = G(x)/G(0)$. Eq. (4.20) now becomes

$$\frac{d^2\tilde{y}}{d\xi^2} - \frac{\tilde{y}(\xi)W^2}{D(\xi)\,\tau_r(\xi)g^2(\xi)} = \varepsilon \frac{\tilde{y}(\xi)}{\delta(\xi)g^2(\xi)}, \qquad (4.24)$$

with

$$\varepsilon = \frac{j\omega W^2}{D(0)}, \qquad (4.25)$$

and $\delta(\xi) = D(\xi)/D(0)$. The a.c. current density reads

$$\tilde{j} = \pm \frac{qn_{i0}^2}{WG(0)} \frac{d\tilde{y}}{d\xi}. \qquad (4.26)$$

We now write the solution of (4.24) as a power series in ε:

$$\tilde{y}(\xi, \varepsilon) = \sum_{m=0}^{\infty} \tilde{y}_m(\xi) \, \varepsilon^m. \tag{4.27}$$

Substitution of (4.27) into (4.24) gives

$$\sum_{m=0}^{\infty} \left[\left(\frac{d^2 \tilde{y}_m}{d\xi^2} - \frac{\tilde{y}_m(\xi) W^2}{D(\xi) \, \tau_r(\xi) g^2(\xi)} \right) \varepsilon^m - \frac{\tilde{y}_m(\xi)}{\delta(\xi) g^2(\xi)} \, \varepsilon^{m+1} \right] = 0. \tag{4.28}$$

Since the coefficient of each power of ε must equal zero, the following infinite set of differential equations is obtained:

$$\frac{d^2 \tilde{y}_0}{d\xi^2} - \frac{\tilde{y}_0(\xi) W^2}{D(\xi) \, \tau_r(\xi) g^2(\xi)} = 0, \tag{4.29}$$

$$\frac{d^2 \tilde{y}_m}{d\xi^2} - \frac{\tilde{y}_m(\xi) W^2}{D(\xi) \, \tau_r(\xi) g^2(\xi)} = \frac{\tilde{y}_{m-1}(\xi)}{\delta(\xi) g^2(\xi)}, \qquad m > 0. \tag{4.30}$$

Note that (4.29) is equal to the equation for the d.c. solution $Y_0(x)$. To solve (4.30) for $\tilde{y}_m(\xi)$ we must have two independent solutions of the corresponding homogeneous equation. One solution is, of course, $\tilde{y}_0(\xi)$ and the second solution $\tilde{y}_s(\xi)$ can be expressed in terms of the other solution $\tilde{y}_0(\xi)$ as follows [4.14]:

$$\tilde{y}_s(\xi) = \tilde{y}_0(\xi) \int_0^\xi \frac{d\xi'}{\tilde{y}_0^2(\xi')}. \tag{4.31}$$

If we consider the fact that the boundary condition (4A.3) at $x = W$ also holds for the d.c. solution $Y(\xi)$, then it is obvious that $\tilde{y}_0(\xi)$ can be written as

$$\tilde{y}_0(\xi) = \frac{\tilde{v}}{V_T} Y_0(\xi), \qquad (4.32)$$

where V_T is the thermal voltage kT/q.

For low frequencies we can truncate the series (4.27) after the second term, i.e. a first-order expansion in $j\omega$. Using the boundary conditions (4.21) and (4.22) it is shown in appendix 4A that the general first-order solution is given by

$$\tilde{y}(\xi) = \tilde{y}_0(\xi)\left\{1 - \frac{j\omega W^2}{D(0)}\left[\left(\int_\xi^{\xi_W}\tilde{y}_0^2(\xi')g^{-2}(\xi')\delta^{-1}(\xi')d\xi'\right)\int_0^\xi\frac{d\xi'}{\tilde{y}_0^2(\xi')}\right.\right.$$

$$+ \left.\left.\int_0^\xi\tilde{y}_s(\xi')\,\tilde{y}_0(\xi')g^{-2}(\xi')\delta^{-1}(\xi')d\xi'\right]\right\}. \qquad (4.33)$$

The general first-order solution (4.33) can now be expressed in terms of the d.c. solution $Y_0(\xi)$. Application of this result implies that we must know the d.c. minority carrier density which, in general, can only be determined by a numerical device simulation. However, for two cases it is possible to obtain analytical results. The first case is for a homogeneously-doped region. This case is of limited practical value but can be used to study the approximations involved in the conventional quasi-static method as given in section 4.2. The second case for which analytical results can be obtained is when bulk recombination is neglected. Neglecting bulk recombination in both the emitter and base regions is a reasonably good approximation for high-speed devices with shallow junctions. This is especially so at frequencies higher than $(2\pi\tau_r)^{-1}$ because in that case the bulk recombination mechanism makes no contribution to the a.c. current, i.e. the region becomes transparent.

4.3.2 Application to a homogeneously-doped region

When we apply the above procedure to a homogeneously-doped region, the coordinate ξ becomes x/W. The quantity $\tilde{y}(x)$ now becomes

$$\tilde{y}(x) = \frac{\tilde{v}}{V_T} Y_0(x) \left[1 + \frac{j\omega L^2}{2D} \left(x \frac{Y'_0(x)}{Y_0(x)} + \frac{T}{Y_0(x)} \sinh\left(\frac{x}{L}\right) \right) \right], \quad (4.34)$$

with

$$T = \left[\frac{\dfrac{\sigma^2 W}{L} - \dfrac{W}{L} + \sigma}{\cosh\left(\dfrac{W}{L}\right) + \sigma \sinh\left(\dfrac{W}{L}\right)} \right] Y_0(W), \quad (4.35)$$

and

$$Y_0(x) = \frac{\cosh\left(\dfrac{W-x}{L}\right) + \sigma \sinh\left(\dfrac{W-x}{L}\right)}{\cosh\left(\dfrac{W}{L}\right) + \sigma \sinh\left(\dfrac{W}{L}\right)} e^{V_j/V_T}. \quad (4.36)$$

In the above equations L and σ are the diffusion length and normalized recombination velocity or saturated drift velocity ($\sigma = SL/D$) respectively, while V_j is the junction voltage.

We now investigate the two assumptions involved in the quasi-static approach. Differentiation of (4.34) and subsequent substitution of this result into the expression for the amplitude of the a.c. current density (4.26) gives

$$\tilde{j}(0) = \frac{\tilde{v}}{V_T} [J(0) + j\omega\phi\tau_{qs}J(0)] \quad (4.37)$$

and

$$\tilde{j}\,(W) = \frac{\tilde{v}}{V_T}\,[\,J\,(W) - j\omega\alpha\tau_{qs}J\,(0)\,] \qquad (4.38)$$

where the transit time τ_{qs} is given by

$$\tau_{qs} \equiv \frac{Q}{J\,(0)} = \frac{L^2}{D}\left(1 - \frac{Y'_0\,(W)}{Y'_0\,(0)}\right). \qquad (4.39)$$

The coefficients α and ϕ are given by

$$\alpha = \frac{\dfrac{WY_0\,(W)}{L^2 Y'_0\,(0)} + \dfrac{Y'_0\,(W)}{Y'_0\,(0)} + \dfrac{T}{LY'_0\,(0)}\cosh\left(\dfrac{W}{L}\right)}{2\left(\dfrac{Y'_0\,(W)}{Y'_0\,(0)} - 1\right)} \qquad (4.40)$$

and

$$\phi = \frac{1 + \dfrac{T}{LY'_0\,(0)}}{2\left(1 - \dfrac{Y'_0\,(W)}{Y'_0\,(0)}\right)}. \qquad (4.41)$$

According to the quasi-static assumption (assumption 1 at the end of section 4.2) it can be obtained from (4.10) that the difference between $\tilde{j}\,(0)$ and $\tilde{j}\,(W)$ is given by

$$\tilde{j}\,(0) - \tilde{j}\,(W) = \frac{\tilde{v}}{V_T}\,J\,(0)\left[1 - \frac{J\,(W)}{J\,(0)} + j\omega\tau_{qs}\right]. \qquad (4.42)$$

According to the quasi-static result, the term $j\omega J\,(0)\tau_{qs}\tilde{v}/V_T$ is the total imaginary current associated with the stored a.c. charge in the quasi-neutral region. Therefore ϕ and α are the fraction of this imaginary current actually flowing through the forward-biased junction and the contact (or reverse-biased junction), respectively. As we have seen

above, according to the quasi-static assumption (assumption 1 at the end of section 4.2) $\phi + \alpha = 1$.

The second essential assumption involved is that the whole charge is supplied through the forward-biased junction, so $\alpha = 0$. In fig. 4.2 the values of ϕ and $\phi + \alpha$ are plotted as a function of the normalized width W/L for two values of the surface recombination velocity. From this figure it can be observed that the quasi-static assumption $\phi + \alpha = 1$ is valid only in the region $W/L << 1$, i.e. with negligible bulk recombination. The reason for this is as follows: As can be seen from (4.34) the minority carrier concentration is out of phase with the applied voltage. This phase shift is proportional to ω and is negative due to the propagation delay of the carriers. Because of this phase shift the a.c. bulk recombination, which in neutral regions is proportional to the minority
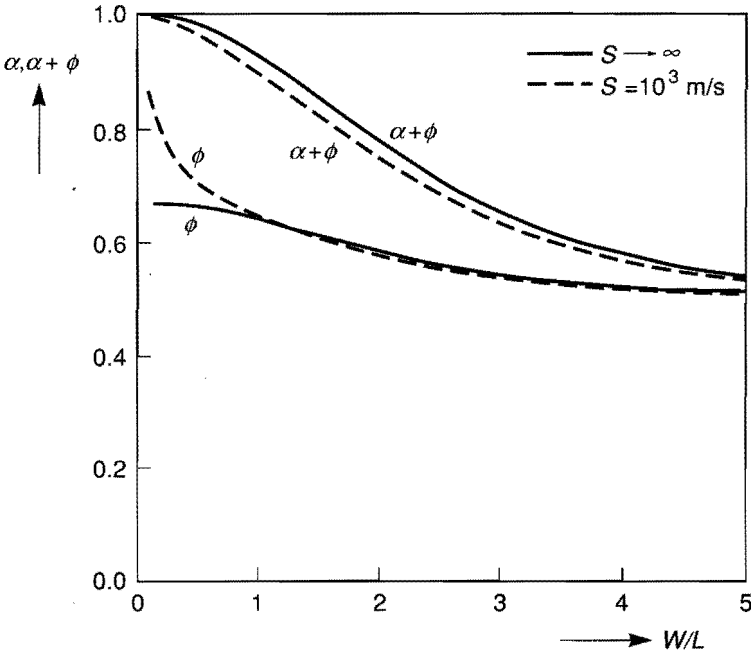
Fig. 4.2. The parameters ϕ and $\phi + \alpha$ plotted as a function of W/L for two values of the recombination velocity. $D = 2\ 10^{-4} m^2/s$, which is a typical value for an emitter region.

