EINDHOVEN UNIVERSITY OF TECHNOLOGY

# Analysis of bipolar transistors

*Citation for published version (APA):*
Slotboom, J. W. (1977). *Analysis of bipolar transistors*. [Phd Thesis 2 (Research NOT TU/e / Graduation TU/e), Electrical Engineering]. Technische Hogeschool Eindhoven. https://doi.org/10.6100/IR5179

*DOI:*
10.6100/IR5179

*Document status and date:*
Published: 01/01/1977

*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

*General rights*
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

*Take down policy*
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

# ANALYSIS OF BIPOLAR TRANSISTORS

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE HOGESCHOOL EINDHOVEN, OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. P. VAN DER LEEDEN, VOOR EEN COMMISSIE AANGEWEZEN DOOR HET COLLEGE VAN DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP DINSDAG 25 OKTOBER 1977 TE 16.00 UUR

DOOR

JAN WILLEM SLOTBOOM

GEBOREN TE UTRECHT

DIT PROEFSCHRIFT IS GOEDGEKEURD
DOOR DE PROMOTOREN

PROF. DR. F.M. KLAASSEN

EN

PROF. D. POLDER

Aan de nagedachtenis van mijn vader
Aan mijn moeder
Aan Annie
Aan Bas en Paul

Het in dit proefschrift beschreven onderzoek werd uitgevoerd in het Natuurkundig Laboratorium van de N.V. Philips' Gloeilampenfabrieken te Eindhoven. Ik ben de directie van dit laboratorium en vooral Ir. L.J. Tummers bijzonder erkentelijk voor de mij geboden gelegenheid dit onderzoek uit te voeren.
Bijzonder veel dank ben ik verschuldigd aan Dr. Ir. H.C. de Graaff voor de bijna dagelijkse discussies en de samenwerking bij de experimentele onderzoekingen, en aan Dr. Ir. P.A.H. Hart voor de vele waardevolle adviezen en daadwerkelijke steun tijdens het onderzoek.
Verschillende essentiële experimenten zijn alleen mogelijk geweest dankzij de bereidheid en toewijding waarmee de heren P.J.W. Jochems, H.G.R. Maas en Drs. A. Schmitz een aantal speciale transistoren hebben gemaakt.
De discussies met Dr. Ir. C. Weber over allerlei numerieke problemen hebben erg verhelderend gewerkt.
Ir. W. Smulders en de heer J. Timmermans ben ik dank verschuldigd voor de vele gesprekken over practische problemen bij het ontwerpen van transistoren.
Bij het maken van de numerieke rekenprogramma's hebben Mej. J. Mombarg en de heren A.C.M. Kilsdonk, J.P.M. Tooten en A.H.M. Goorman zeer gewaardeerde hulp verleend.
De heren R.J. van der Wal en A.P.M. van 't Hof ben ik veel dank verschuldigd voor de meetapparatuur die zij ontwikkeld hebben en waardoor nauwkeurige metingen van de stroom-spanningskarakteristieken mogelijk waren.
Tot slot wil ik ook alle anderen die op enigerlei wijze hebben bijgedragen tot het tot stand komen van dit proefschrift heel hartelijk bedanken.

CONTENTS

1. INTRODUCTION . . . . . 1
1.1. General outline . . . . . 1
1.2. Basic device equations . . . . . 2
1.3. Doping profile . . . . . 6
1.4. Mobility . . . . . 10
1.2. Recombination . . . . . 13

2. NUMERICAL ANALYSIS OF BIPOLAR TRANSISTORS . . . . . 17
2.1. Calculations without high doping effects . . . . . 17
2.1.1. The Gummel method . . . . . 17
2.1.2. Two-dimensional transistor calculations . . . . . 21
2.1.2.1. Introduction . . . . . 21
2.1.2.2. "Iterative scheme for one- and two-dimensional d.c. transistor simulation" (1969) . . . . . 22
2.1.2.3. "Computer-aided two-dimensional analysis of bipolar transistors" (1973) . . . . . 25
2.2. Calculations including high doping effects . . . . . 28
2.2.1. Discussion of TRAP 1 . . . . . 28
2.2.2. Example: I.C. transistor . . . . . 31

3. EXPERIMENTS AND CALCULATIONS OF BANDGAP NARROWING IN SILICON . . . . . 43
3.1. Introduction . . . . . 43
3.2. "Measurements of bandgap narrowing in Si-bipolar transistors" (1976) . . . . . 45
3.3. "The pn-product in silicon" (1977) . . . . . 48
3.4. Discussion . . . . . 48

4. EMITTER EFFICIENCY . . . . . 51
4.1. Introduction . . . . . 51
4.2. "Minority carrier injection into heavily doped silicon" (1977) . . . . . 55
4.3. "The emitter efficiency of bipolar transistors: Theory and Experiments" (1977) . . . . . 57
4.4. "Some aspects of LEC transistor behaviour" (1976) . . . . . 59

REFERENCES . . . . . 62

5. REPRINTS OF PUBLICATIONS . . . . . 65
5.1. "Iterative scheme for 1- and 2-dimensional dc transistor simulation", J.W. Slotboom, Electr. Lett., Vol. 5, No. 26, Dec. 1969. . . . . . 65
5.2. "Computer-aided two-dimensional analysis of bipolar transistors", J.W. Slotboom, IEEE Transact. on El. Dev., Vol. ED-20, No. 8, Aug. 1973. . . . . . 71

5.3. "Measurements of bandgap-narrowing in Si bipolar transistors", J.W. Slotboom, H.C. de Graaff, Solid-State Electr., Vol. 19, p. 857, 1976. . . . . . 96

5.4. "The pn-product in silicon", J.W. Slotboom, Solid-State Electr., Vol. 20, p. 279, 1977. . . . . . 102

5.5. "Bandgap narrowing in silicon bipolar transistors". J.W. Slotboom, H.C. de Graaff, IEEE Transact. on Electr. Dev., Vol. ED-24, Aug. 1977. . . . . . 107

5.6. "Minority carrier injection into heavily doped silicon", J.W. Slotboom, Solid-State Electr., Vol 20, p. 167, 1977. . . . . . 112

5.7. "The emitter efficiency of bipolar transistors: Theory and Experiments", H.C. de Graaff, J.W. Slotboom, A. Schmitz, Solid-State Electr., Vol. 20, 1977. . . . . . 116

5.8. "Some aspects of LEC transistor behaviour", H.C. de Graaff, J.W. Slotboom, Solid-State Electr., Vol. 19, p. 809, 1976. . . . . . 123

SUMMARY . . . . . . 129

SAMENVATTING . . . . . . 131

# 1. INTRODUCTION

## 1.1. General outline

Analysis of semiconductor devices is an important means of obtaining a clearer picture of what is really going on inside the device, and how this is related with doping profile, geometry, lifetime etc. and with external electrical characteristics measured at the device terminals. Owing to the nonlinear character of this problem it is not possible to solve it analytically without introducing a large number of simplifications e.g. low injection, constant mobilities and lifetimes, simplified doping profiles etc. In the classical analysis of pn-junction devices [1] the device is divided into neutral and space charge regions. In each region the equations are solved analytically using assumptions that are reasonable for that region, and these solutions are fitted at the boundaries. In spite of the simplifications used, this solving procedure often gives a very practical and useful description.

Since the invention of the transistor there has been a continuous trend in technology towards reduced device dimensions. This may be illustrated by a reduction of the thickness of the base region in a high-frequency bipolar transistor from several microns towards one tenth of a micron or even less. This rapid technological development has made several of the above-mentioned classical assumptions problematic, while at the same time the perfection of device design needed a more quantitative and generally applicable description of device behaviour. Circuit analysis programs moreover, which are necessary for the analysis and design of large integrated circuits, also require accurate device characterization.

With the help of modern computers and numerical analysis techniques it is possible to analyse the transistor problem while taking into account realistic device data and the basic physical equations as accurately as they are known. Possible improvements in physical formulation or the influence of one particular parameter can easily be investigated. In 1964 Gummel [2] published *a self-consistent iterative scheme for one-dimensional steady-state transistor calculations* which described for the first time a general way of solving the transistor problem while considering the transistor as an entity and applying boundary conditions only at points representing contacts. This work started a whole train of numerical investigations for other semiconductor devices as well, and the method has come into general use for device analysis. Right at the beginning of this development it became clear that this new method of analysis was seriously hindered by a number of typical numerical problems such as:

- the small differences between nearly equal numbers;
- carrier concentrations that vary many decades over small distances;
- numerical instability of the difference equations;
- slow convergence rate of iterative processes.

In addition to these problems gradually several weaknesses in the physical formulation manifested themselves such as:

- influence of impurity concentration on bandgap and carrier lifetime due to fundamental physical phenomena (high doping effects);
- influence of the fabrication process on doping profile, carrier lifetime, etc.;
- uncertainties concerning the temperature dependence of carrier mobilities and lifetimes.

In this thesis recent developments of transistor analysis will be discussed in the light of several of the author's contributions in this field.
Chapter 1 presents the basic device equations, including the influence of heavy doping, and touches on quantities such as doping profile, mobility and recombination. In chapter 2 the numerical analysis of bipolar transistors is studied with emphasis on the numerical problems. In the second part of this chapter a recently developed one-dimensional transistor analysis program (TRAP 1) is presented and calculated results are compared with measurements. This program uses experimentally determined values for the bandgap narrowing and minority carrier lifetime as a function of impurity concentration.
In chapter 3 measurements and calculations concerning bandgap narrowing are discussed in detail.
Chapter 4 describes the direct consequences of this effect for the emitter efficiency in normal transistors as well as in a new transistor structure, which has an extra emitter region with a low impurity concentration, and the theory is compared with experiments.
Reprints of the publications discussed in the foregoing chapters are included in chapter 5.

## 1.2. Basic device equations

The basic physical equations describing the flow of electrons and holes in the semiconductor under steady state conditions are [3] :

$$\frac{dE}{dx} = \frac{q}{\epsilon}\ (p - n + N_D - N_A) \qquad \text{(Poisson's equation)} \qquad (1.1)$$

(Continuity equations)

$$\frac{dJ_p}{dx} = -qR \qquad (1.2)$$

$$\frac{dJ_n}{dx} = qR \qquad (1.3)$$

where the electric field $E = -d\Psi/dx$ and the hole and electron current densities are

$$J_p = -q\mu_p \, p \frac{d\varphi_p}{dx} \tag{1.4}$$

$$J_n = -q\mu_n \, n \frac{d\varphi_n}{dx} \tag{1.5}$$

and further

$R$ = the generation-recombination rate
$N_D - N_A$ = the net ionized impurity concentration
$\mu_p, \mu_n$ = hole and electron mobility
$\varphi_p, \varphi_n$ = hole and electron quasi-Fermi potential (see below).

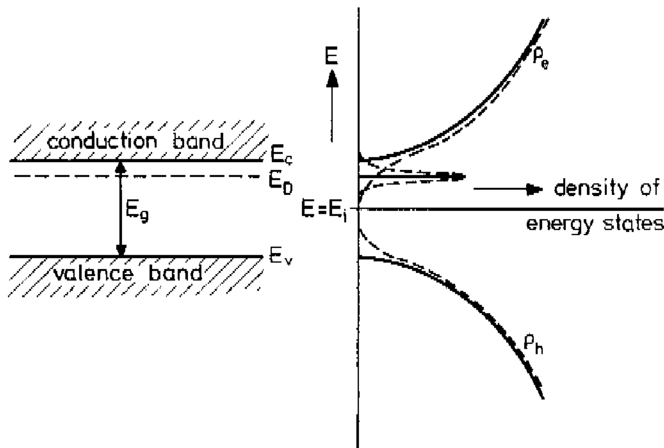



Fig. 1.1. Schematic representation of the bandgap $E_g$ and the density of energy states functions for electrons ($\rho_e$) and holes ($\rho_h$) for an n-type semiconductor. At high impurity concentrations an impurity band and tails are formed.