carrier concentration, also has an imaginary component. This means that not only the stored charge, but also bulk recombination, contributes to the imaginary terminal currents. Since these terms are opposite in sign they partly compensate each other so, $\phi + \alpha < 1$. We can illustrate this effect with an often discussed example. Consider a homogeneously-doped n-side of a long diode (i.e. $W/L_p \to \infty$). In this case (4.34) reduces to

$$\tilde{p}(x) = \frac{\tilde{v}}{V_T} p\,(x)\left(1 - j\omega\tau_p \frac{x}{2L_p} \right), \qquad (4.43)$$

where $\tilde{p}(x)$ and $p\,(x)$ are the a.c. and d.c. hole concentrations respectively. From the above equation it can be seen that the a.c. hole concentration does indeed have a negative phase shift. This phase shift varies linearly with x. When we substitute (4.43) into the following a.c. continuity equation for holes

$$\frac{d\tilde{j}_p\,(x)}{dx} = -q\tilde{p}(x)\left(\frac{1}{\tau_p} + j\omega \right), \qquad (4.44)$$

we obtain (to first order in $j\omega$)

$$\frac{d\tilde{j}_p\,(x)}{dx} = -q\,\frac{\tilde{v}}{V_T} p\,(x)\left\{ (\,\frac{1}{\tau_p} + j\omega\left(1 - \frac{x}{2L_p} \right) \right\}. \qquad (4.45)$$

In an interval Δx around a certain location x_1, $\Delta\tilde{j}_p$ equals

$$\Delta\tilde{j}_p\,(x_1) = \frac{\tilde{v}}{V_T} \Delta Q\,(x_1)\left\{ (\,\frac{1}{\tau_p} + j\omega\left(1 - \frac{x_1}{2L_p} \right) \right\}, \qquad (4.46)$$

with $\Delta Q\,(x_1) = qp\,(x_1)\Delta x$. This shows that the imaginary part of $\Delta\tilde{j}_p\,(x_1)$ not only consists of the charge storage term $j\omega\Delta Q(x_1)\tilde{v}/V_T$, but also of a term which is the current due to recombination of the out-of-phase hole concentration. As can be seen from fig. 4.2, in this case $(W/L >> 1)$ $\phi \to 0.5$ and $\alpha \to 0$. This reflects the well-known result

that in a very wide diode the minority-carrier current flows only through one contact and the imaginary component is only one-half of the quasi-static result [4.15].

With respect to the validity of the two assumptions mentioned at the end of section 4.2, the case of a fully transparent region ($W/L << 1$) is opposite to the case of a fully non-transparent region ($W/L >> 1$). In the former case the quasi-static assumption (i.e. $\alpha + \phi = 1$) is valid, but the assumption that the stored charge is supplied entirely through the forward-biased junction (i.e. $\alpha = 0$) does not apply. In the latter case the stored charge is indeed supplied through the forward-biased junction, but in that case the quasi-static assumption is not valid.

4.3.3 Application to a recombination-free region

When bulk recombination is negligible the perturbation analysis as outlined in section 4.3.1 can also be analytically applied to an arbitrarily-doped region. In appendix 4B it is derived that the a.c. currents can be written as

$$\tilde{j}(0) = \frac{\tilde{v}}{V_T} J(0)[1 + j\omega(1 - \alpha)\tau_{qs}], \qquad (4.47)$$

and

$$\tilde{j}(W) = \frac{\tilde{v}}{V_T} J(0)[1 - j\omega\alpha\tau_{qs}], \qquad (4.48)$$

where α is given by

$$\alpha = \frac{k}{G_t} \frac{\displaystyle\int_0^W c(x)\left[\int_0^x G(x')dx'\right]dx}{\displaystyle\int_0^W c(x)dx}. \qquad (4.49)$$

In (4.49) $c(x)$ is the minority carrier concentration (see eq. (4B.9)) and G_t is the Gummel number, i.e. the integral of $G(x)$ over the interval $[0, W]$. See (4B.3) for a definition of k.

If we apply the above analysis to the base region of a bipolar device, then (4.48) gives the collector current and α in (4.49) gives the fraction of the stored minority-carrier charge in the quasi-neutral base which is supplied from the collector. When the above analysis is applied to the emitter, (4.48) gives the hole current density at the emitter contact. In both cases the parameter α has to be obtained from (4.49).

In order to obtain a simple expression for α in the base region, defined as α_b, we adopt the widely used exponential approximation for the base doping profile $N(x) = N_0 \exp(-\eta x/W)$. Assuming that D_n and n_{ie} are $x-$ independent, the analytical evaluation of (4.49) is straightforward and results in:

$$\alpha_b = \frac{\eta(1 + D/WS) - 2}{\eta(1 + D/WS) - 1} \qquad (4.50)$$

for $\eta >> 1$ and $WS/D >> 1$. Note that for the base region S denotes the saturated drift velocity of the carriers. For $\eta = 0$, α_b becomes

$$\alpha_b = \frac{1}{3}\left[1 - \frac{2}{(WS/D + 1)(WS/D + 2)}\right]. \qquad (4.51)$$

In fig. 4.3 the value of α_b is plotted for two cases: $S = 10^5 m/s$ and $S \rightarrow \infty$. We see from fig. 4.3 that for the latter case with $\eta = 0, \alpha_b = 1/3$, as it should do [4.15]. For a homogeneously-doped base (i.e. $\eta = 0$) the major part of the base charge is supplied from the emitter while in the case of modern transistors, for which $4 < \eta < 8$, the major part of the base charge is supplied through the collector. The reason for this is that the minority-carrier flow is enhanced by the built-in field towards the collector-base junction, together with the fact that the minority-carrier distribution is shifted towards the collector-base junction. At this point it is interesting to compare the obtained results with the results in [4.9,4.10,4.11,4.15]. It can be shown that, in the case of negligible bulk recombination, Klose and Wieder [4.10]
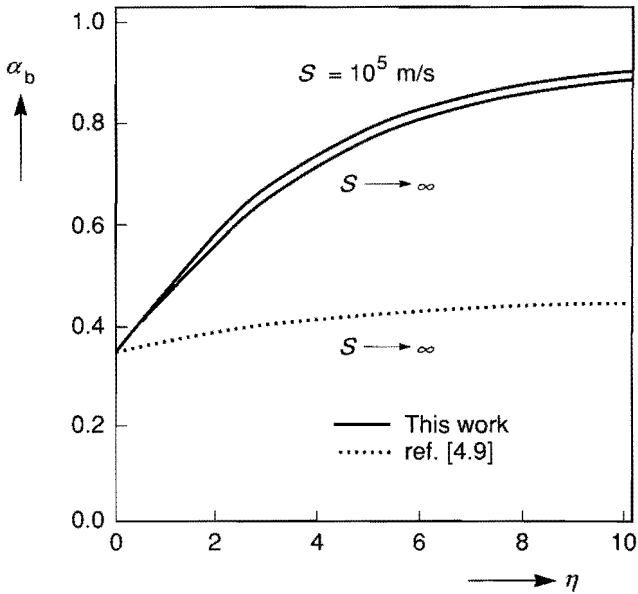
Fig. 4.3. The parameter α_b as a function of η for $S \to \infty$ and for $S = 10^5 m/s$. The dashed line is obtained by adopting the result of [4.9], and using $S \to \infty$.

arrive at the same general expression (4.49) for α_b. However, Fossum and Veeraraghavan [4.9] have used the weighting function for a homogeneously-doped base in their expression for the currents, instead of the more general function $G(x)/G_i$ (see eq. (4.49)). In the case of an exponentially doped base this leads to a great underestimation of α_b, as can be seen in fig. 4.3. The results in both [4.11] and [4.15] are based on an exact solution of the continuity equation in an exponentially doped region with $S \to \infty$. The result for α_b is then obtained by a series expansion of this solution. For $\eta >> 1$ this yields expression (4.50) with $S \to \infty$, as it should be.

An important parameter which is not accounted for in the conventional quasi-static approach is the phase of the transconductance

Modelling of small-signal a.c. transistor properties   *153*
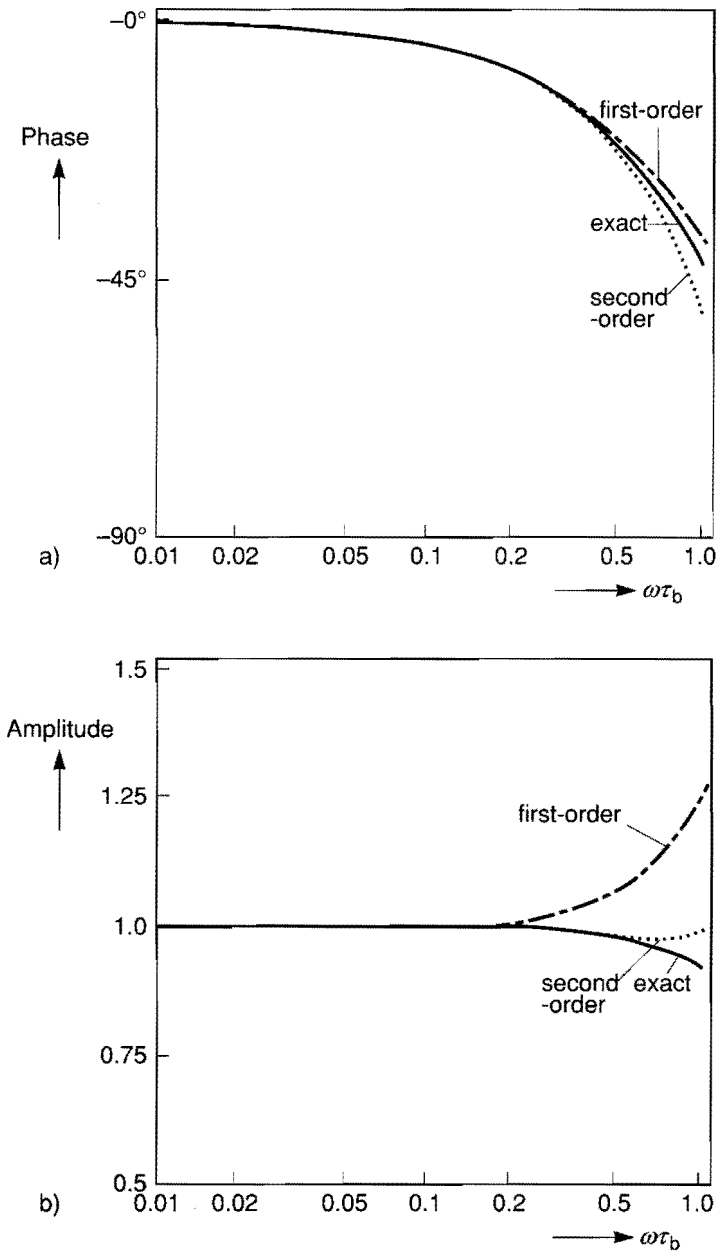
Fig. 4.4. The phase (a) and normalized amplitude (b) of the transconductance for an exponentially-doped region with $\eta = 5$ and $S \to \infty$. The exact results, as given in [4.15], are also denoted.

$\tilde{j}(W)/\tilde{v}$, i.e. the excess phase shift. The above analysis yields $-\arctan(\omega\alpha_b\tau_b)$ for this phase. In fig. 4.4a this phase is plotted as a function of $\omega\tau_b$ for $\eta = 5$ and $S \to \infty$. The exact phase, as given in [4.15], is also denoted. We see that this phase is accurately described by our first-order expansion for frequencies up to $(2\pi\tau_b)^{-1}$. Note that the frequency $(2\pi\tau_b)^{-1}$ is usually considerably higher than the cut-off frequency f_T because τ_b is only a part of the total transit time. However, when we consider the amplitude of the transconductance (fig. 4.4b) we see that the first-order solution increases considerably near $\omega = \tau_b^{-1}$, while the exact solution decreases. By including a second-order term, i.e. $m = 2$ in (4.30), this difference can be significantly reduced (see fig. 4.4b). Application of the procedure as outlined in section 4.3 yields

$$\tilde{j}\ (W) = \frac{\tilde{v}}{V_T}\ J_0[1 - j\alpha_b(\omega\tau_b) - \gamma_b(\omega\tau_b)^2], \qquad (4.52)$$

with

$$\gamma_b = \frac{\eta^2 - 2\eta - 4}{2\eta^2 - 4\eta + 2}, \qquad (4.53)$$

for $\eta >> 1$ and $WS/D >> 1$. In [4.5] the higher-order effects on the collector current are taken into account by a differential operator. After transforming this operator to the frequency domain and expanding this expression around $\omega = 0$, we arrive for $\eta >> 1$ at (4.53).

For the emitter region we assume a Gaussian profile for $0 < x < d$ and a constant doping level N_0 for $d \le x < W$. See fig. 4.5 for a plot of the profile. The characteristic length of the Gaussian profile can be determined from the ratio $N(0)/N_0$. Typically, for modern emitters this ratio is around 0.01. In fig. 4.6 the dashed line gives α_e as a function of d/W for $S \to \infty$. At a high doping density ($N_0 = 10^{26}m^{-3}$) the assumption of constant bandgap narrowing over a two-decade fall-off in the doping density is very inaccurate. In the doping range of $10^{24} - 10^{26}m^{-3}$ the bandgap narrowing can be expressed as a power function of the dope [4.17]. This implies that when $N(x)$ is a Gaussian

function, $G(x)$ is also Gaussian. The characteristic length of $G(x)$ can also be obtained from $G(0)/G_0$. In the doping range given above, $G(0)/G_0 = 0.1$. Taking this into account, we obtain the solid line in fig. 4.6. From this figure we see that a proper inclusion of the dopant dependence of bandgap narrowing in the calculation of α_e is important. Obviously, for $S \to \infty$ and $d = 0$, $\alpha_e = 1/3$. We see that for a non-uniformly doped emitter the value of α_e is lower than in the case of a uniformly-doped emitter. The reason for this effect is that the polarity of the built-in field is such that the minority carriers are driven towards the base-emitter junction. This is the opposite of the situation in the base. In the case of a non-uniformly doped base α_b is higher than in a homogeneously-doped base, as shown in fig. 4.3, because the minority-carrier flow is enhanced by the built-in field towards the collector-
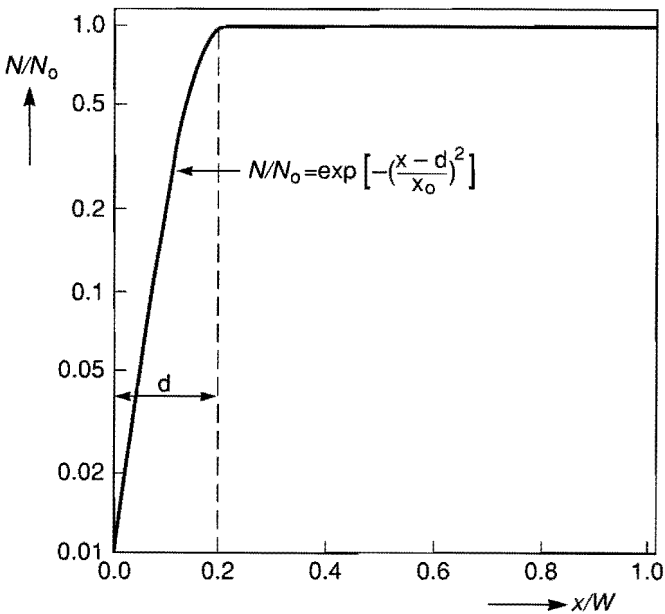

Fig. 4.5. Plot of the Gaussian emitter profile. The origin corresponds to the edge of the emitter- base depletion layer in the emitter.
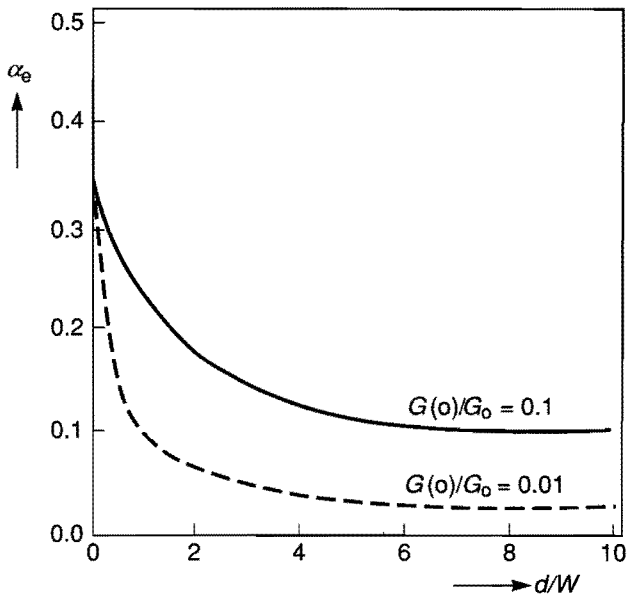
Fig. 4.6. The parameter α_e as a function of d/W for $S \to \infty$ and $N(0)/N_0 = 0.01$ with (solid line) and without (dashed line) inclusion of the dope dependence of the bandgap narrowing.

base junction. As we can see in fig. 4.6, the value of α_e decreases strongly when d/W is increased from zero. The reason for this is shown in fig. 4.7 where the minority carrier density is given for $d/W = 0.1$ (solid lines) and for the homogeneous case $d = 0$ (dashed line). We see that the extra minority charge is almost entirely located in the region $0 < x < d$ near the emitter-base junction. This extra minority charge near the junction, together with the aiding built-in field in that region, causes a significantly larger part of the emitter charge to be reclaimable through the emitter-base junction when d/W is increased from zero. This effect is more pronounced in the case of constant bandgap narrowing because, as can be observed from fig. 4.7, the peak of the mi-
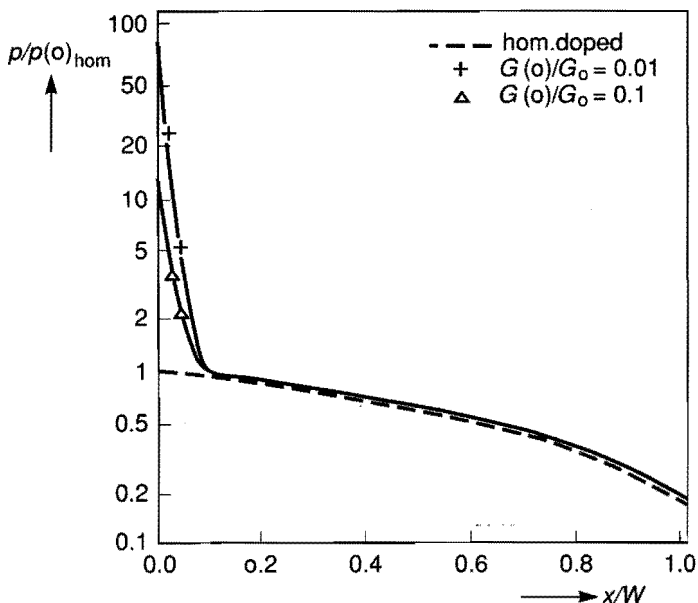
Modelling of small-signal a.c. transistor properties   *157*

Fig. 4.7. The minority carrier density in the emitter as a function of x/W. The values are normalized to the density at $x = 0$ for the case of a homogeneously-doped emitter. The dashed line gives the result for a homogeneously-doped emitter, while the solid lines give the results for $d/W = 0.1$ and $N(0)/N_0 = 0.01$. Furthermore, $S = 10^4 m/s$ and $W = 0.1\mu m$.

nority carrier density is then a factor 10 higher than in the case of dope-dependent bandgap narrowing. To indicate the influence of the emitter surface recombination, in fig. 4.8 the values of α_e (including dope-dependent bandgap narrowing) are shown as a function of d/W for three values of the surface recombination velocity S.

4.4 DISCUSSION AND CONCLUSIONS

In section 4.2 we have seen that the quasi-static approach, on which (4.1) and (4.2) are based, involves two assumptions:

1) The minority carrier density is in phase with the applied junction voltage (eq. (4.6)), i.e. the quasi-static assumption.

2) The stored minority charge in the emitter and base regions is entirely supplied through the emitter-base junction.