In equilibrium the carrier concentrations are given by (fig. 1.1):

$$n_0 = \int_{E_c}^{\infty} \rho_e(E) \, f(E) \, dE \tag{1.6}$$

$$p_0 = \int_{-\infty}^{E_v} \rho_h(E) \left\{1 - f(E)\right\} dE \ . \tag{1.7}$$

The density of energy states for the electrons ($\rho_e$) and the holes ($\rho_h$) are usually approximated by a square root dependence on energy. The Fermi-Dirac distribution function f(E) gives the probability that an energy level E is occupied by an electron:

$$f(E) = \frac{1}{1 + \exp\left\{(E - F)/kT\right\}} \tag{1.8}$$

and F is the Fermi energy level.
The position of the Fermi level in the bandgap can be derived from eqs. (1.6) and (1.7) and the charge neutrality condition:

$$p_0 - n_0 + N_D - N_A = 0 \; . \tag{1.9}$$

When the allowed energy levels are at a distance of more than about 2 kT/q from F, then the Fermi-Dirac distribution can be approximated by the Maxwell-Boltzmann function:

$$f(E) = \exp\left\{-(E-F)/kT\right\} \; . \tag{1.10}$$

The material is then called non-degenerate. Assuming Maxwell-Boltzmann statistics eqs. (1.6) and (1.7) can be integrated and the result written as

$$n_0 = N_c \exp\left\{-(E_c - F)/kT\right\} \tag{1.11}$$

$$p_0 = N_v \exp\left\{-(F - E_v)/kT\right\} \tag{1.12}$$

where $N_c$ and $N_v$ are the effective density of states in the conduction and valence bands.
For an intrinsic semiconductor the hole and electron carrier concentrations are equal:

$$p_0 = n_0 = n_{io} \tag{1.13}$$

It is possible to express the carrier concentration in extrinsic material in the intrinsic carrier concentration $n_{io}$ :

$$n_0 = n_{io} \exp\left\{-(E_i - F)/kT\right\} = n_{io} \exp\left\{q(\Psi - \varphi)/kT\right\} \tag{1.14}$$

$$p_0 = n_{io} \exp\left\{-(F - E_i)/kT\right\} = n_{io} \exp\left\{q(\varphi - \Psi)/kT\right\} \tag{1.15}$$

where $E_i$ is the Fermi level in intrinsic material:

$$E_i = \tfrac{1}{2}(E_c + E_v) + \tfrac{1}{2}kT \ln(N_v/N_c) \tag{1.16}$$

and $\Psi = -E_i/q$ and $\varphi = -F/q$ are the electric and Fermi potentials. The pn-product is:

$$p_o n_o = N_C N_V \exp(-E_g / kT) \tag{1.17}$$

where the bandgap $E_g(T) = E_c - E_v$ (see fig. 1.1).

Optical absorption measurements on pure silicon show that at temperatures above 250 K the bandgap can be approximated by a linear function of temperature (see 5.3, fig. 3):

$$E_g(T) = E_{go} - \alpha T \tag{1.18}$$

with $E_{go} = 1.206$ eV and $\alpha = 2.8 \cdot 10^{-4}$ eV / K .
Measurements of the pn-product as a function of temperature in pure silicon (see 5.3) are described by

$$p_o n_o = n_{io}^2 = 9.61 \cdot 10^{32}\, T^3 \exp(-1.206 / kT) \,. \tag{1.19}$$

This is in reasonably good agreement with the theoretical formula [4] assuming parabolic density of states functions independent of the impurity concentration. However it will be shown in chapter 3 that this is not justified in het case of impurity concentrations above $10^{17}$ cm$^{-3}$. An impurity band and bandtails are then formed which reduce the bandgap. This effect increases with higher impurity concentrations and as a result the bandgap and also the pn-product are not constant throughout the device but are position-dependent. For practical purposes it is useful to approximate the pn-product by

$$p_o n_o = n_i^2(N,T) = 9.61 \cdot 10^{32}\, T^3 \exp\left\{-q V_{go}(N) / kT\right\} \tag{1.20}$$

where $V_{go}(N)$ is interpreted as the dope-dependent bandgap (see chapter 3).

In non-equilibrium it is useful to define quasi-Fermi potentials $\varphi_p$ and $\varphi_n$ by analogy with the Fermi level in equilibrium:

$$n = n_i \exp\left\{q(\Psi - \varphi_n) / kT\right\} \tag{1.21}$$

$$p = n_i \exp\left\{q(\varphi_p - \Psi) / kT\right\} \,. \tag{1.22}$$

These two expressions assume Maxwell-Botzmann statistics and the only difference compared with eqs. (1.14) and (1.15) is that $n_i$ is dope-dependent because the density of states functions depend on the doping level and therefore on position. Substitution of eqs. (1.21) and (1.22) into the expressions for the current denstities gives

$$J_p = q\mu_p\, p\,(E + \frac{kT}{q}\, \frac{1}{n_i}\, \frac{dn_i}{dx}) - q\, D_p\, \frac{dp}{dx} \qquad (1.23)$$

$$J_n = q\mu_n\, n\,(E - \frac{kT}{q}\, \frac{1}{n_i}\, \frac{dn_i}{dx}) + q\, D_n\, \frac{dn}{dx} \qquad (1.24)$$

where use has been made of the Einstein relationship

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = \frac{kT}{q}\ . \qquad (1.25)$$

In the current density equations an extra term $kT/q \cdot d(\ln n_i) / dx$ appears which describes the position-dependent bandgap throughout the device. These transport equations in heavily doped silicon have been discussed in more detail by van Overstraeten c.s. [5], Mock [6] and by Marshak and van Vliet [7].

## 1.3. Doping profile

For detailed numerical analysis of semiconductor devices it is necessary to know accurately the doping profile. Unfortunately it is very difficult and laborious to determine.
Several of the most commonly used techniques are:

- stripping of thin layers after anodic oxidation and measuring the incremental sheet resistance [8];
- measuring the spreading resistance when stepping with probes along a bevelled angle [9];
- sputtering the wafer with an ion beam and measuring the secondary ions by means of a mass spectrometer (SIMS [10]);
- measuring the capicitance of a Schottky diode as a function of voltage.

The problem to derive the doping profile directly from the diffusion equations, taking into account temperature, diffusion time, ambient etc., has not yet been solved satisfactorily, although recently [11] a process modelling of this type has been claimed to give quantitatively reliable results.

Another more indirect way is to approximate the doping profile by a set of analytical expressions [12]. The parameters in these expressions are chosen in agreement with sheet resistances, epitaxial layer thickness and resistivity, junction depths etc., and are further derived from electrical measurements at the device terminals. Based on this idea we developed a method that will be illustrated for a double-diffused NPN transistor (see fig. 1.2).

First only the doping profile in the base and neighbouring junctions will be considered. It is assumed that in this region the profile N(x) can be approximated

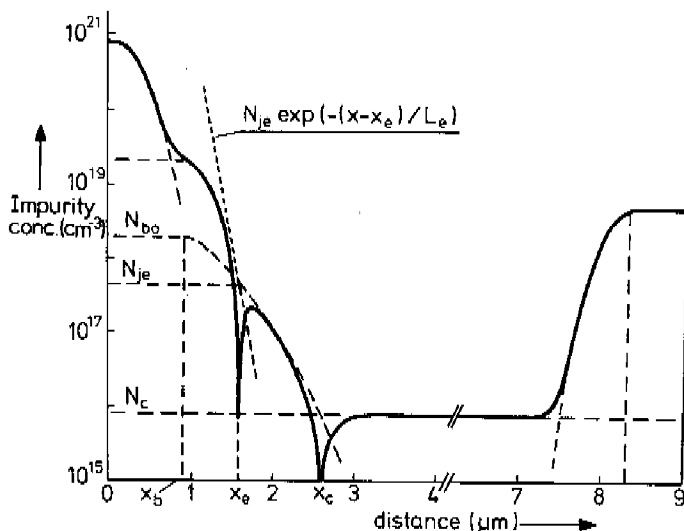



Fig. 1.2. Modelled doping profile of an NPN transistor.

by the following expressions

$$N(x) = N_D(x) - N_A(x)$$

$$N_D(x) = N_{je} \exp\left\{-(x - x_e)/L_e\right\} + N_c \tag{1.26}$$

$$N_A(x) = \begin{cases} N_{bo} & \text{for } 0 \leqslant x \leqslant x_b \\ N_{bo} \exp\left\{-\left[(x - x_b)/L_b\right]^2\right\} & \text{for } x \geqslant x_b \end{cases}$$

The interpretation of the parameters in these expressions follows from fig. 1.2. In order to determine these parameters we assumed that the junction depths $x_e$ and $x_c$, the impurity concentration in the collector $N_c$ and the sheet resistance of the base underneath the emitter

$$R_{\square EB}^{-1} = \int_{x_e}^{x_c} q\,\mu_p\,N\,dx \tag{1.27}$$

($\mu_p$ is dope-dependent, see fig. 1.5) are known. The gradient of the doping profile in the junction regions, which can be derived from capacitance measurements (see e.g. [13], [14]), is important for the analysis of the cut-off frequency $f_T$ and should

therefore also be taken into account for the parameter determination. For small forward and reverse bias the junctions can be considered as linear and the gradient of the profile in the junctions can be derived from the slope of a $C^{-3}$ versus V plot (see fig. 1.3).

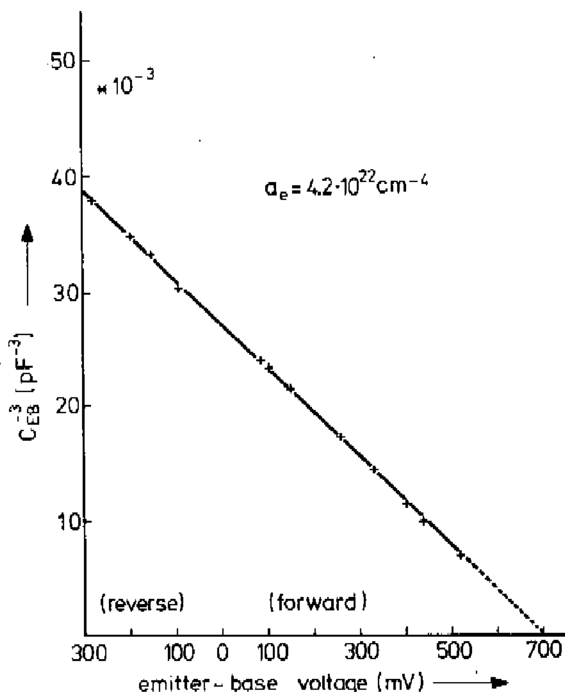



Fig. 1.3. $C^{-3}$ (V) plot of measured emitter-base capacitance for low voltages. $a_e$ is the gradient of the doping profile in the emitter-base junction (emitter junction area $A_E = 3.6 \cdot 10^{-5}$ $cm^2$).

Because $x_e$, $x_c$ and $N_c$ are known, the other five parameters in eq. (1.26) can be determined from the following relations:

$$N(x_e) = 0, \ N(x_c) = 0, \ \left.\frac{dN}{dx}\right|_{x_e} = a_e, \ \left.\frac{dN}{dx}\right|_{x_c} = a_c \text{ and eq. (1.27).}$$

Once the parameters $N_{je}$, $L_e$, $N_{bo}$, $x_b$ and $L_b$ have been calculated the profile in the base and junction regions is given by eq. (1.26).

The profile in the emitter is usually approximated by

$$N_D(x) = \begin{cases} N_{eo} & \text{for } 0 \leqslant x \leqslant x_e \\ N_{eo} \exp\left\{-[(x-x_e)/l_e]^2\right\} + N_c & \text{for } x \geqslant x_e \end{cases} \quad (1.28)$$

where $N_{eo}$ and $l_e$ are chosen such that $dN/dx \mid_{x_e} = -N_{je}/L_e$ and the sheet resistance of the emitter is

$$R_{\square E}^{-1} = \int_0^{x_e} q\,\mu_n\,N\,dx\ . \quad (1.29)$$

On the collector side the profile can be extended straight-forwardly in accordance with the data on the epitaxial layer thickness, substrate concentration and out-diffusion.

In fig. 1.2. an example of a similar procedure is shown where in this case, as a refinement, in the emitter profile two gaussian curves are used because extra data from profile measurements, made by the anodic oxidation and stripping method, were available. This detail is relevant to the calculation of the recombination in the emitter region (see section 4.1).

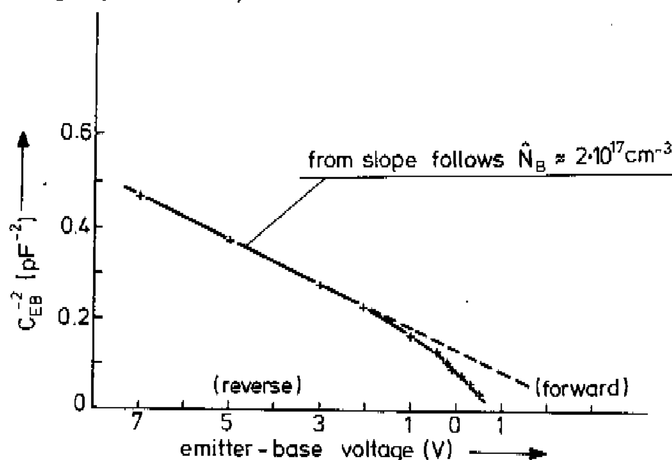



Fig. 1.4. $C^{-2}$ (V) plot of measured emitter-base capacitance for reverse voltages (emitter junction area $A_E = 3.6 \cdot 10^{-5}$ cm$^2$).

The junction capacitances have been measured at higher reverse biases, which made it possible [14] to derive from the $C^{-2}$ (V) plot the maximum base impurity concentration ($\hat{N}_B$) (see fig. 1.4), and the collector impurity concentration $N_C$. The derived values agree well with the modelled profile.

## 1.4. Mobility

The mobility of the electrons and holes in the crystal depends upon several scattering mechanisms which are difficult to describe theoretically. For the device analysis it is better to use empirical expressions based on experimental data (for a review of mobility data see [15]). The dependence of the mobility on the *impurity concentration* is quite well known from experiments [16] and can be described by the following fit formula [17] :

$$\mu(N) = \frac{\mu_{max} - \mu_{min}}{1 + (N / N_{ref})^{\alpha}} + \mu_{min} \; . \tag{1.30}$$

With this expression it is possible to approximate the experimental data with the parameter values as given in Table 1. The corresponding mobility curves are shown in fig. 1.5.

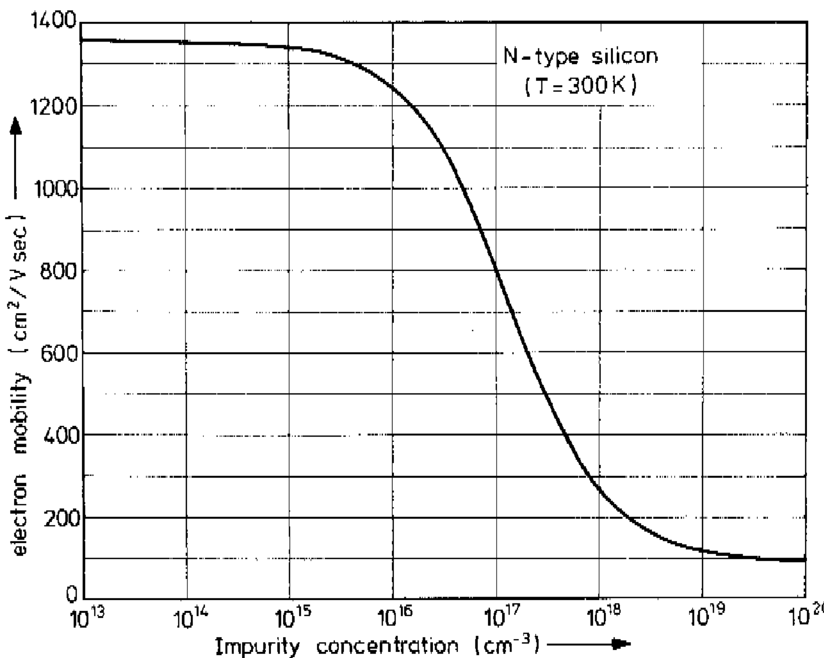



Fig. 1.5a. Electron mobility as a function of impurity concentration.

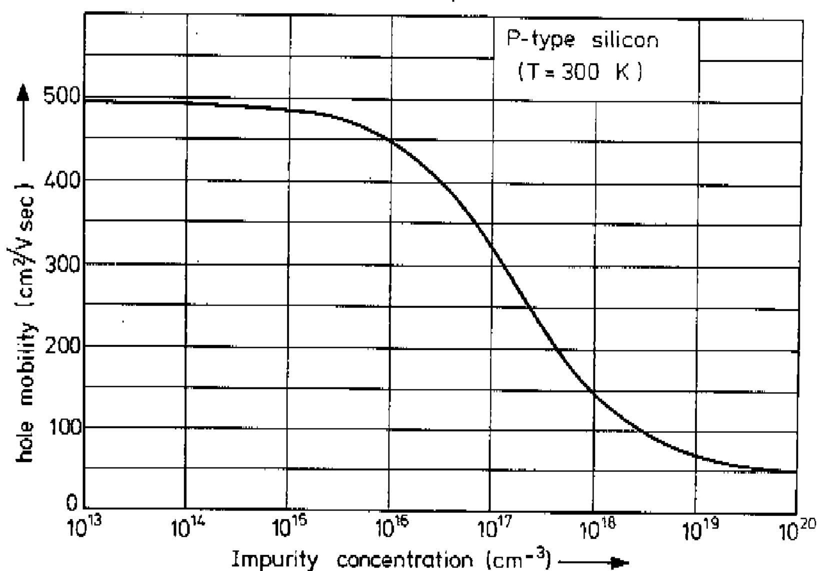



Fig. 1.5b. Hole mobility as a function of impurity concentration.

**Table 1** (T = 300 K)

| | $\mu_{max}$ (cm²/Vsec) | $\mu_{min}$(cm²/Vsec) | $\alpha$ | $N_{ref}$(cm⁻³) | $v_m$(cm/sec) | $\beta$ |
|---|---|---|---|---|---|---|
| holes | 495 | 47.7 | 0.76 | $1.9 \cdot 10^{17}$ | $9.5 \cdot 10^6$ | 1 |
| electrons | 1360 | 92 | 0.91 | $1.3 \cdot 10^{17}$ | $1.1 \cdot 10^7$ | 2 |

Several papers have dealt with the experimental determination of the drift velocity as a function of *electric field* [18] (see fig. 1.6). This field dependence can be approximated [17] by

$$\mu(E) = \frac{\mu_0}{\{1 + (E / E_c)^{\beta}\}^{1/\beta}} \tag{1.31}$$

where $\mu_0 = v_m/E_c$ and $v_m$ is the saturated drift velocity at high electric fields. The values for the parameters $v_m$ and $\beta$ are given in Table 1.
For the combined effect of N and E on the mobility the dope-dependent value $\mu$(N) from eq. (1.30) is substituted for $\mu_0$ in eq. (1.31).

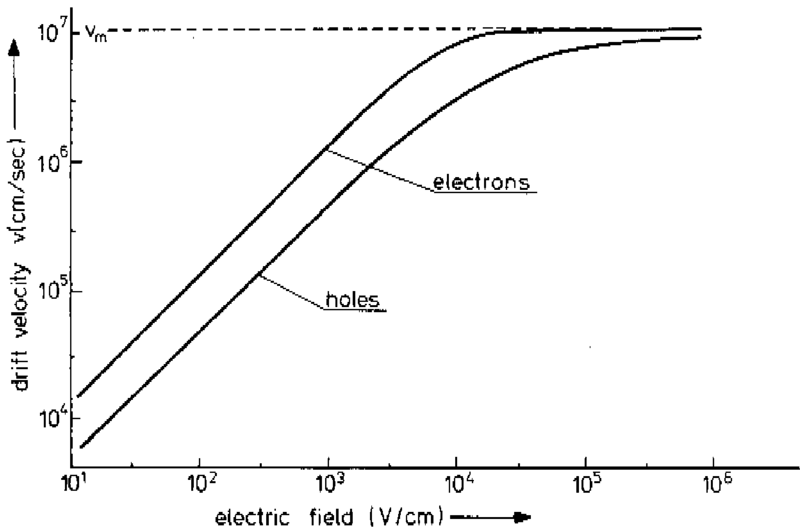



Fig. 1.6. Carrier drift velocity as a function of electric field.

Although the literature contains little quantitative data about the influence of *temperature* on mobility at different impurity concentrations [19] it is an important parameter for the temperature behaviour of the transistor characteristics. Usually it is assumed that the total mobility is caused by lattice scattering ($\mu_L$) and impurity scattering ($\mu_I$) and given by the following relationship :

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I} . \tag{1.32}$$

The temperature dependence of $\mu_L$ is given by [20] :

$$\mu_L (T) = \mu_L (T = 300) \cdot \left( \frac{T}{300} \right)^{-\eta}$$

where $\eta$ is 2.7 for holes and 2.5 for electrons and $\mu_L$ (T = 300 K) equals $\mu_{max}$ in eq. (1.30).

The temperature dependence of $\mu_I$ is modelled as :

$$\mu_I (T) = \mu_I (T = 300) \cdot \exp \left\{ 0.5 \cdot 10^{-3} (T - 300) \right\} . \tag{1.33}$$

In fig. 1.7 the result of this modelled temperature dependence of the electron and hole mobility is shown.

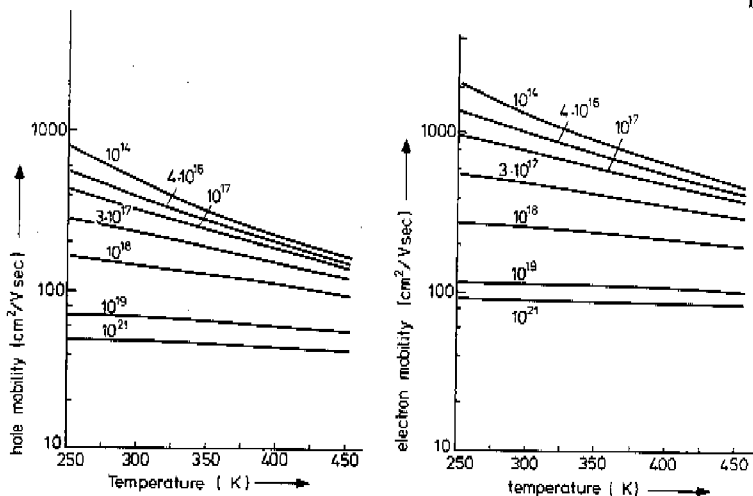



Fig. 1.7. Carrier mobilities as a function of temperature with the impurity concentration ($cm^{-3}$) as a parameter.

## 1.5. Recombination

Recombination in bipolar transistors is important because it determines particularly the base current and its dependence upon voltage and temperature. In [21] is shown that the total steady-state recombination rate of electrons and holes is given by the following expression

$$R = (pn - n_i^2)\ [(B + C_n n + C_p p) + \{\tau_{po}(n + n_1) + \tau_{no}(p + p_1)\}^{-1}] \quad (1.34)$$

which can be considered as a generalization of the usual Hall-Shockley-Read recombination because also direct band to band and Auger recombination are included. The band to band recombination, characterized by the coefficient B, will be neglected because for silicon the lifetimes involved are much larger than experimentally determined values.

Hall, Shockley and Read [22] have shown that recombination-generation via traps having energy levels within the forbidden energy gap, due to imperfections in the periodicity of the crystal lattice, play a dominant role in silicon devices. This mechanism is described by the following expression :

$$R_{HSR} = \frac{pn - n_i^2}{\tau_{po}(n + n_1) + \tau_{no}(p + p_1)} \quad (1.35)$$

the lifetimes $\tau_{po}$, $\tau_{no}$ being given by $(v_{th} N_t \sigma_{p,n})^{-1}$ where $v_{th}$ is the thermal velocity ($\sim 10^7$ cm/sec), $N_t$ the trap density and $\sigma_{p,n}$ the hole and electron capture cross-sections. The carrier concentrations $p_1$ and $n_1$ are the hole and electron concentrations that would exist if the Fermi level were at the trap level $E_t$ :

$$n_1 = n_i^2/p_1 = n_i \exp\left\{(E_t - E_i)/kT\right\} . \tag{1.36}$$

The three parameters $E_t$, $\tau_{po}$ and $\tau_{no}$ depend strongly upon the processing of the device and will vary throughout the device. As a rule not much is known quantitatively about these parameters and therefore as a first approximation it will be assumed that

$$\begin{aligned} \tau_{po} &= \tau_{no} = \tau_0 \\ p_1 &= n_1 = n_i \end{aligned} \tag{1.37}$$

Now the recombination is characterized by only one parameter $\tau_0$. Measurements of IV characteristics indicate that the lifetime in the junction is about $1 - 0.1$ $\mu$sec. Furthermore it is plausible that the trap density $N_t$ will increase for higher impurity concentrations and therefore $\tau_0$ will decrease. In practical devices $\tau_0$ varies roughly between about 20 $\mu$sec for $N = 10^{14} - 10^{15}$ towards 1 $\mu$sec for $N \approx 10^{17}$ cm$^{-3}$.

Recent experiments (see fig. 1.8) show that the minority carrier lifetime decreases strongly for high impurity concentrations above about $10^{17} - 10^{18}$ cm$^{-3}$ [23, 24, 25]. This phenomenon is ascribed to Auger recombination, where an electron recombines with a hole and the excess energy is transferred to another carrier as kinetic energy (this is the inverse of the impact ionization process). This recombination mechanism is described by (see eq. (1.34))

$$R_{AU} = C_n (n^2 p - n n_i^2) + C_p (n p^2 - n_i^2 p) . \tag{1.38}$$

The coefficients $C_p$ and $C_n$ have not only been derived from lifetime measurements of minority carriers in heavily doped regions [23, 24] but also for pn plasmas with high carrier concentrations [26, 27] and agree rather well with each other ($C_p \approx C_n \approx 1.5 \cdot 10^{-31}$ cm$^6$sec$^{-1}$). In [24] is shown that $C_p$ and $C_n$ are nearly independent of temperature measured for 77,300 and 400 K. Measurements of the recombination-diffusion length [25] as a function of impurity concentration are also shown in fig. 1.8 (these measurements have been converted to lifetimes using $L = \sqrt{D\tau}$ together with the mobility values given in fig. 1.5).