As a result of the first assumption the sum of the imaginary terminal currents for either holes or electrons is determined only by the stored charge in the device. In section 4.3. we have seen that this is true only when bulk recombination is negligible. If this is not the case, the imaginary terminal currents also consist of a term due to bulk recombination. The second assumption is necessary because, due to the integration of the continuity equation over the volume of the device, this approach only yields the sum of the terminal currents. In order to obtain expressions for $\tilde{i}_b$ and $\tilde{i}_c$, an additional assumption must be made
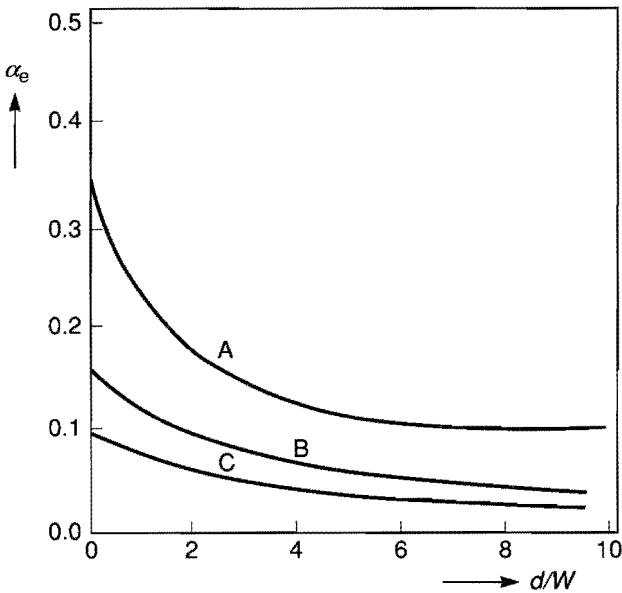
Fig. 4.8. The parameter α_e as a function of d/W for $S \to \infty$ (A), $S = 10^3 m/s$ (B) and $S = 5 \; 10^2 m/s$ (C). In these cases $G(0)/G_0 = 0.1$.

about the distribution of the total hole and electron currents over the contacts. In fact, assumption 2 is arbitrary and, as we have seen in section 4.3, is generally not valid.

In section 4.3 we have presented a completely different approach to the determination of the small-signal a.c. currents from the results of a d.c. calculation. This method is not based on the quasi-static assumption or on the integration of the continuity equations and therefore does not involve the previous assumptions. Unlike the quasi-static approach this method also yields internal a.c. quantities such as the complex carrier densities and current densities. As a result of the fact that we have taken account of only the first-order term in $j\omega$, the results presented are low-frequency approximations. A drawback of the method, as presented here, is that only calculations for a one-dimensional problem can be performed.

In view of the above considerations, in the rest of this section we discuss the use of the quasi-static cut-off frequency $f_{T_{qs}}$ (from (4.2) as a figure of merit for the characterization of high-frequency properties of advanced transistors. We confine ourselves to the common-emitter configuration, which means that the a.c. voltage difference between emitter and collector is zero (see fig. 4.9) . In this case the imaginary base current consists of terms associated with hole storage and recombination in the emitter, base and collector regions. The imaginary emitter current consists of terms associated with electron storage and recombination in the emitter and base regions, while the imaginary collector current is due to electron storage and recombination in the base and collector regions.

From both small-signal a.c. measurements and small-signal a.c. simulations (non-quasi-static) it is found that the current gain as a function of frequency does indeed accurately follow a first order relation of the form

$$h_{FE} = \frac{h_{FE0}}{1 + j\omega h_{FE0}\,\tau_{AC}}, \qquad (4.54)$$

provided the d.c. current gain h_{FE0} is much larger than unity [4.18,4.19]. The quantities h_{FE0} and the characteristic time constant
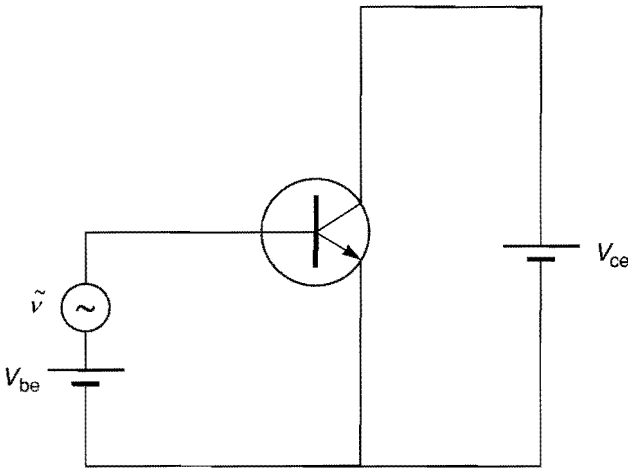
Fig. 4.9. Common-emitter configuration.

τ_{AC} can be extracted from the measured or simulated curve. An example is given in fig. 4.10, which shows the current gain as a function of the frequency at $V_{be} = 0.825V$ and $V_{ce} = 3.825V$ as obtained from one-dimensional numerical a.c. simulations. The corresponding doping profile is given in fig. 4.11b. Both the numerical a.c. results and the first-order approximation (4.54) are plotted. The measured (or simulated) quantity h_{FE0} gives information about the low-frequency behaviour of the device, while the measured (or simulated) value of τ_{AC} provides information about the high-frequency behaviour. The measured (or simulated) cut-off frequency f_T is related to τ_{AC} by $f_T = (2\pi\tau_{AC})^{-1}$. This cut-off frequency is used as the figure of merit for the characterization of the high-frequency behaviour of the device. The interesting question now is how accurately the quasi-static cut-off frequency f_{Tqs} approximates to the actual cut-off frequency f_T. To this end we compared the value of f_T as extracted from numerical small-signal
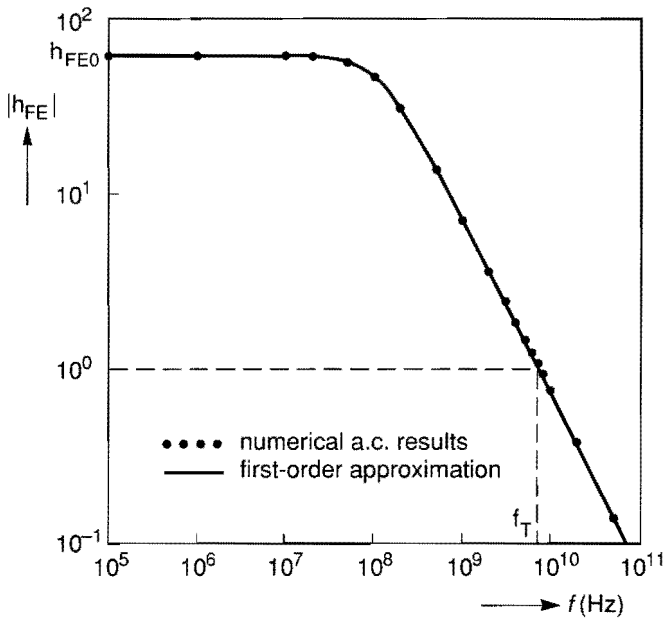
Fig. 4.10. Modulus of the current gain as a function of the frequency. $V_{be} = 0.825V$ and $V_{ce} = 3.825V$. Both the numerical a.c. results and the first-order approximation (4.54) are denoted.

a.c. calculations with the quasi-static cut-off frequency for two profiles and three emitter widths. In fig. 4.11, f_T and f_{Tqs}, both obtained from 2D numerical simulations, are plotted for three emitter widths, viz. $H_e = 0.5\mu m$, $2\mu m$ and ∞. For the case of an infinitely wide emitter the collector current density is multiplied by $2\mu m$. In fig. 4.12 the same curves are given for a shallow device. From figs. 4.11 and 4.12 we observe that also for laterally downscaled devices the quasi-static expression (4.2) gives a fairly good indication of f_T and can therefore be used as a figure of merit for the characterization of the a.c. properties of downscaled devices. From both figures it is observed that, especially at high bias conditions, the quasi-static quantity f_{Tqs} is slightly less than

Chapter 4

Fig. 4.11. f_{Tqs} and f_T as a function of the collector current for three emitter widths, viz. $H_e = 0.5\mu m$, $2\mu m$ and ∞ (a). The vertical doping profile, i.e. along cross-section A-A in fig. 1.3, is shown in (b). The solid lines denote f_{Tqs} and the symbols denote f_T.
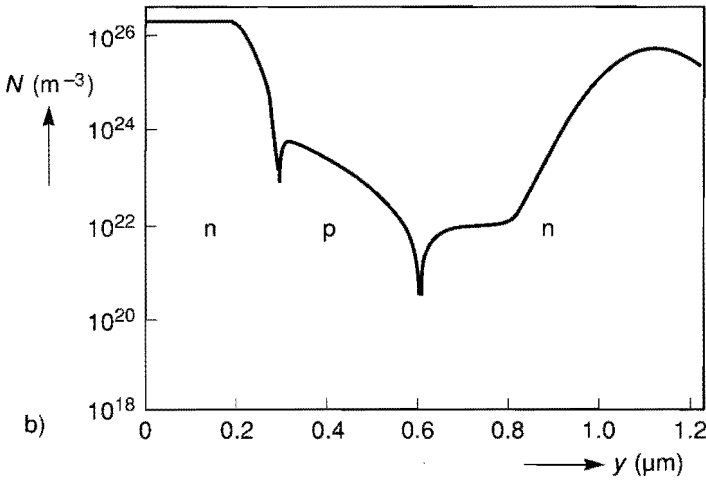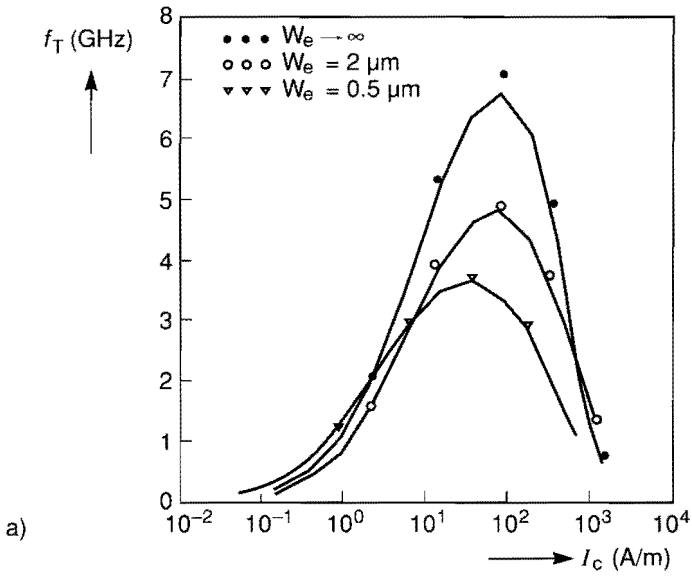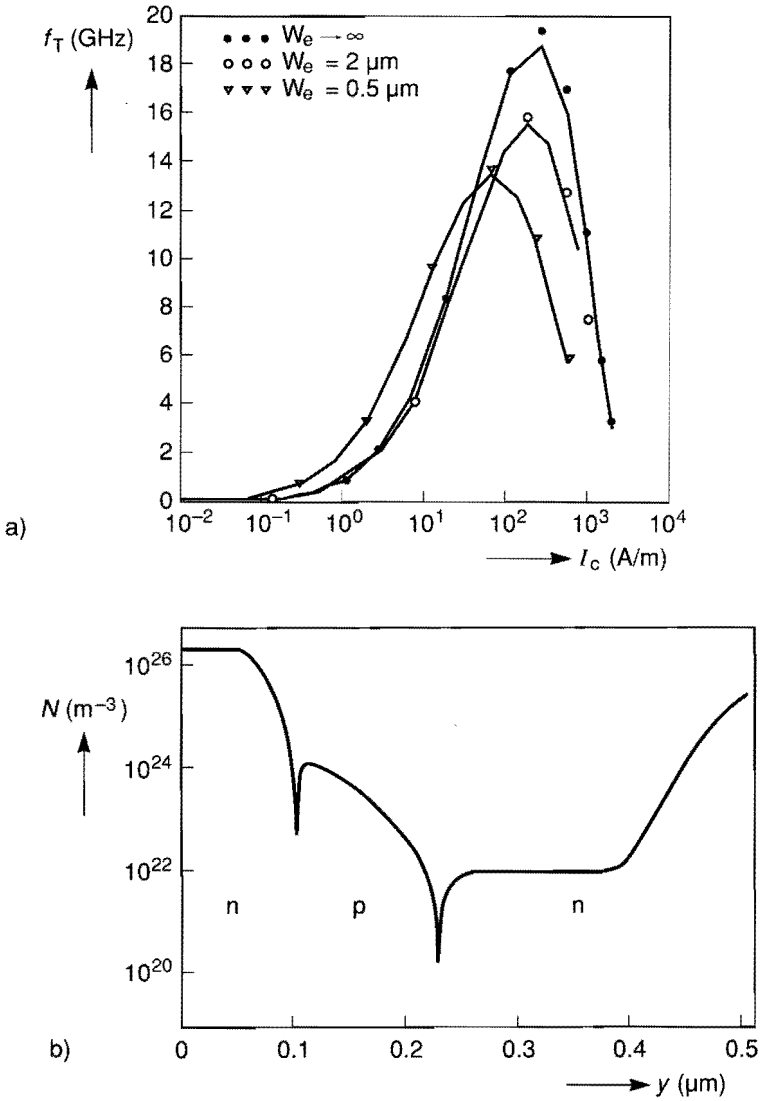
Modelling of small-signal a.c. transistor properties 163

Fig. 4.12. $f_{T_{qs}}$ and f_T as a function of the collector current for three emitter widths, viz. $H_e = 0.5 \mu m$, $2 \mu m$ and ∞. The vertical doping profile is given in (b). The solid lines denote $f_{T_{qs}}$ and the symbols denote f_T.

f_T, which would be the cut-off frequency as obtained from a.c. measurements or a.c. simulations.

In view of the analysis in sections 4.2 and 4.3 we will now investigate the discrepancy between f_T and $f_{T_{qs}}$. First we investigate the base current. If the emitter transit time plays a significant role in the total transit time, the quasi-static expression (4.14) for the base current may not be valid for two reasons:

1) Unlike the situation in the base region, in the emitter region the hole recombination length is not much larger than the emitter depth, so the emitter cannot considered to be fully transparent. This implies that for the emitter region the quasi-static assumption (assumption 1) is not fully justified.

2) According to the results in section 4.3.3 the imaginary hole current at the emitter contact is not equal to zero. This means that, in principle, assumption 2 is not valid either.

By analogy with (4.37) we write for the a.c. hole current injected into the emitter

$$\tilde{i}_{be} = \tilde{v} \left(\frac{dI_{be}}{dV} \right)_{V_0} [1 + j\omega\phi_e\bar{\tau}_e], \tag{4.55}$$

where I_{be} is the d.c. hole current injected into the emitter and $\bar{\tau}_e = (dQ_e/dI_{be})_{V_0}$. According to the analysis in section 4.3.3, $\phi_e = 0.5$ for a homogeneously-doped wide emitter ($W_e > L_p$) and $\phi_e = 1 - \alpha_e$, where α_e is given by (4.49), for a transparent emitter. According to the conventional quasi-static approach as described in section 4.2, $\phi_e = 1$. In order to estimate the magnitude of this effect on the cut-off frequency we assume $I_{be} = I_b$, i.e. the d.c. base current consists of hole injection into the emitter only. Now the conventional quasi-static result for the base current (4.14) is replaced by:

$$\tilde{i}_b = \tilde{v} \left(\frac{dI_b}{dV} \right)_{V_0} [1 + j\omega h_{FE0}\{\tau_{qs} - (1 - \phi_e)\tau_e\}], \tag{4.56}$$

where τ_e, by analogy with the definition (4.2) of τ_{qs}, is given by $\tau_e = (dQ_e/dI_c)_{V_0}$. So, $\tau_e = \bar{\tau}_e / h_{FE0}$.

The quasi-static expression for the collector current (4.15) is replaced by

$$\tilde{i}_c = \tilde{v} \left(\frac{dI_c}{dV} \right)_{V_0} [1 - j\omega(\alpha_b\tau_b + \tau_{cb})], \qquad (4.57)$$

with $\tau_b = (dQ_b/dI_c)_{V_0}$ and α_b being given by (4.49). The additional time constant τ_{cb} is associated with the transit time through the collector-base depletion layer. By combining (4.56) and (4.57) we obtain the following modified expression for the current gain

$$h_{FE} = \frac{h_{FE0}[1 - j\omega(\tau_{cb} + \alpha_b\tau_b)]}{1 + j\omega h_{FE0}\tau_{qs}[1 - (1 - \phi_e)\tau_e/\tau_{qs}]}. \qquad (4.58)$$

For the case of a homogeneously-doped transparent emitter, for which $1 - \phi_e = 1/3$, such an expression has already been given in [4.18].

When we compare (4.58) with (4.1) we observe two differences:

1) In (4.58) the imaginary component of the collector current causes a deviation of the first-order behaviour for $\omega \geq (\tau_{cb} + \alpha_b\tau_b)^{-1}$. However, since $\tau_{cb} + \alpha_b\tau_b$ is smaller than τ_{qs}, this deviation will be significant only at frequencies around f_T or higher.

2) In the denominator the time constant τ_{qs} in (4.1) is replaced by $\tau_{qs} - (1 - \phi_e)\tau_e$. This, of course, yields no deviation from the first-order behaviour but it implies that the cut-off frequency resulting from (4.58) is larger than the quasi-static cut-off frequency f_{Tqs}. This is in agreement with the observations from figs. 4.11 and 4.12 that the actual cut-off frequency is somewhat larger than f_{Tqs}.

In fig. 4.13 the current gain is plotted as a function of $\omega\tau$ using the conventional expression (4.1) and the modified expression (4.58). It is assumed that $\tau_e = \tau_b = \tau/3$ and $\tau_{cb} = \tau_{qs}/6$, while the depletion capacitances can be neglected. These are reasonable values at a bias condition around which the maximum cut-off frequency is obtained for modern devices. The value of ϕ_e is taken to be 2/3, while $\alpha_b = 1/3$, i.e. the val-
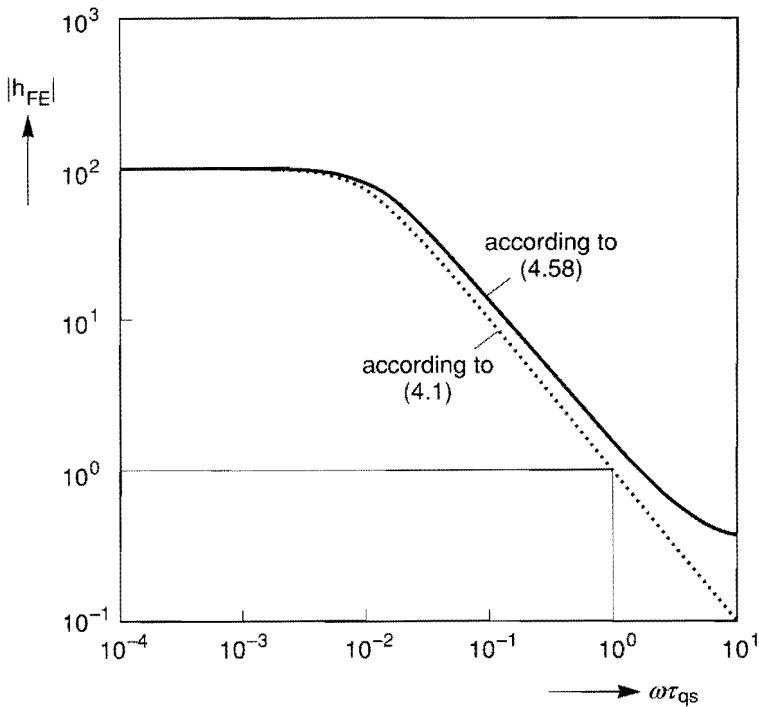
Fig. 4.13. Modulus of the current gain as a function of $\omega\tau_{qs}$ according to (4.1) and (4.58). The d.c. current gain is 100, while $\tau_b = \tau_e = \tau_{qs}/3$ and $\tau_{cb} = \tau_{qs}/6$.

ues for the case of homogeneously-doped transparent regions. The dashed line is obtained by using (4.1) and the solid line is obtained from (4.58). We see that the modified expression (4.58) gives a 10 % larger value for the cut-off frequency than the conventional expression (4.1). However, except for epitaxially-grown emitters, the assumption of a homogeneously-doped emitter is not a very practical one. As we have seen in section 4.3.3, in the case of diffused or implanted emitters the major part of the emitter minority charge is located near the emitter-base junction. This, together with the fact that the polarity of the built-in field in the emitter is such that the minority carriers are driven towards the emitter-base junction, causes ϕ_e to be even larger than 2/3. As an indication, from fig. 4.8 a value of around 0.85 is found for ϕ_e