It appears from these experimental data that they agree rather well and spread relatively little at high impurity concentrations, whereas at the lower concentrations the spread is much larger. This is probably connected with the different origins: at

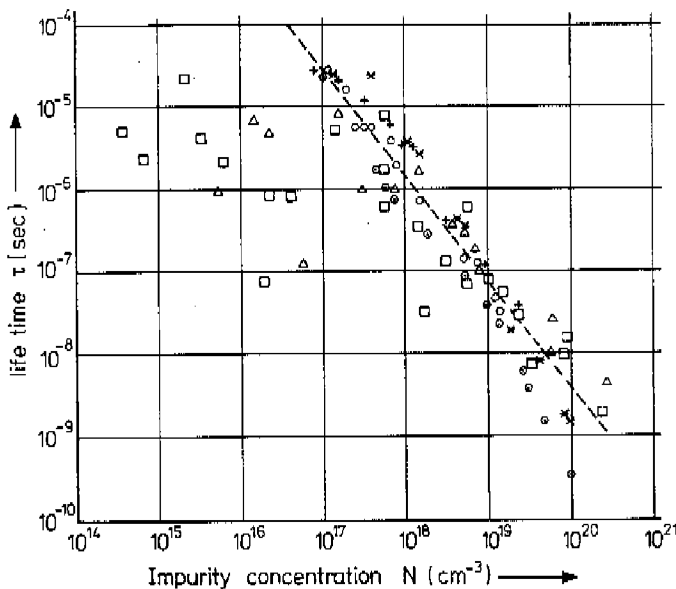



Fig. 1.8. A survey of experimentally derived minority carrier lifetimes as a function of impurity concentration (T = 300 K) :
○, + are measurements of Beck and Conradt [23] ;
□, △ are measurements of Iles and Soclof [25] ;
⊗, ✳ are measurements of Dziewior and Schmid [24] ;
(○, □ and ⊗ are lifetime measurements in n-type silicon and the other symbols represent lifetimes in p-type silicon).

high impurity concentrations Auger recombination dominates (a basic physical phenomenon) while at low impurity concentrations recombination via traps is important (a more process-dependent phenomenon).

In the analysis of NPN transistors both recombination mechanisms are taken into account as

$$R = \frac{pn - n_i^2}{\tau_0\,(p + n + 2\,n_i)} + 2.25\,10^{-19}\,p\,N^{1.36} \,. \tag{1.39}$$

The second term is a fit of the experimental data at high impurity concentrations. The surface recombination is characterized by the surface-recombination velocity s defined by

$$J_p\,(0) = -q\,s\,(p(0) - p_0(0)) \tag{1.40}$$

for holes, recombining at the surface ($x = 0$) of the n-type emitter. The concentrations $p(0)$ and $p_0(0)$ are the holes at $x = 0$ under bias and equilibrium conditions. When the surface-recombination dominates the electrical device characteristics, it has to be considered in more detail and can no longer be characterized by one constant parameter $s$.

# 2. NUMERICAL ANALYSIS OF BIPOLAR TRANSISTORS

## 2.1. Calculations without high doping effects

### *2.1.1. The Gummel method*

The classical formulation of the basic equations governing the carrier flow in one dimension and in the steady state in semiconductor devices is, in normalized form:

$$\frac{d^2\Psi}{dx^2} = e^{\Psi - \varphi_n} - e^{\varphi_p - \Psi} - N \tag{2.1}$$

$$\frac{dJ_p}{dx} = -R \qquad \text{with } J_p = -\frac{1}{\gamma_p} e^{\varphi_p - \Psi} \frac{d\varphi_p}{dx} \tag{2.2}$$

$$\frac{dJ_n}{dx} = R \qquad \text{with } J_n = -\frac{1}{\gamma_n} e^{\Psi - \varphi_n} \frac{d\varphi_n}{dx} \tag{2.3}$$

where $n = \exp(\Psi - \varphi_n)$, $p = \exp(\varphi_p - \Psi)$ and $\gamma_p^{-1}$ and $\gamma_n^{-1}$ are normalized mobilities. The norm factors used are given in Table 2.

**Table 2 (T = 300 K)**

| description | normalized quantity | normalization factor | |
|---|---|---|---|
| | | symbol | numerical value |
| position coordinates | $x, y$ | $L_D = \sqrt{\frac{kT}{q}\frac{\epsilon}{q\,n_{io}}}$ | $3.64 \cdot 10^{-3}$ cm |
| time | $t, \tau$ | $L_D^2 / D_0$ | $1.33 \cdot 10^{-5}$ sec |
| potentials | $\Psi, \varphi_p, \varphi_n, V$ | $kT/q$ | 0.02588 V |
| carrier diffusion constants | $\gamma_p^{-1}, \gamma_n^{-1}$ | $D_0$ | 1 $cm^2$/sec |
| carrier mobilities | $\gamma_p^{-1}, \gamma_n^{-1}$ | $qD_0/kT$ | 38.64 $cm^2$/Vsec |
| carrier concentrations | $p, n$ | $n_{io}$ | $1.22 \cdot 10^{10}$ $cm^{-3}$ |
| impurity concentrations | $N, N_D, N_A$ | $n_{io}$ | $1.22 \cdot 10^{10}$ $cm^{-3}$ |
| generation-recombination rate | $R$ | $D_0\, n_{io}/L_D^2$ | $9.20 \cdot 10^{14}$ $cm^{-3}sec^{-1}$ |
| electric field | $E$ | $kT/qL_D$ | 7.11 V/cm |
| capacitance | $C$ | $\epsilon/L_D$ | $2.75 \cdot 10^{-10}$ $F/cm^2$ |
| current densities | $J, J_p, J_n$ | $qD_0\, n_{io}/L_D$ | $5.36 \cdot 10^{-7}$ $A/cm^2$ |

This system of equations can be considered as three second-order nonlinear differential equations in $\Psi$, $\varphi_p$ and $\varphi_n$, while for $R$ and $\gamma_p$, $\gamma_n$ the nonlinear functions as described in sections 1.4 and 1.5 are used.

As mentioned in chapter 1, Gummel made an important breakthrough with his numerical iteration scheme for solving this problem for a one-dimensional transistor structure. He formulated the problem as a boundary value problem, specifying two sets of boundary conditions for the potentials at the emitter and collector contact following from the assumption of charge neutrality ($p - n + N = 0$) and equilibrium ($pn = n_{io}^2$) at these boundaries. Moreover the quasi-Fermi potential for the majority carriers in the base is specified at a point B in the base region representing the base contact (see fig. 2.1). Since the base current is relatively small the quasi-Fermi potential in the p-type base region will be nearly constant. Consequently the choice of the base point B is not critical.

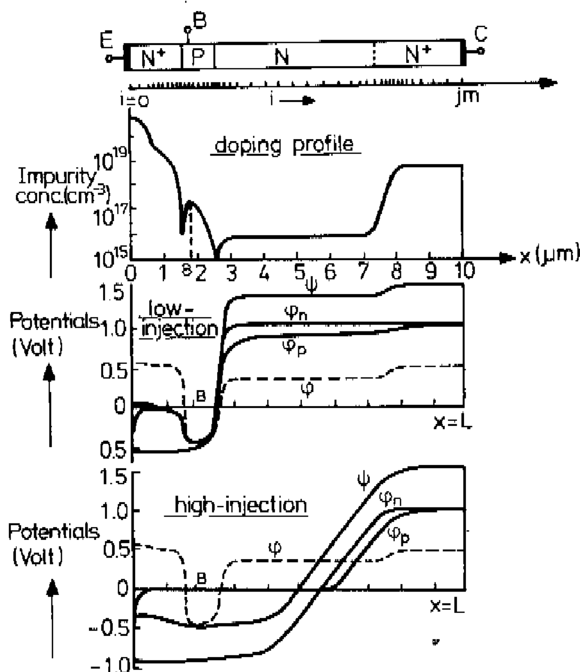



**Fig. 2.1.** Doping profile and calculated potential distributions of $\Psi$, $\varphi_p$ and $\varphi_n$ for low injection ($V_{EB}$ = 520 mV) and high injection ($V_{EB}$ = 880 mV). The reverse-bias collector base voltage is 1V. The dashed line gives the Fermi potential in equilibrium. The non-uniform step size distribution is schematically indicated ($i = 0, 1, \ldots, jm$).

These three differential equations are approximated by three sets of difference equations after discretizing the variables on a finite number of mesh points and are solved successively as follows:

(1) Poisson's equation is considered as a nonlinear two-point boundary value problem in $\Psi$ with the boundaries $x = 0$ and $x = L$. The mathematical aspects of this problem have been considered in detail in [28]. Assuming an approximate solution $\Psi(x)$ of eq. (2.1), a correction $\delta(x)$, given by

$$\Psi_{new}(x) = \Psi(x) + \delta(x)\,, \tag{2.4}$$

must satisfy after Newton-linearization the following equation:

$$\frac{d^2\delta}{dx^2} - \delta\left(e^{\varphi_p - \Psi} + e^{\Psi - \varphi_n}\right) = -\frac{d^2\Psi}{dx^2} + e^{\Psi - \varphi_n} - e^{\varphi_p - \Psi} - N \tag{2.5}$$

or in finite difference approximations:

$$a_{i-1}\,\delta_{i-1} - a_i\delta_i + a_{i+1}\,\delta_{i+1} = b_i \tag{2.6}$$

$(i = 1, 2, \ldots jm - 1)$
where $a_{i-1} = a_{i+1} = 1$, $a_i$ and $b_i$ are functions of $\varphi_p$, $\varphi_n$ and $\Psi$.
In matrix notation eq. (2.6) is

$$\mathbf{A\delta} = \mathbf{b} \tag{2.7}$$

where the coefficient matrix A is tridiagonal. This set of equations can easily be solved by gaussian elimination [28]. For $\varphi_p$, $\varphi_n$ and $\Psi$ first guesses or results from a previous iteration are used. With the calculated corrections $\delta_i$ the "old" $\Psi_i$ values are corrected. Although an accurate solution of Poisson's equation alone requires an iterative process, only one $\delta$-iteration is often enough, because $\varphi_p$ and $\varphi_n$ have yet to be changed in the next step of the overall iteration procedure.

(2) With the above described improved electric potential, the continuity equations are solved by integration, which avoids numerical instability (see 2.1.2.2). The electron current density equation is integrated for $0 \leqslant x \leqslant L$:

$$e^{-\varphi_n(x)} = \int_0^x J_n(t)\,\gamma_n(t)\;e^{-\Psi(t)}\,dt + \text{Const.} \tag{2.8}$$

$J_n(x)$ and the integration constant are found from integration of the recombination (see eq. (2.3)) and from the boundary conditions. When recombination can be neglected ($J_n$ = const.) the well-known relationship (unnormalized) follows:

$$J_n = \frac{q\,n_{io}^{\,2}}{\int_0^L \frac{1}{D_n}\,p\,dx}\left\{e^{-qV_{EB}/kT} - e^{-qV_{CB}/kT}\right\}\,. \tag{2.9}$$

The denominator is usually called the "Gummel number" ($G_B$) and is in general a function of $V_{EB}$ and $V_{CB}$.
For low injection the hole concentration equals the net base impurity concentration and then eq. (2.9) is the usual Moll-Ross relation [29].

(3) In a similar way, but now separately in the emitter-base region ($x \leqslant B$) and in the base-collector region ($x \geqslant B$), the hole current density equation is integrated:

$$e^{\varphi_p(x)} = -\int_0^x J_{pe}(t)\,\gamma_p(t)\;e^{\Psi(t)}\,dt + \text{Const.} \qquad (x \leqslant B) \tag{2.10}$$

$$e^{\varphi_p(x)} = -\int_B^x J_{pc}(t)\,\gamma_p(t)\;e^{\Psi(t)}\,dt + \text{Const.} \qquad (x \geqslant B)\;. \tag{2.11}$$

This gives a discontinuity for the calculated hole current density at the base point B, which, in physical terms, means the lateral supply of base current.

With the new values derived for $\varphi_p(x)$ and $\varphi_n(x)$ a new iteration can be started by going back to (1). When sufficient accuracy has been reached, the calculation is stopped and results are printed out.

The recombination term (R) and the mobilities ($\gamma_p^{-1}$, $\gamma_n^{-1}$) are calculated at an appropriate stage of the iteration cycle. It is clear that adding these nonlinearities generally will degrade the overall convergence rate, particularly when they strongly influence the potential distributions. Mock [30] has shown by a perturbation method that convergence failures in this method may be expected when the carrier diffusion lengths become small compared to the device dimensions. In general, experience with bipolar transistors has shown that the convergence rate of the above described "Gummel method" can become unpractically low for:

- power device structures of several hundreds of microns thickness,
- high bias conditions, when the electric field is relocated because of the high current densities (see fig. 2.1).

Other difficulties with this method were that terms like exp ($\pm\,\Psi$) and exp ($\pm\,\varphi$) can easily cause overflow of computer capacity (at high reverse voltages $\Psi$ and $\varphi$ can have normalized values of 200 or more). Differences of nearly equal numbers in the r.h.s. of eqs.(2.8), (2.10) and (2.11) can also cause severe difficulties. Notwithstanding the difficulties this solution method has successfully been used for analysing the Kirk effect [31] in the base collector region of bipolar transistors. The electron current flow in an NPN transistor influences the charge distribution and therefore the electric field in the collector base junction. This phenomenon will be illustrated with a simplified example.

For a lightly doped collector region with impurity concentration $N_c$, the width of the space charge layer ($x_d$) is not only a function of the applied voltage $V_{CB}$, but also of the electron current $J_n$ (= $J_c$) flowing through that junction. Integration of Poisson's equation and assuming that the electrons move with the saturated drift velocity in the space charge region, it follows that:

$$x_d \doteq \sqrt{\frac{2\epsilon}{q N_c} \; \frac{V_{CB} + V_0 - \rho\, w\, J_C}{(1 - J_C/J_1)}} \qquad (2.12)$$

where $\rho\, w\, J_C$ is the voltage drop in the collector and $V_0$ the built-in voltage of the junction. The critical current density $J_1$ is given by

$$J_1 = q\, N_c\, v_m \; . \qquad (2.13)$$

From eq. (2.12) it follows that $J_C$ has two effects: first the electrons, necessary to carry the current, partly compensate the fixed ionized impurities (this is given by the term $(1 - J_C/J_1)$) which tends to increase $x_d$. And secondly, the ohmic voltage drop in the collector region tends to decrease the depletion region width. At low reverse voltages the depletion layer disappears when

$$J_C > q\, \mu_n\, N_c\, (V_{CB} + V_0)\, /\, w \; . \qquad (2.14)$$

Then the transistor is said to operate in the "injection mode" or "quasi-saturation" because extra electrons ($n > N_c$) are needed in some part of the collector epitaxial layer in order to carry this current flow. At the same time extra holes are injected from the base and will neutralize these electrons. This extra hole charge storage strongly influences the electrical transistor characteristics. The boundary between base and collector is no longer well defined and the nonlinear dependence of the drift velocity on the electric field makes numerical calculations necessary for a detailed investigation. The results of these calculations have been used extensively for developing approximate analytical models, as is clearly described in the work of de Graaff [32].

### *2.1.2. Two-dimensional transistor analysis*

#### 2.1.2.1. Introduction

In order to describe lateral effects such as inhomogeneous current density distribution (current crowding), injection along the emitter-base side wall junction and lateral charge storage (the "2D Kirk effect") it was necessary to solve the basic device equations in two dimensions. This is a difficult problem because it is not possible to integrate the current density equation as is done in the 1D Gummel method. Particularly the difference approximation for the current density equation has to be chosen carefully in order to prevent numerical instability.

Two papers in which the author has presented a solution of this problem will be discussed.

2.1.2.2. "Iterative scheme for one- and two-dimensional d.c. transistor simulation" (see 5.1).

The problem of the two-dimensional transistor structure is formulated as a set of three coupled partial differential equations (see 5.1 eqs. 10, 11, 12), in the variables $\Psi$, $\Phi_p$ and $\Phi_n$. The last two variables are exponentials of the quasi-Fermi potentials:

$$\Phi_p = \exp(\varphi_p)$$

$$\Phi_n = \exp(-\varphi_n) \quad .$$

As in the 1D Gummel method, Poisson's equation is linearized by considering a small perturbation $\delta$ (x, y) on the electric potential $\Psi$ (x, y). The resulting linearized equation in $\delta$ (x, y) and the two continuity equations are linear elliptic differential equations of the following form:

$$\frac{\partial}{\partial x}\left(a\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(c\frac{\partial u}{\partial y}\right) + f\,u = g(x,y) \qquad (2.15)$$

where $a > 0$, $c > 0$ and $f \leqslant 0$. This "self adjoint" form is very attractive for numerical solution because for differential equations of this type standard theory on numerical analysis gives theorems about the existence of a unique solution and convergence properties of iterative solution methods [33].

A non-uniform rectangular mesh is chosen covering the two-dimensional structure. Varga [33] shows that by means of integrating eq. (2.15) over a small box around each mesh point (i, j), difference equations can be derived, given by

$$\mathbf{A}\,\mathbf{u} = \mathbf{b} \qquad (2.16)$$

where **A** is a symmetrical coefficient matrix with the following properties:

$a_{ii} > 0$, $a_{ij} < 0$ ;

$a_{ii} \geqslant \sum_{j \neq i} |a_{ij}|$ with strict inequality for some i (strictly diagonally dominant);

A is "irreducible", which means that eq. (2.16) cannot be reduced into two or more isolated problems.

The matrix A with these properties is positive definite.
Because A is a large and sparse matrix with many elements equal to zero, it is

preferred for reasons of storage and operation time on the computer to solve eq. (2.16) by iterative solution techniques instead of by direct solution methods. It can be shown [33, 34] that the method of successive over-relaxation for a matrix with the above described properties is convergent and the convergence rate can be optimized by choosing an optimal relaxation parameter.

Apart from the difference in the method of solving each of the three differential equations, the overall iteration scheme is the same as the one-dimensional Gummel method.

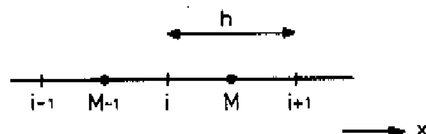



Fig. 2.2.

In this paper (see 5.1) the current density at a point M midway between two mesh points i and i + 1 (see fig. 2.2) is approximated by

$$J_M = -\gamma_M^{-1} \; e^{-\Psi_M} \; \frac{\Phi_{i+1} - \Phi_i}{h} \quad . \tag{2.17}$$

As will be shown, this approximation avoids numerical instability, but it is not very accurate when there are large differences in potentials between two neighbouring points. Because of the importance of the discretization of the current density equation, this will be discussed in more detail.

Consider the one-dimensional continuity equation for holes:

$$\frac{dJ_p}{dx} = -R \tag{2.18}$$

where

$$J_p = -\gamma_p^{-1} \; e^{-\Psi} \; \frac{d\Phi_p}{dx} \tag{2.19}$$

or when we use p instead of $\Phi_p$ as variable

$$J_p = \gamma_p^{-1} \; (pE - \frac{dp}{dx}) \; . \tag{2.20}$$

It is assumed that $\gamma_p$ and $\Psi$ are known functions of position.
Integrating eqs. (2.19) and (2.20) gives :

$$\Phi_p(x_i) = \Phi_p(x) + \int_{x_i}^{x} \gamma_p(s)\, e^{\Psi(s)}\, J_p(s)\, ds \tag{2.21}$$

and

$$p(x_i) = \left[\exp\left\{-\int_{x_i}^{x} E(t)\, dt\right\}\right] p(x) + \int_{x_i}^{x} \gamma_p(s)\, J_p(s) \exp\left\{-\int_{x_i}^{x} E(t)\, dt\right\} ds. \tag{2.22}$$

From eq. (2.22) it is clear that the hole concentration at any point $x_i$ is coupled via positive coefficients with the hole concentration at any other point x. The same is true for the variable $\Phi_p$. Notice that in eq. (2.22) the coefficients are unequal, whereas in eq. (2.21) they are equal. The formulation in the variable $\Phi_p$ will appear to lead to a symmetric matrix for the difference equations and the formulation in p will not. We shall discuss several possible difference approximations for the current density $J_M$ at point M.

(1) Conventional difference approximation:

$$J_M = \gamma_M^{-1} \left\{\frac{1}{2}\,(p_i E_i + p_{i+1}\; E_{i+1}) - \frac{p_{i+1} - p_i}{h}\right\} \tag{2.23}$$

or

$$p_{i+1} = \left(\frac{1 + \frac{1}{2} E_i h}{1 - \frac{1}{2} E_{i+1} h}\right) p_i - \frac{J_M\, \gamma_M}{1 - \frac{1}{2} E_{i+1} h} \tag{2.24}$$

whenever $| E_{i+1} h | > 2$ (unnormalized this means that $| E_{i+1} h | > 2\ kT / q$) there is no positive coupling. This violates the physical property of the original differential equation (see eq. (2.22) ). When substituting this approximation in eq. (2.24) the resulting matrix A only satisfies the necessary conditions for convergence and stability when at every mesh point $| E_i\, h | < 2$. In practice, failure to satisfy this condition will result in numerical instability, leaving no alternative but to choose locally very small step sizes.
Approximation (2.24) can be derived from eq. (2.22) assuming that $J_p$ = constant ($= J_M$), $\gamma_p$ = const. and $| hE | \ll 1$ in each cell.

(2) The difference approximation given in eq. (2.17).
The coupling between $\Phi_i$ and $\Phi_{i+1}$ is positive, as it has to be, and gives for the continuity equation a positive definite coefficient matrix.
Eq. (2.17) can be derived from eq. (2.21) under the assumptions that in each cell $J_p$ = const. ($= J_M$), $\gamma_p$ = const. and exp ($\Psi$) is a linear function in x (which is equivalent to saying $| hE | \ll 1$). This forces one again to use very small step sizes at

places where $\Psi$ changes considerably (e.g. in the reverse-biased base collector junction). The conclusion is that although this difference approximation will always be stable and convergent, the validity range is still small.

(3) The difference approximation suggested by Scharfetter and Gummel [35]. In their one-dimensional large-signal calculations on a silicon Read Diode Oscillator Scharfetter and Gummel introduced a new difference approximation for the current density equation which was numerically stable under all conditions. It is assumed that in each cell $J_p$ = constant (= $J_M$), $\gamma_p$ = const. (= $\gamma_M$) and E = constant (= $E_M$). From eq. (2.22) it follows then that

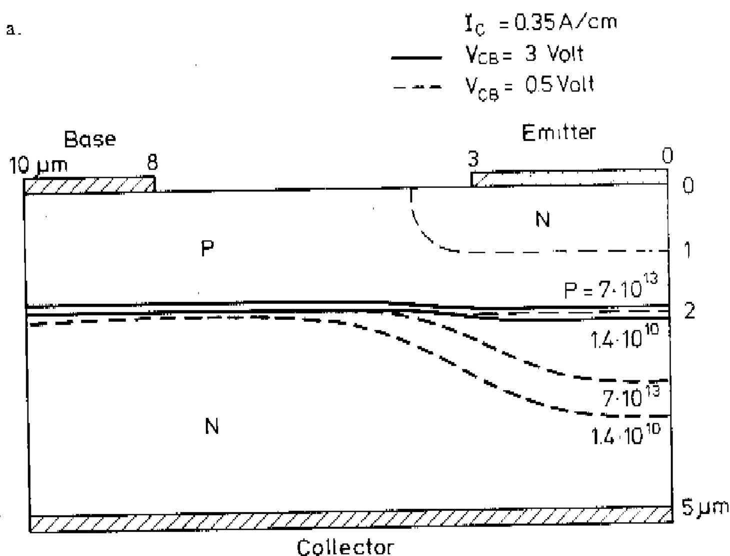
$$J_M = \gamma_M^{-1} E_M \frac{p_{i+1} \exp(-hE_M) - p_i}{\exp(-hE_M) - 1} \tag{2.25}$$

The positive coupling between $p_i$ and $p_{i+1}$ is always present. Because the assumption of E = const. is much less restrictive than $|hE| \ll 1$, the validity range of this approximation is much larger. Unfortunately substitution of eq. (2.25) in the continuity equation results in a matrix A that is not symmetric and not diagonally dominant. However starting with eq. (2.21) and using the same assumptions does give a symmetric relation and a definite positive coefficient matrix A for the continuity equation. It can be shown (see 5.2) that this matrix is similar to the one that would follow from eq. (2.25) and thus it can be said that the difference approximation of eq. (2.25) is the most practical one from a numerical as well as a physical point of view.