Modelling of small-signal a.c. transistor properties   *167*

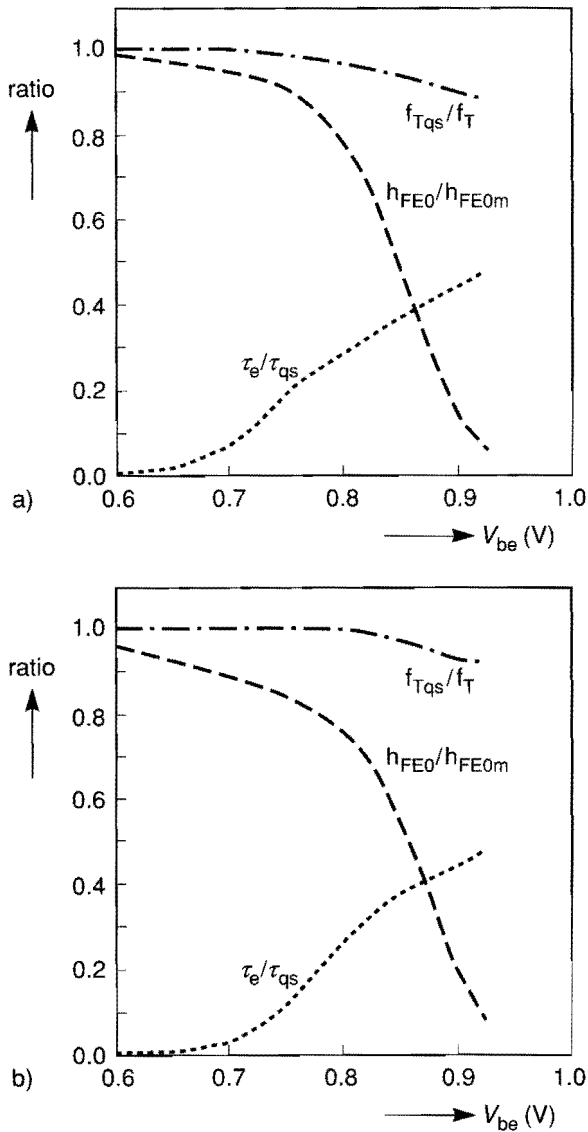Fig. 4.14. The ratios f_{Tqs}/f_T, τ_e/τ_{qs} and h_{FE0}/h_{FE0m} plotted as a function of base-emitter voltage. The collector-base voltage is $3V$. h_{FE0m} is the maximum value of the low-frequency current gain as a function of V_{be}. Case a) is for the doping profile given in fig. 4.11 and b) for the doping profile in fig. 4.12.

(in this case $\phi_e = 1 - \alpha_e$). Under the assumptions $\tau_e = \tau_b = \tau_{qs}/3$, this would give an increase of around 5% for the cut-off frequency. This increase is in the same range as found for the numerical calculations shown in figs. 4.11 and 4.12. This can be more clearly seen in fig. 4.14 where the ratio f_{Tqs}/f_T is plotted as a function of the base-emitter voltage. In this figure we also see that the difference between f_T and f_{Tqs} increases with increasing base-emitter voltage. Using the modified expression (4.58) for the current gain, this can be understood as follows: As explained above, the emitter transit time τ_e is given by $\tau_e = \bar{\tau}_e / h_{FE0}$, where $\bar{\tau}_e = dQ_e/dI_b$. Since the emitter is highly doped, high-injection effects play no significant role in that region and $\bar{\tau}_e$ is virtually constant up to very high bias conditions. However, h_{FE0} decreases at a high base-emitter voltage due to high injection in the base or quasi-saturation. This is shown in fig. 4.14. As a result τ_e increases and the relative contribution of the emitter transit time to the total transit time also increases (see fig. 4.14). Hence, from (4.58) it follows that the difference between the cut-off frequency and f_{Tqs} increases. Nevertheless, we can conclude that at all practical bias conditions the quasi-static cut-off frequency f_{Tqs} gives a good indication of the actual cut-off frequency and can therefore be used as a figure of merit for the characterization of the a.c. properties of downscaled devices.

REFERENCES

4.1 R. Beaufoy and J.J. Sparkes, A.T.E.J., 13, p. 310 (1957)

4.2 J.W. Slotboom, Thesis, Eindhoven University of Technology, Eindhoven (1977)

4.3 D.J. Roulston, S.G. Chamberlain and J. Sehgal, IEEE Trans. Electron Devices, ED-19, p. 809 (1972)

4.4 J.J. te Winkel, Advances in Electr. and Electr. Physics, 39, p. 253 (1975)

4.5 H.C. de Graaff and F.M. Klaassen, Compact Transistor Modelling for Circuit Design, Wien: Springer, 1990

4.6 D.J. Hamilton, F.A. Lindholm and J.A. Narud, Proc. IEEE, 52, p. 239 (1964)

4.7 H.C. de Graaff and G.A.M. Hurkx, Proc. ESSDERC '88, Montpellier, p. 707 (1988)

4.8 G.A.M. Hurkx, Solid-St. Electron., 31, p. 1269 (1988)

4.9 J.G. Fossum and S. Veeraraghavan, IEEE Electron Device Letters, EDL-7, p. 652 (1986)

4.10 H. Klose and A.W. Wieder, IEEE Trans. Electron Devices, ED-34, p. 1009 (1987)

4.11 Te Winkel, IEEE Trans. Electron Devices, ED-20, p. 389 (1973)

4.12 J.A. Seitchik, L.A. Arledge Jr. and P. Yang, IEEE Trans. Electron Devices, ED-33, p.61 (1986)

4.13 G.F. Carrier and C.E. Pearson, Ordinary Differential Equations, Waltham: Blaisdell Publishing Company, 1968, p. 79

4.14 R.P. Agnew, Differential Equations, New York: Mc Graw-Hill Book Company Inc, 1942, p. 117

4.15 J. Lindmayer and C.Y. Wrigley, Fundamentals of Semiconductor Devices, Princeton: D. Van Nostrand Company, 1965, p. 50

4.16 C.T. Kirk, IRE Trans. Electron Devices, ED-9, p. 164 (1962)

4.17 J.W. Slotboom and H.C. de Graaff, Solid-St. Electron., 19, p. 857 (1976)

4.18 J.J.H. v.d. Biesen, IEEE Trans. Computer-Aided Design, CAD-8, p. 855 (1988)

4.19 M.P.G. Versleijen, Philips Nat. Lab. Report no. 6478 (1990)

4.20 H.K. Gummel, Proc. IEEE, 57, p. 2159 (1969)

APPENDIX 4A

The general solution of (4.30) reads [4.14]

$$\tilde{y}_m(\xi) = A_{m,1}\,\tilde{y}_0(\xi) + A_{m,2}\,\tilde{y}_s(\xi) - \int_0^\xi [\tilde{y}_0(\xi)\,\tilde{y}_s(\xi') - \tilde{y}_0(\xi')\,\tilde{y}_s(\xi)]$$

$$\times \tilde{y}_{m-1}(\xi')\delta^{-1}(\xi')g^{-2}(\xi')d\xi', \qquad (4A.1)$$

which is a recurrence relation between the functions $\tilde{y}_m(\xi)$ and $\tilde{y}_{m-1}(\xi)$. The constants $A_{m,1}$ and $A_{m,2}$ can be determined from the boundary conditions for $\tilde{y}(\xi, \varepsilon)$. These are

$$\tilde{y}(0, \varepsilon) = \frac{\tilde{v}}{V_T}\,Y_0(0) = \frac{\tilde{v}}{V_T}\,e^{V_j/V_T}, \qquad (4A.2)$$

and

$$\left(\frac{1}{\tilde{y}}\frac{d\tilde{y}}{d\xi}\right)_{\xi=\xi_W} = \frac{-WS}{D(\xi_W)\,g(\xi_W)} = \left(\frac{1}{Y_0}\frac{dY_0}{d\xi}\right)_{\xi=\xi_W} \qquad (4A.3)$$

where S is the saturated drift velocity when this is applied to the base [4.16] and S is the surface recombination velocity if $x = W$ corresponds to the emitter contact. The coordinate ξ equals ξ_W for $x = W$. In the case of low frequencies we can truncate the series (4.27) after the second term, i.e. a first-order expansion in $j\omega$. We choose $\tilde{y}_0(0) = \tilde{y}(0)$, so $\tilde{y}_s(0) = 0$ and (4A.2) holds both for $\tilde{y}_0(\xi_W)$ and $\tilde{y}_s(\xi_W)$. From $\tilde{y}_s(0) = 0$ and (4A.1) it follows that $A_{m,1} = 0$. Expression (4A.1) now becomes:

$$\tilde{y}(\xi) = \tilde{y}_0(\xi)\left\{1 + \frac{j\omega W^2}{D(0)}\left[\left(A_{m,2} + \int_0^\xi \tilde{y}_0^2(\xi')g^{-2}(\xi')\delta^{-1}(\xi')d\xi'\right)\int_0^\xi \frac{d\xi'}{\tilde{y}_0^2(\xi')}\right.\right.$$

$$- \int_0^\xi \tilde{y}_s\,(\xi')\,\tilde{y}_0\,(\xi')g^{-2}(\xi')\delta^{-1}(\xi')d\xi' \Bigg]\Bigg\}. \qquad (4A.4)$$

The application of boundary condition (4A.3) gives

$$A_{m,2} = -\,\tilde{y}_s\,(\xi)\int_0^{\xi_W}\tilde{y}_0^2(\xi')g^{-2}(\xi')\delta^{-1}(\xi')d\xi'. \qquad (4A.5)$$

Substitution of the above expression into (4A.4) gives (4.33).