#### 2.1.2.3. "Computer-aided two-dimensional analysis of bipolar transistors" (see 5.2).

The method for solving numerically the two-dimensional semiconductor steady-state transport equations for bipolar transistors is described in this paper in detail and is a continuation of the paper discussed above. The Scharfetter-Gummel difference approximation for the current density (see above (3) ) is shown to be a remarkable improvement for the two-dimensional analysis as well. In addition two rectangular meshes are used, one fine mesh for Poisson's equation adapted to follow the rapid potential changes, and another coarser mesh for discretization of the continuity equations. In the last case the step sizes may be larger because the current densities do not vary so rapidly.
In fig. 9 of this paper the overall convergence rate is given as a function of the forward bias conditions. In much the same way as in the one-dimensional Gummel method, the convergence rate becomes worse at higher currents. A number of computer plots illustrate the internal and the external behaviour of an NPN transistor. Electron current flow lines as a function of biasing are shown in fig. 11 and fig. 13 (5.2). Current crowding along the emitter-base junction takes place due to a voltage drop in the base region underneath the emitter. When the base-collector junction is swamped with electrons and holes, the electron current starts

a.




b.

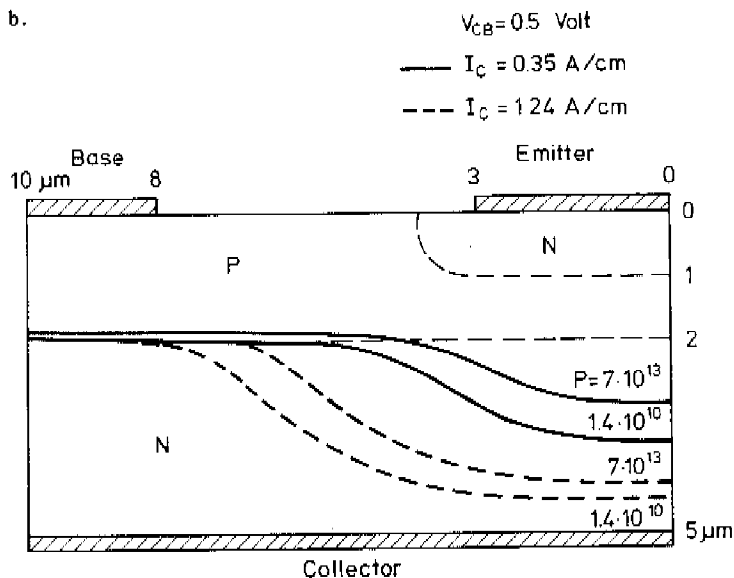



Fig. 2.3. Lines of constant hole concentrations. In (a) it is shown that when $V_{CB}$ is increased the transistor comes out of saturation and in (b) it is seen that many holes are stored laterally, which increase with current density.

spreading directly behind the emitter junction. This introduces a strong lateral change in the carrier concentrations (see 5.2, fig. 17) and the electric potential (see 5.2, fig. 15). Fig 16 (5.2) shows that a lot of hole charge also is stored laterally, out of the "active" transistor region. Fig. 2.3 shows how the collector current and the applied collector base voltage influence this stored hole charge.
In fig. 18 (5.2) a comparison is made for the $f_T$ characteristic between the two-dimensional structure and the corresponding one-dimensional transistor. In the latter case the $f_T$ falls off earlier and more steeply because of the larger collector series resistance. The one-dimensional calculated $f_T$ values can easily be corrected for this effect by taking the voltage drop in the one-dimensional structure ($J_C \rho w$) equal to that in the two-dimensional structure ($I_C R_S$) [36] :

$$J_C \rho w = I_C R_S \tag{2.26}$$

where $R_S$ is the collector spreading resistance [37]. In this way the crosses (x) in fig. 2.4 are obtained. For very high current densities the predicted curve deviates appreciably because the current spreading is then a function of current and voltage and $R_S$ will change too.

At the beginning of the paper reprinted in 5.2 a short, incomplete review of the literature on numerical transistor calculations is given. More recent work in the

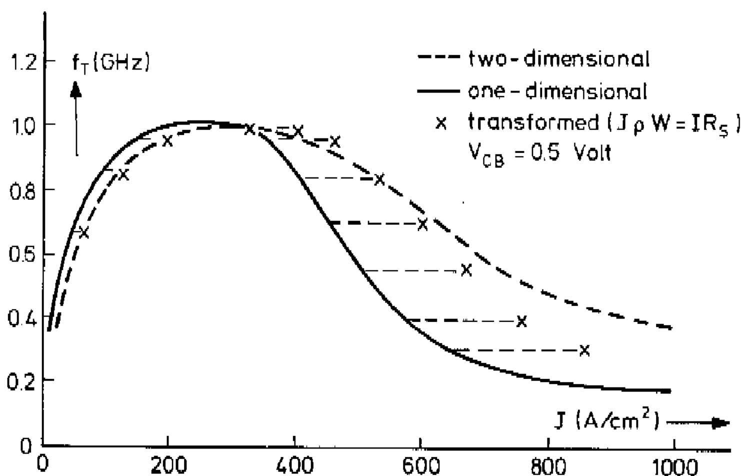



Fig. 2.4. The one-dimensionally calculated $f_T$ characteristic is corrected for collector spreading resistance (see the crosses).

field of bipolar transistors has been published by Heimeier [38], Manck and Engl [39], Gaur and Navon [40], Jesshope, Zaluska and Kemhadjian [41], Mock [42], Hachtel and Mack [43] and many others. For a more complete review and discussion of each of these publications is referred to [44, 45]. An interesting development is the use of the finite-element method [46].

## 2.2. Calculations including high doping effects

### *2.2.1. Discussion of TRAP1*

Recently we have developed an efficient one-dimensional TRansistor Analysis Program (TRAP1) which incorporates a number of considerable improvements, both physical and numerical, compared to the original Gummel method. The influence of the impurity concentration on the bandgap and on the minority carrier lifetime has been taken into account in accordance with the experimental results for these phenomena (see section 1.5 and Chapter 3).
But the numerical solution procedure, too, has become much more efficient because of a simplifying assumption first reported by Ruch and Scharfetter [47], which is an attractive compromise between accuracy and computational effort. It is assumed that the hole quasi-Fermi potential is constant throughout the transistor (see fig. 2.1):

$$\varphi_p(x) = 0 \tag{2.27}$$

This means that internal recombination is neglected and only two of the three differential equations are left:

$$\frac{d^2\Psi}{dx^2} = n_i\, e^{\Psi - \varphi_n} - n_i\, e^{-\Psi} - N \tag{2.28}$$

$$J_n = -\gamma_n^{-1}(E, N, T)\; n_i\, e^{\Psi - \varphi_n} = \text{const.} \tag{2.29}$$

This assumption is justified because in "useful" bipolar transistors the recombination currents are roughly an order of magnitude smaller than the main current ($J_n$). Or, in other words, the base current is usually only a small fraction of the collector current (e.g. $h_{FE} > 20$). Once this problem has been solved the recombination currents are calculated from a perturbation of this solution, the internal recombination this time being fully taken into account. The resulting function $\varphi_p(x)$ deviates only in the heavily doped n-regions from the a priori assumption (2.27) and its effect on $\Psi(x)$ and $J_n$ can be neglected (see fig. 2.1). This has been verified with the Gummel method and deviations were found to be less than a few percent. Eqs. (2.28) and (2.29) are linearized by considering small perturbations $\delta(x)$ and $\epsilon(x)$ on $\Psi(x)$ and $\varphi_n(x)$, according to the Newton method.

This results in the following difference equations:

$$\begin{aligned} a1_i\,\delta_{i-1} + a2_i\,\delta_i + a3_i\,\delta_{i+1} + \quad & + a4_i\,\epsilon_i &= d_i \\ b1_i\,\delta_{i-1} + b2_i\,\delta_i + b3_i\,\delta_{i+1} + \quad & + b4_i\,\epsilon_{i-1} + b5_i\,\epsilon_i + b6_i\,\epsilon_{i+1} &= e_i \end{aligned} \tag{2.30}$$

$(i = 1, 2, 3, \ldots jm-1)$

Use has been made of the Scharfetter-Gummel difference approximation for the current density equation. The coefficients in eq. (2.30) are formulated in such a way that only the carrier concentrations and potential differences between two neighbouring mesh points are involved. This prevents overflow of computer capacity. The complete set of difference equations is solved simultaneously by means of Gauss elimination.

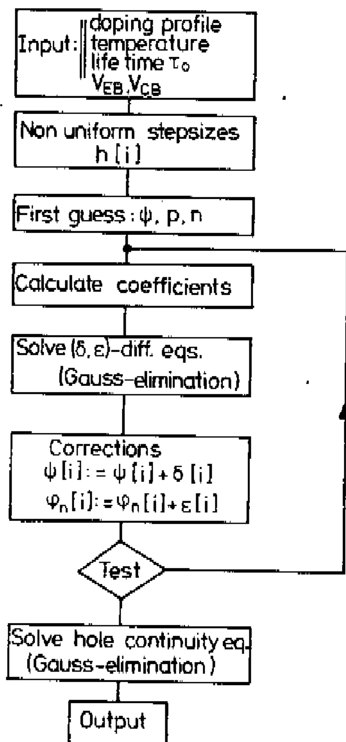



Fig. 2.5. Flow diagram of TRAP1.

A simplified flow diagram of the solution procedure of TRAP1 is shown in fig. 2.5. The required input data are the doping profile, the transistor temperature, the Hall-Shockley-Read lifetime $\tau_0$ (see 1.5) and the applied voltages.
Depending on the doping profile a non-uniform step-size distribution is automatically chosen. For this purpose the total number of mesh points (usually 100 to 200) and the minimum and maximum allowed step-size must be specified, after which the step-size distribution is chosen in a rather pragmatic way with small steps where the profile is steep (junctions) and larger steps in regions where the profile did not change strongly. The changes in the step-sizes must be smooth (see fig. 2.1).
As a first guess for $\Psi$ (x), p (x) and n(x), results from a previous bias condition can be used or directly generated from simplified physical models, e.g. complete depletion, charge neutrality, constant quasi-Fermi potentials, etc.
The calculation of the coefficients contains many exponentials, which for small arguments can be expanded in order to save computing time.
For the solution of the linearized set of difference equations (2.30) a procedure based on Gauss elimination has been written.
Usually three iterations are sufficient to solve the non-linear problem of eqs. (2.28) and (2.29). This is illustrated in fig. 2.6 in a comparison with the Gummel method. The computing time is more than ten times shorter.

The continuity equation for the holes is discretized using the difference equation (2.25) and also solved directly. The two recombination mechanisms discussed in section 1.5 and a finite surface recombination velocity s at the emitter contact are taken into account.

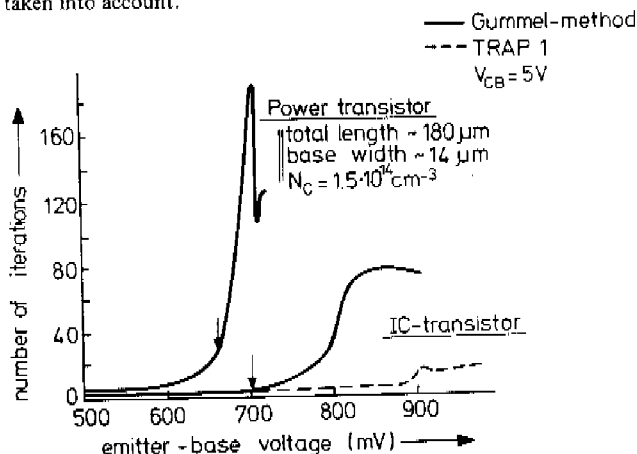



Fig. 2.6. The number of iterations needed to solve the transistor problem for a given biasing as found with the original Gummel method and with the method used in TRAP1. (The arrows correspond to the maximum in current gain).

The TRAP1 program is equipped with a number of extra features, such as PNP or NPN structures, with or without high doping effects, avalanche effect, perturbation of the dc solution (for capacitance and $f_T$ calculations), emitter contact resistance, crowding, etc. In general it is very simple to investigate the effect of one particular parameter without changing the others. The program has been used for analysing a large number of transistor structures, including IC transistors, high-frequency transistors, power transistors and transistors with epitaxial layers in the emitter region ($I^2L$, LEC). Several of the possibilities and results will be shown in the next section. It is noted that no fitting has been used in these calculations apart from the parameter values already discussed (chapter 1).