APPENDIX 4B

In the case of negligible bulk recombination (4.29) reduces to the Laplace equation and yields for $Y_0\,(\xi)$:

$$Y_0\,(\xi) = (1 - a\,\xi)\,e^{V_j/V_T}, \qquad (4B.1)$$

with

$$a = \frac{kWG(0)}{G_t}, \qquad (4B.2)$$

and

$$k = \frac{1}{1 + \dfrac{G(W)D\,(W)}{G_t\,S}}. \qquad (4B.3)$$

Using (4.32) and (4.31) we find

$$\tilde{y}_0\,(\xi) = \frac{\tilde{v}}{V_T}\,(1 - a\,\xi)e^{V_j/V_T}. \qquad (4B.4)$$

and

$$\tilde{y}_s(\xi) = \left(\frac{\tilde{v}}{V_T} e^{V_j/V_T} \right)^{-1} \xi. \tag{4B.5}$$

Differentiation of (4.33) yields:

$$\tilde{y}'(\xi) = \tilde{y}_0'(\xi) + \frac{j\omega W^2}{D(0)} \left[-\tilde{y}_s'(\xi) \int_\xi^{\xi_W} \tilde{y}_0^2(\xi') g^{-2}(\xi') \delta^{-1}(\xi') d\xi' \right.$$

$$\left. - \tilde{y}_0'(\xi) \int_0^\xi \tilde{y}_s(\xi') \tilde{y}_0(\xi') g^{-2}(\xi') \delta^{-1}(\xi') d\xi' \right]. \tag{4B.6}$$

From the above expression it follows that

$$\tilde{y}'(\xi_W) = \tilde{y}_0'(\xi_W) \left[1 - \frac{j\omega W^2}{D(0)} \int_0^{\xi_W} \tilde{y}_s(\xi') \tilde{y}_0(\xi') g^{-2}(\xi') \delta^{-1}(\xi') d\xi' \right]. \tag{4B.7}$$

Substitution of (4B.4) and (4B.5) into (4B.7) gives

$$\tilde{y}'(\xi_W) = -a \frac{\tilde{v}}{V_T} e^{V_j/V_T} \left[1 - \frac{j\omega W^2}{D(0)} e^{-V_j/V_T} \int_0^{\xi_W} \xi' Y_0(\xi') g^{-2}(\xi') \delta^{-1}(\xi') d\xi' \right].$$

$$\tag{4B.8}$$

After transformation to the variable x and using

$$c(x) = \frac{Y_0(x) n_{i0}^2}{G(x) D(x)}$$

$$= n_{i0}^2 \, e^{V_j/V_T} \left(\frac{1 - \dfrac{k}{G_t} \displaystyle\int_0^x G(x')dx'}{G(x)D(x)} \right), \qquad (4\mathrm{B}.9)$$

and

$$|J(0)| = \frac{qn_{i0}^2}{G_t} \, k \, e^{V_j/V_T}, \qquad (4\mathrm{B}.10)$$

we arrive at

$$\tilde{j}(W) = \frac{\tilde{v}}{V_T} J(0) \left[1 - \frac{j\omega qk}{|J(0)|G_t} \int_0^W c(x)\left(\int_0^x G(x')\,dx' \right) dx \right]. \qquad (4\mathrm{B}.11)$$

In the above expressions $c(x)$ is the minority-carrier density and $J(0)$ is the d.c. current density, given by (4.16) and differentiation of (4B.1). The transit time τ_{qs} is, as usual, defined as:

$$\tau_{qs} \equiv \frac{q}{|J(0)|} \int_0^W c(x)dx. \qquad (4\mathrm{B}.12)$$

If we define α as

$$\alpha \equiv \frac{k}{G_t} \frac{\displaystyle\int_0^W c(x)\left[\int_0^x G(x')dx' \right] dx}{\displaystyle\int_0^W c(x)dx}, \qquad (4\mathrm{B}.13)$$

(4B.11) can be written as

$$\tilde{j}(W) = \frac{\tilde{v}}{V_T} J(0)[1 - j\omega\alpha\tau_{qs}]. \qquad (4B.14)$$

Similarly, by the substitution of (4B.4) and (4B.5) into (4B.6) for $\xi = 0$ we arrive at

$$\tilde{y}'(0) = -a\frac{\tilde{v}}{V_T} e^{V_j/V_T} -$$

$$\frac{\tilde{v}}{V_T} \frac{j\omega W^2}{D(0)} \int_0^{\xi_W} (1 - a\xi') Y_0(\xi') g^{-2}(\xi') \delta^{-1}(\xi') d\xi'. \qquad (4B.15)$$

Substitution of (4B.8) into (4B.15) gives:

$$\tilde{y}'(0) = \tilde{y}'(\xi_W) - \frac{\tilde{v}}{V_T} \frac{j\omega W^2}{D(0)} \int_0^{\xi_W} Y_0(\xi') g^{-2}(\xi') \delta^{-1}(\xi') d\xi'. \qquad (4B.16)$$

Using (4.26) and subsequent transformation to the variable x gives:

$$\tilde{j}(0) - \tilde{j}(W) = \frac{\tilde{v}}{V_T} j\omega J(0)\tau_{qs}, \qquad (4B.17)$$

so $\tilde{j}(0)$ can be written as

$$\tilde{j}(0) = \frac{\tilde{v}}{V_T} J(0)[1 + j\omega(1 - \alpha)\tau_{qs}]. \qquad (4B.18)$$

Chapter 5

EPILOGUE

In chapter one we distinguished between downscaling-related problems in the field of numerical device simulations and in the field of compact transistor modelling. In the numerical simulations the problem is to incorporate all relevant physical phenomena properly into the differential equations and boundary conditions. In compact modelling we can distinguish between problems related to the incorporation of distinct physical phenomena into the models and problems related purely to geometrical effects, i.e. related only to the transistor dimensions.

The topic of chapter two, sidewall effects on the d.c. currents in bipolar transistors, is a clear example of the latter category. In order to obtain practically useful expressions from the two-dimensional analytical calculations in this chapter, we had to make fairly rigorous assumptions, such as uniformly-doped rectangular regions. These assumptions are even more rigorous than those required in a one-dimensional calculation. For instance, as we have seen in chapter three, analytical expressions for the one-dimensional current density and minority-carrier density can also be obtained for an arbitrary dopant distribution. An analogous approach for a two-dimensional region with an arbitrary dopant profile (e.g. the base region) has not succeeded. By using a coordinate transformation similar to that used in chapter four, an attempt has been made to remove the first-order term (i.e. drift term) from the continuity equation so that in steady-state the resulting differential equation is a Laplace equation. However, in that case the boundaries of the region are no longer rectangular and, therefore, a conformal mapping technique does not lead to useful results.

Despite the rigorous assumptions in chapter two, the functional relations between process parameters and sidewall effects are found to be equally applicable to the case of a realistic dopant distribution. These expressions provide an insight into the influence of process variables on the sidewall effects on the currents and, hence, on the current

gain variation with lateral emitter scaling. They can be used, for instance, in a so-called process block in a circuit simulator. Such a process block contains all the necessary information about the scaling properties of the electrical transistor parameters in order to enable the circuit designer to choose the desired device geometry.

An interesting aspect of the above sidewall effects is that they cannot be uniquely defined, i.e. their magnitude depends on which part of the transistor is attributed to the sidewall and which part to the bottom. For instance, if in chapter two we had not defined the emitter area as the total area of the n^{++} region (see fig. 2.1), but as the contact area (which is easier to measure), the total sidewall effects on the base and collector currents would have been larger. This holds not only for sidewall effects on the currents, but also for other sidewall effects such as those on the capacitances and charges. This implies that a statement about the magnitude of a sidewall effect without a corresponding definition of the sidewall is meaningless. However, as we have seen in chapter two, the ratio of two quantities which scale with the emitter bottom area, such as the current gain, does not suffer from measurement uncertainties, because errors in the estimated emitter dimensions are approximately cancelled out. Because such a ratio depends only on process quantities it can be used as a figure of merit for process characterization. An example of such a quantity is the characteristic emitter width H_β for the characterization of the scaling properties of the current gain.

Sidewall effects on the capacitances, charges, cut-off frequency and Early voltage are not treated in this thesis. As discussed above, the determination of the sidewall effects on the capacitances and charges, which scale with the emitter area too, suffer from the same problem as the sidewall effects on the d.c. currents. However, the cut-off frequency can possibly be defined as the ratio of two quantities which scale with the emitter area, as can be seen as follows: In chapter four we have shown that a quasi-static approach for the total transit time and, therefore, for the cut-off frequency can also be used for downscaled devices. Similar to the approach for the d.c. currents used in chapter two, it might be possible to describe the total charge in the device as the sum of a term which scales with the bottom area and a term which

scales with the perimeter length. Using the relation (4.2), i.e. $\tau_{qs} = dQ/dI_c$, it might then be possible, similar to the definition of H_β, to define a characteristic emitter width H_{f_T} which characterizes the scaling properties of the cut-off frequency for a certain process. An approach such as this might also be possible for the Early voltage.

Chapter three deals with a distinct physical phenomenon, namely tunnelling in p-n junctions. Tunnelling becomes important due to the increase in doping level required for properly downscaled devices. This chapter deals both with the incorporation of these effects into a numerical device simulator and with the compact modelling of tunnelling effects in a p-n junction. The incorporation of a quantum-mechanical effect such as tunnelling into a conventional numerical device simulator is an inherently difficult task because we have to incorporate a pure wave aspect of matter into a semi-classical, particle-based description of the electrical behaviour of semiconductors. This chapter presents the basis for a recombination model containing both conventional Shockley-Read-Hall recombination and trap-assisted tunnelling. Unlike previously published models for trap-assisted tunnelling (e.g. ref. [3.13]), this model is based on the recognition that, basically, trap-assisted tunnelling and SRH recombination are the same recombination mechanisms, i.e. recombination via traps. The difference is in the character of the initial state of the electrons and holes. This recognition enabled us to unify the two mechanisms into one expression. However, in order to arrive at a model readily suitable for implementation in a device simulator the following additional work has to be done:

1) The recombination rate at a certain location, as given by (3.14) together with (3.22) and (3.23), depends on the average electric field and contains an integration over a part of the depletion layer. This means that the recombination rate at a certain mesh point j not only depends on local quantities (e.g. n_j, p_j, ψ_j, $\underline{E}_j$), but also on quantities at other mesh points. In addition, the boundary of the depletion layer must also be determined and an integration from this boundary to mesh point j must be carried out. This would be fairly time-consuming in a numerical simulation. The expressions for the tunnelling carrier densities therefore have to be simplified in order to depend on local quantities only.

2) As discussed in section 3.5, generation due to band-to-band tunnelling and tunnelling-enhanced emission from traps must also be included. These effects are important under reverse-bias conditions. Band-to-band tunnelling can easily be included by an additional term of the form of (3.48). Tunnelling-enhanced emission from traps, which can be considered as the inverse process of capture by tunnelling, must be incorporated into the expression for transitions via traps (cf. expression (3.14)).

The implications of tunnelling effects on device behaviour, such as sidewall effects on the non-ideal base current due to tunnelling, have not been discussed in chapter three. The reason for this is not that they are considered to be unimportant but the lack of adequate physical models. Before investigating these sidewall effects we had to establish physical models for these tunnelling effects, which are necessary for acquiring an insight into the importance of these effects in bipolar transistors, both qualitatively and quantitatively.

The work described in chapter four can also be classified as a modelling problem related to geometrical effects. In view of downscaling, the foundations of the quasi-static approach to determine a.c. quantities on the basis of a quasi-static analysis are reviewed and the validity of the approximations involved is discussed. Furthermore, a mathematically more rigorous low-frequency description of a.c. currents is given, based on a perturbation analysis of the time-dependent continuity equations.