### 2.2.2. Example: IC transistor

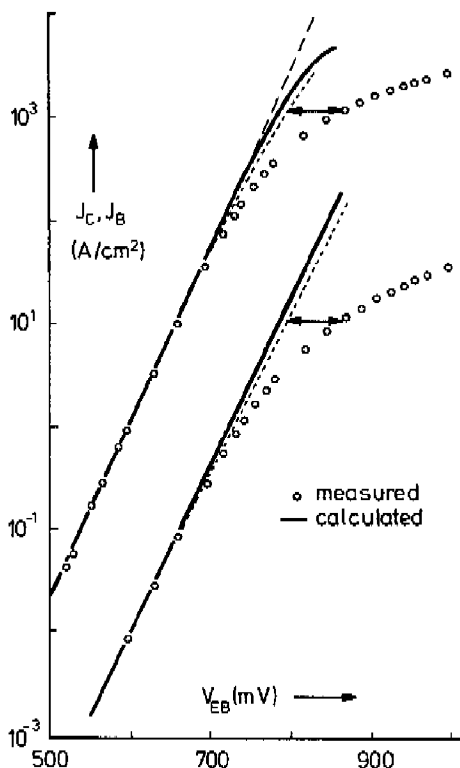



Fig. 2.7. Measured and calculated $J_C$, $J_B$ ($V_{EB}$) characteristics.

The electrical characteristics of an integrated circuit transistor have been calculated with TRAP1 and compared with measurements. The doping profile of this transistor has been modelled as discussed in section 1.3 (see fig. 1.2).

In fig. 2.7 the $J_C$-$V_{EB}$ and $J_B$-$V_{EB}$ characteristics are shown. At high current densities pulse measurements [48] were made in order to prevent heating of the transistor. The dotted lines are derived from the measured values by correcting for the resistive voltage drop in the emitter and base region.

The current gain $h_{FE}$ (= $I_C$ / $I_B$) as a function of collector current is given in fig. 2.8. It appears that a lifetime of 1 μsec. in the Hall, Shockley and Read recombination mechanism fits the measured values well. The effect of changing this lifetime is demonstrated in fig. 2.9. Reducing it means an increase of recombination in the emitter-base junction and makes the non-ideal base current component larger; it therefore causes a stronger fall-off in $h_{FE}$ at low currents. At about 700 mV the current gain has its maximum value; at higher currents it falls off owing to extra hole charge storage. The current gain usually increases with temperature (see fig. 2.10) due to two counteracting phenomena, as will be discussed in 4.2. At high currents the temperature coefficient changes sign because of an increasing voltage drop in the ohmic collector region.

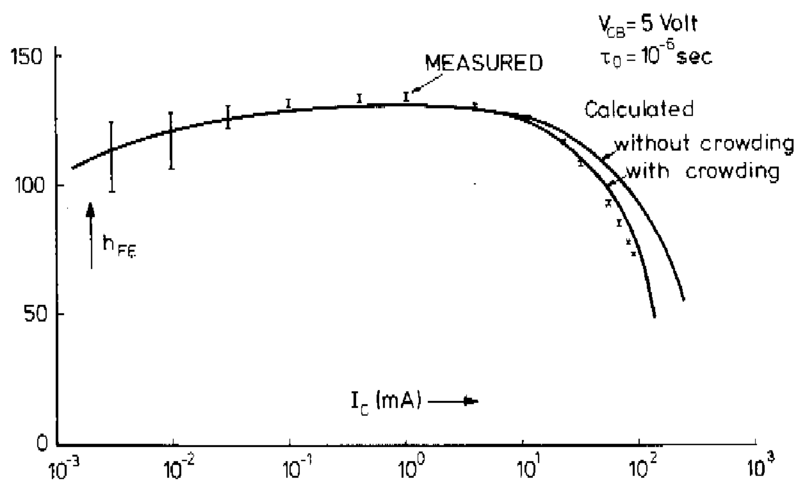



Fig. 2.8. Measured and calculated current gain as a function of collector current. The influence of crowding is also shown.

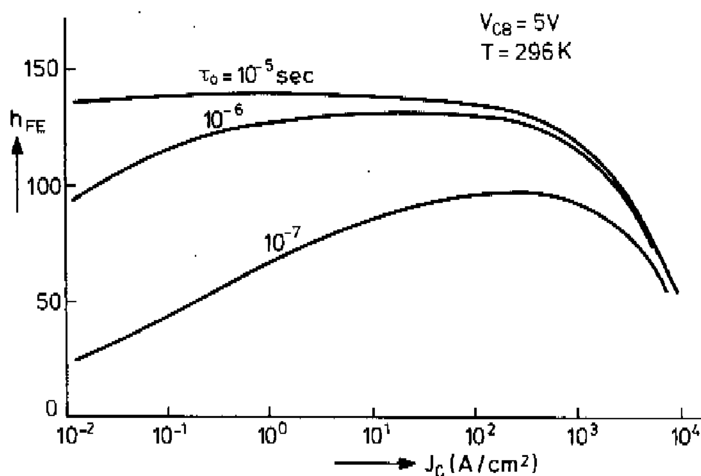



Fig. 2.9. Influence of H.S.R. lifetime $\tau_0$ on the calculated current gain.

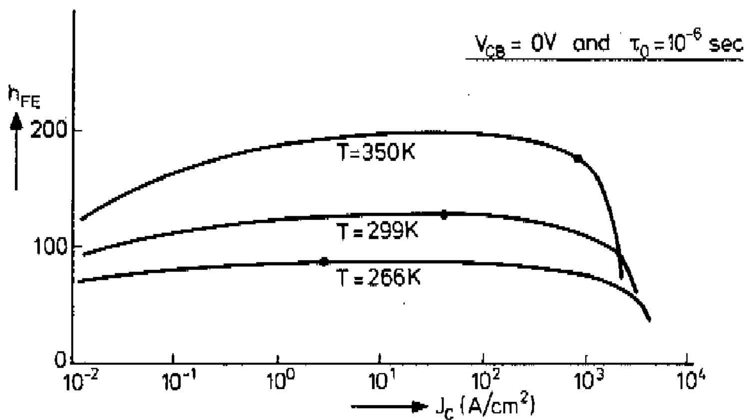



Fig. 2.10. Influence of temperature on the calculated current gain (the dots correspond to $V_{EB}$ = 700 mV).

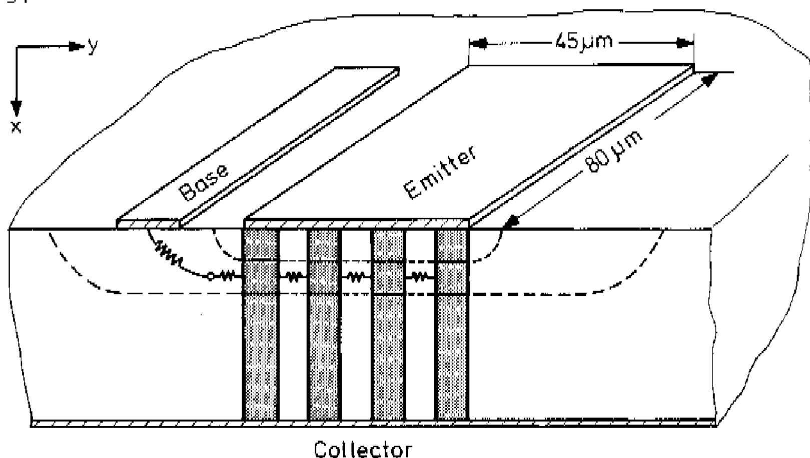



Fig. 2.11. Cross section of bipolar transistor. The coupled one-dimensional transistor sections are schematically indicated.

Current crowding caused by lateral voltage drop in the base region has already been mentioned in the discussion of the two-dimensional analysis (2.1.2.3). It can be approximated by putting a number of one-dimensional transistors in parallel, coupled through bias dependent base resistances (fig. 2.11). These are calculated from the base sheet resistance

$$R_{\square EB}^{-1} = q \int_0^L \mu_p \, p \, dx \tag{2.31}$$

valid for the local one-dimensional transistor under consideration. The influence of this crowding on the current gain characteristic is shown in fig. 2.8 for an emitter width of 45 μm. At the high current densities in the corner of the emitter-base junction the voltage drop is much less than would be predicted from a model with constant current gain and sheet resistance [49]. This is also illustrated in the internal lateral current distribution (fig. 2.12).

In fig. 2.13 calculated and measured $I_C$-$V_{CE}$ characteristics are shown. Like the recombination currents the avalanche effect is included in the program as a first-order approximation. This seems a reasonable approximation as long as the current ($J_G$) generated in the avalanche process is small compared with the electron current flow ($J_n$). This avalanche-generated current density is calculated from

$$J_G = \alpha J_C \int_B^L \exp(-\beta / |E|) \, dx \tag{2.32}$$

where $\alpha = 2.4 \cdot 10^6$ cm$^{-1}$ and $\beta = 1.6 \; 10^6$ V/cm.

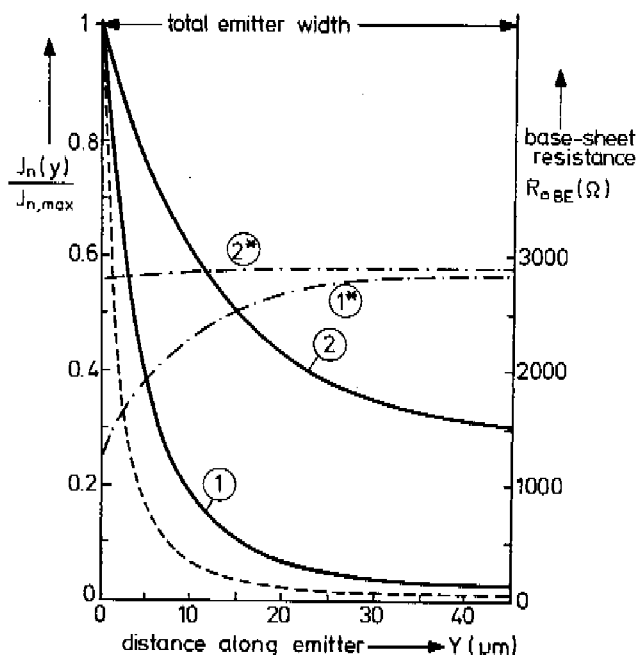



Fig. 2.12. Current crowding along the emitter-base junction.

1. $J_n(y) / J_{n,max}$ for $I_C = 52$ mA and $J_{n,max} = 10^4$ A/cm² at y = 0. The base-sheet resistance ($R_{\square EB}$) is also shown as a function of position along the emitter junction (the dashed line 1*). The dotted line gives the corresponding current crowding when it is assumed that $R_{\square EB}$ and the current gain are constant along the junction ($I_C = 52$ mA, $J_{n,max} = 2 \cdot 10^4$ A/cm²).
2. $J_n(y) / J_{n,max}$ for $I_C = 6$ mA, $J_{n,max} = 350$ A/cm² and $R_{\square EB}$ is given by the dashed line 2*.

Particularly in the case of thin collector epilayers an interesting phenomenon [50] can occur at high current levels when the space charge in the depleted collector region

$$\rho = q\,(N_c - J_n / q\, v_m) \tag{2.33}$$

changes sign. Then the maximum electric field shifts from the collector-base junction towards the $nn^+$ interface (see fig. 2.14). The avalanche-generated base current flows

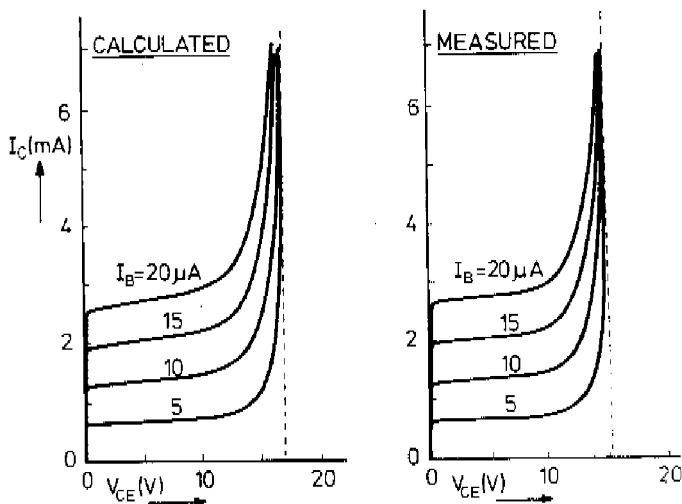



Fig. 2.13. Comparison between measured and calculated $I_C$-$V_{CE}$ characteristics.