From a practical point of view the work described in this chapter might seem less valuable. However, the conclusion that for downscaled devices the quasi-static approach remains a good approximation for the cut-off frequency is important for modelling the scaling properties of the cut-off frequency. The quasi-static approach is also important for our insight into the contributions of the different transistor regions to the a.c. properties of the device. In fact, relation (4.2), i.e. $\tau_{qs} = dQ/dI_c$, offers the possibility to investigate the contributions to τ_{qs} from the different transistor regions. If (4.2) were not valid, and our insight into the a.c. behaviour of a bipolar transistor therefore had to be acquired from small-signal a.c. simulations alone, this would be a very difficult task.

SAMENVATTING

Ten gevolge van de schaalverkleining van IC-processen wordt de beschrijving van het elektrische gedrag van bipolaire transistoren ingewikkelder. Bij de modellering van dit gedrag kan men onderscheid maken tussen problemen van verschillende aard. Eén klasse omvat problemen die voortvloeien uit het optreden van fysische verschijnselen die tot dusver onbelangrijk waren. In het bijzonder de noodzakelijke verhoging van de doteringsniveaus en de hiermee gepaard gaande sterke elektrische velden leiden bij schaalverkleining onder meer tot het optreden van tunneleffecten en het ontstaan van hete ladingsdragers. Een andere klasse van modelleringsproblemen is uitsluitend het gevolg van de verkleining van de transistorafmetingen. Als gevolg van de laterale schaalverkleining kan men bij geavanceerde bipolaire processen niet meer volstaan met eendimensionale beschouwingen, maar moet men overgaan op tweedimensionale berekeningen. In dit proefschrift komen beide soorten problemen ter sprake.

In hoofdstuk 2 worden randeffecten op de statische basis- en collectorstroom in verticale npn-transistoren beschreven. Met behulp van een eenvoudig tweedimensionaal transistormodel zijn analytische berekeningen uitgevoerd om inzicht te verkrijgen in het verband tussen deze randeffecten en procesparameters, zoals transistorafmetingen en doteringsniveaus. Deze berekeningen hebben, te zamen met numerieke simulaties, geleid tot praktische formules die ook voor meer ingewikkelde transistorstructuren gelden. De berekeningen tonen aan dat het randeffect op de gatenstroom in de emitter niet alleen het gevolg is van de gateninjectie vanuit de emitter-basis-overgang aan de emitterrand, maar dat ook de gateninjectie vanuit de bodem van deze overgang, binnen een zekere afstand tot de rand, een compenserende werking heeft op dit randeffect. De berekeningen laten ook zien dat de recombinatiestroom aan het basiscontact exponentieel afneemt met de afstand tussen dit contact en de emitterrand. Vervolgens worden de verkregen relaties tussen procesparameters en randeffecten getoetst aan metingen aan verschillende typen van bipolaire transistoren. Hierbij wordt ook nader ingegaan op de experimentele methoden ter bepaling van deze randeffecten. Uit zowel metingen als berekeningen

is gebleken dat, als gevolg van deze effecten, bij verkleining van de laterale emitterafmetingen de stroomversterking van sommige typen van transistoren toeneemt en van andere afneemt.

In hoofdstuk 3 worden tunneleffecten in een pn-overgang beschreven. Het tunneleffect is een quantummechanisch verschijnsel waarbij deeltjes door een potentiaaldrempel heen dringen ("tunnelen"). In dit geval zijn de deeltjes elektronen en de wordt de potentiaaldrempel gevormd door de verboden zone tussen valentie- en geleidingsband. Voor bipolaire transistoren zijn deze effecten belangrijk bij de emitter-basis overgang. Met name aan de emitterrand, waar de basisdotering het hoogst is, kunnen als gevolg hiervan relatief grote stromen lopen. Zowel tunnelen van band naar band als tunnelen via energietoestanden in de verboden zone (recombinatiecentra) komen in dit hoofdstuk aan de orde. Experimenteel en theoretisch is onderzocht of een bekende formule voor tunnelen van band naar band, afgeleid voor het geval dat het elektrische veld constant is, gebruikt kan worden voor de beschrijving van de tunnelstroom in een pn-overgang, waarbij het elektrische veld niet constant is. Het blijkt dat deze formule inderdaad toegepast kan worden, mits hierin voor het elektrische veld de maximale waarde bij de overgang wordt gebruikt en niet de gemiddelde waarde. Vervolgens wordt een nieuw recombinatiemodel voorgesteld dat, naast het conventionele Shockley-Read-Hall recombinatieproces, ook het tunnelen via recombinatiecentra beschrijft. Een vergelijking van metingen met numerieke simulaties laat zien dat het voorgestelde model de experimenteel gevonden spannings- en temperatuurafhankelijkheid van de niet-ideale diodestroom goed beschrijft. Ten slotte wordt de fysische basis voor een nieuw compact diodemodel ter beschrijving van de stroom-spanningskarakteristiek van een gesperde pn-overgang besproken. Dit model berust op een analytische oplossing van de continuïteitsvergelijking in het ruimteladingsgebied rondom een pn-overgang en beschrijft zowel tunneleffecten als lawinevermenigvuldiging. Uit een vergelijking van metingen met modelberekeningen volgt dat in het model, naast tunnelen van band naar band, ook tunnelen via recombinatiecentra meegenomen moet worden.

In hoofdstuk 4 zijn, met het oog op schaalverkleining, de aannamen in de quasi-statische beschrijving van het klein-signaal wisselspanningsgedrag van verticale npn-transistoren onderzocht. Ook wordt een nieuwe methode voor de berekening van dit gedrag besproken. Deze methode berust op een eerste-orde storingsrekening en kan dienen om, naast de gebruikelijke berekening van de afsnijfrequentie, ook op een eenvoudige manier faseverschuivingen en interne wisselspanningsgrootheden (zoals complexe stroomdichtheden en concentraties) te verkrijgen uit een statische berekening. Ten slotte zijn quasi-statische berekeningen van de afsnijfrequentie vergeleken met de resultaten van niet-quasi-statische simulaties van het wisselspanningsgedrag voor kleine signalen. Hieruit blijkt dat, ook voor moderne transistoren, de quasi-statische afsnijfrequentie goed overeen komt met berekeningen van de afsnijfrequentie uit wisselspanningssimulaties.

CURRICULUM VITAE

G.A.M. Hurkx

24-9-1956	Geboren te Best
1968-1969	MAVO St. Thomas van Aquino te Eindhoven
1969-1973	Van der Puttlyceum te Eindhoven - Eindexamen HAVO
1973-1977	Instituut voor Hoger Beroepsonderwijs te Eindhoven - Eindexamen HTS Werktuigbouwkunde - Afstudeerrichting: Meet- en Regeltechniek
1977-1979, 1982-1985	Technische Universiteit Eindhoven - Doctoraalexamen Technische Natuurkunde (met lof) - Afstudeerrichting: Theorie v.d. Vaste Stof
vanaf 1979	In dienst bij Philips Natuurkundig Laboratorium te Eindhoven

DANKWOORD

Hierbij wil ik mijn collega's en promotoren hartelijk bedanken voor hun hulp bij het tot stand komen van dit proefschrift. Voorts wil ik Marij bedanken voor haar geestelijke en daadwerkelijke steun hierbij. Tenslotte bedank ik langs deze omweg mijn ouders, die mij altijd hebben gestimuleerd om te studeren. Zonder deze houding zou dit proefschrift er nooit gekomen zijn.

STELLINGEN

behorende bij het proefschrift

On the Modelling of Downscaled Bipolar Transistors

van

G.A.M. Hurkx

27 November 1990

De methode van Fossum en Veeraraghavan voor de berekening van de niet-quasi-statische collectorstroom is alleen juist voor een transistor met een homogeen gedoteerde basis.

J.G. Fossum and S. Veeraraghavan, IEEE Electron Device Lett., EDL-7, p. 652 (1986)

2

Het model dat Woo, Plummer en Stork voorstellen om de invloed van het Poole-Frenkel-effect op het Shockley-Read-Hall-recombinatieproces te beschrijven, is niet juist.

J.C.S. Woo, J.D. Plummer and J.M.C. Stork, IEEE Trans. Electron Devices, ED-34, p. 130 (1987)

3

"Zener breakdown" bestaat niet.

4

Ten onrechte beweert Bennett herhaaldelijk dat de conventionele fysische modellen voor device-simulatie, met daarin realistische waarden van de parameters, niet toereikend zijn om de gemeten stroomversterking van bipolaire transistoren zelfs maar bij benadering te berekenen.

H.S. Bennett, IEEE Trans. Electron Devices, ED-30, p. 920 (1983)

H.S. Bennett and D.E. Fuoss, IEEE Trans. Electron Devices, ED-32, p. 2069 (1985)

H.S. Bennett, Solid-State Electron., 30, p. 1137 (1987)

5

De invloed van lawinevermenigvuldiging in de externe basis van moderne transistoren op de Early-spanning en de collector-emitter-doorslagspanning dient nader onderzocht te worden.

6

Het model dat Jo en Burk voorstellen om de interne basisweerstand van een
bipolaire transistor te beschrijven, berust op een foutieve uitdrukking voor de
spanningsafhankelijkheid van de basisstroom.

M. Jo and D.E. Burk, IEEE Trans. Electron Devices, ED-37, p. 202 (1990)

7

Om onhandige formuleringen te vermijden, dient het werkwoord "tunnelen",
voor de beschrijving van de quantummechanische doorboring van een
potentiaalberg door een deeltje, in de Nederlandse taal ingevoerd te worden.

8

De toenemende projectmatige aanpak van wetenschappelijk onderzoek, met de
bijbehorende subsidieregelingen en mijlpalen, bewijst de wetenschap een slechte
dienst; niet alleen omdat vaak op vooraf vastgestelde tijdstippen vooraf
omschreven resultaten moeten zijn behaald, maar ook omdat het onderzoekers
vaak teveel afhoudt van hun eigenlijke taak.

9

Bij de beoordeling van wetenschappelijke prestaties aan de hand van de Science
Citation Index verkrijgt men een beter beeld indien verwijzingen die verschijnen
binnen een zekere tijd na publicatie van het betreffende artikel (bijv. 2 à 3 jaar)
minder zwaar meegerekend worden.

10

Mededelingen op automatisch bewegende deuren dienen op enkele meters
afstand daarvan leesbaar te zijn.