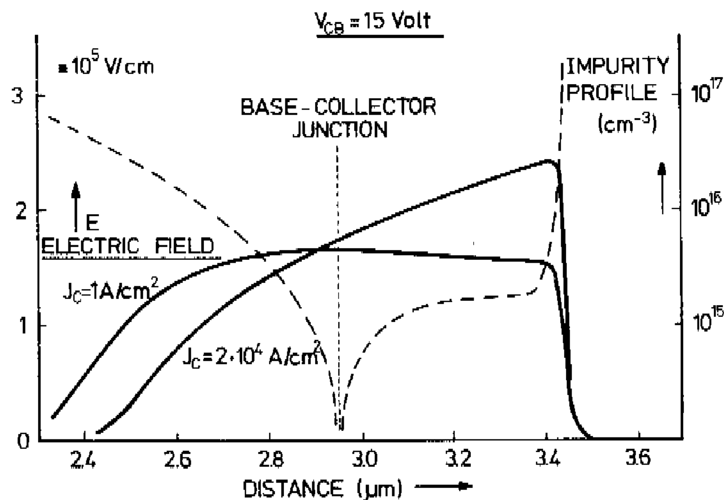



Fig. 2.14. The same doping profile as is shown in fig. 1.2 is used, the only difference being that the collector epilayer is made very thin (about 0.5 μm, see the dashed line). At high current densities the maximum of the electric field is at the $nn^+$ barrier.

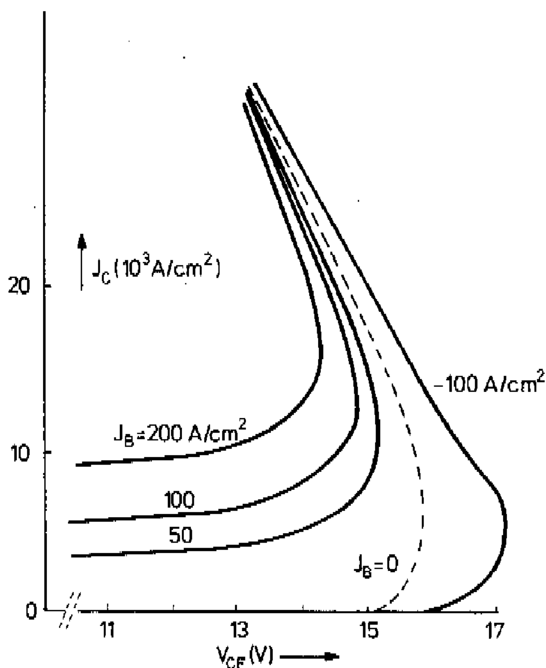



Fig. 2.15. Calculated $J_C$-$V_{CE}$ characteristics showing a negative resistance due to avalanche-generation at the $nn^+$ barrier.

in a direction opposite to the normal base current and tends to decrease the total base current or may even invert its direction. When the total base current is kept constant the collector current increases and also $|E|_{max}$. This generates even more avalanche current and leads to an unstable situation which finally (probably together with temperature effects) causes second-breakdown of the transistor. This is illustrated in the calculated $I_C$-$V_{CE}$ characteristic shown in fig. 2.15.

Usually it is assumed that the maximum of the recombination takes place at the emitter-base junction where p = n, because in this case the H.S.R. recombination term has a maximum. However, when Auger recombination together with dope-dependent bandgap narrowing and lifetime are taken into account, another peak of the recombination takes place deeper in the neutral emitter region (see fig. 2.16 and 5.6). This recombination peak is exactly proportional to exp $(q\, V_{EB} / kT)$ while the space charge recombination, controlled by the H.S.R. lifetime $\tau_0$, is proportional

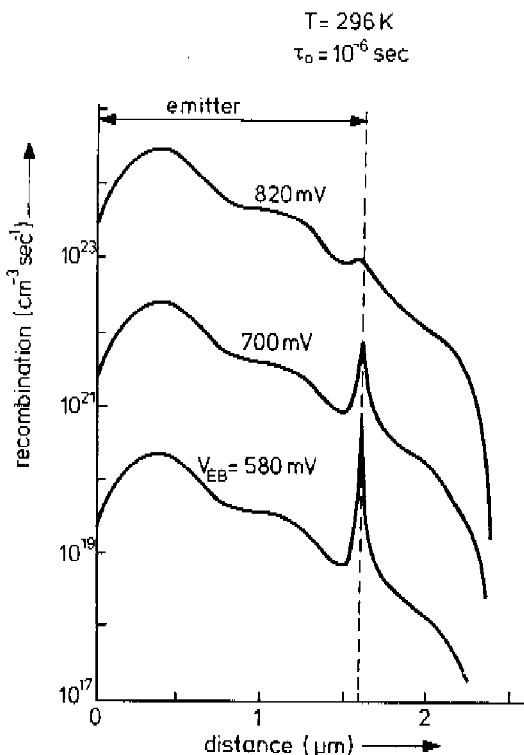



Fig. 2.16. Recombination rate in the emitter-base region for three emitter-base voltages. The recombination peak in the space-charge layer is due to recombination via traps ($\tau_0$) and the other peak deeper in the neutral emitter is the result of bandgap narrowing and Auger recombination.

to exp (q $V_{EB}$ / m kT) with $1 < m < 2$. When the forward bias and temperatures are low, recombination in the space charge region via traps is dominant, but at higher forward voltages and higher temperatures Auger recombination in the neutral bulk is dominant.
Experimental verification of the Auger recombination peak will be discussed in section 4.1.

From a perturbation of the dc solution, approximations for several important ac quantities can be made in accordance with the charge control principle, which assumes that the internal charge distribution follows instantaneously the applied

voltages and controls the currents without time delay (for a critical discussion of the charge control principle see [51]. The total transit time from emitter to collector contact is then given by

$$\tau_{ec} = \frac{1}{2\pi f_T} = \frac{\Delta Q_{tot}}{\Delta J_C}\bigg|_{V_{CE} = \text{const}} \tag{2.34}$$

and the emitter-base capacitance by

$$C_{EB} = \frac{\Delta Q_{EB}}{\Delta V_{EB}} \quad . \tag{2.35}$$

The total stored charge $\Delta Q_{tot}$ and the charge $\Delta Q_{EB}$ stored in the emitter region are defined by

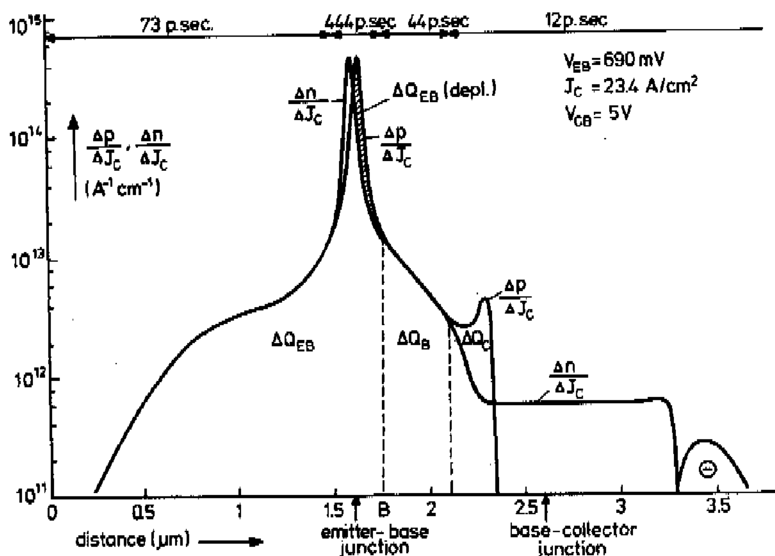



**Fig. 2.17. Perturbations of the hole and electron carrier concentrations as a function of position in the transistor when a small perturbation $\Delta J_C$ of the given bias condition is applied.**

$$\Delta Q_{tot} = q \int_0^L \Delta p \, dx \tag{2.36}$$

$$\Delta Q_{EB} = q \int_0^B \Delta p \, dx \tag{2.37}$$

where $\Delta p$ (x) is the change in hole concentration caused by a small perturbation in the applied emitter base voltage.

In fig. 2.17 is shown how the electron and hole carrier concentrations change due to a perturbation $\Delta J_C$ of a bias condition given by $J_C = 23.4$ A/cm$^2$ and $V_{CB} = 5$ V. Integration of $q \, \Delta p / \Delta J_C$ over the whole transistor gives the total delay time $\tau_{ec}$ (see eqs. (2.34) and (2.36) ). Because the total transistor is charge-neutral, integration of $q \, \Delta n / \Delta J_C$ should give the same result (in TRAP1 a comparison of both calculations is used as an indication of the numerical accuracy). For low forward bias most of the charge is stored in the emitter-base junction, while at higher current the charge storage in the base and collector regions becomes dominant.

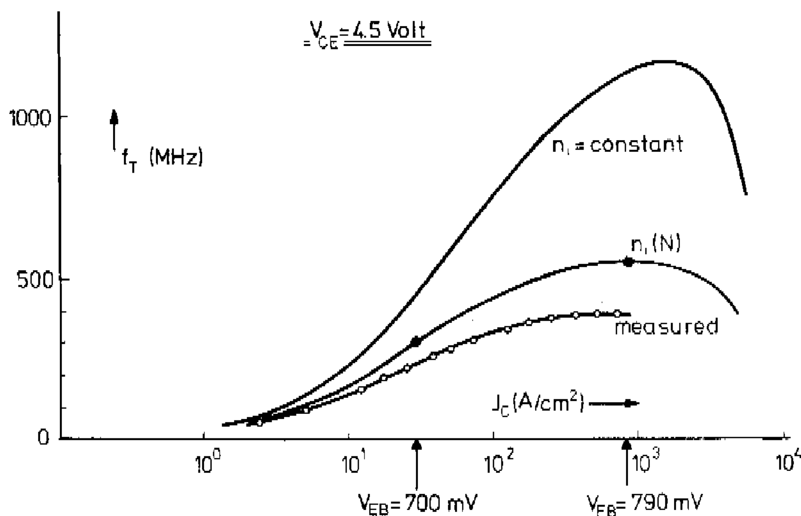



Fig. 2.18. Measured and calculated $f_T$ characteristics (the measurements have been done by B.C. Bouma).

In fig. 2.18 the calculated $f_T$ ($J_C$) characteristic is compared with the measurements and also with a calculation where no bandgap narrowing is involved (no corrections have been made for the emitter side-wall capacitance nor for the base-collector

capacitance outside the active transistor area). Bandgap narrowing increases particularly the charge storage in the region of the emitter-base junction and reduces therefore the $f_T$ values [52].

Fig. 2.19 shows the emitter-base capacitance calculated with the help of eqs. (2.35) and (2.37).

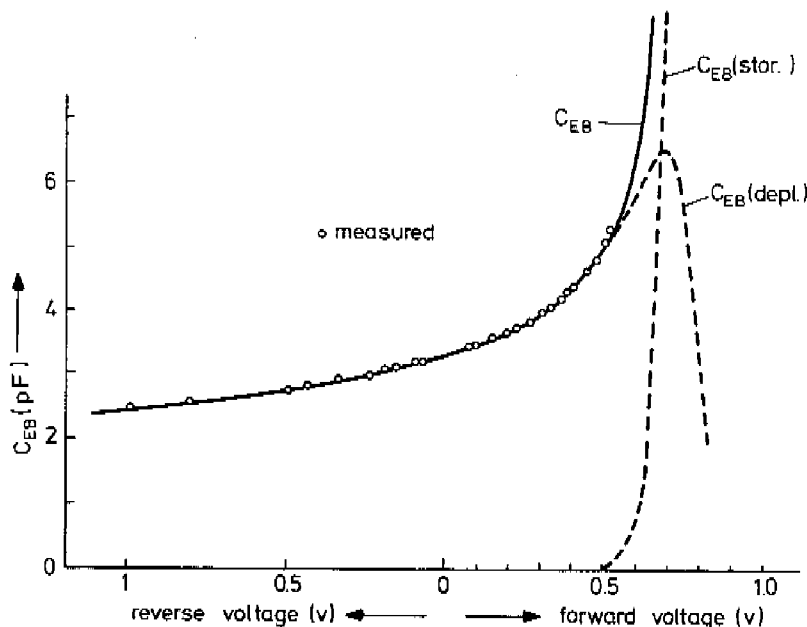



Fig. 2.19. Emitter-base capacitance as a function of voltage.

From the calculations it is easy to separate the total stored charge into different charge packets (see fig. 2.17):

$$\Delta Q_{tot} = \Delta Q_{EB} + \Delta Q_B + \Delta Q_C \,, \tag{2.38}$$

each correlated with a time constant $\tau_E$, $\tau_B$ and $\tau_C$.
The important emitter-base charge storage $\Delta Q_{EB}$ can be further divided into

$$\Delta Q_{EB}\,(\text{depl}) = q \int_0^B |\Delta p - \Delta n| / 2\, dx \tag{2.39}$$

$$\Delta Q_{EB}\,(\text{stor}) = \Delta Q_{EB} - \Delta Q_{EB}\,(\text{depl})\ , \tag{2.40}$$

where $\Delta Q_{EB}$ (depl) is the result of changes in space charge and called usually the "depletion capacitance"- or "transistion region capacitance"-charge, while $\Delta Q_{EB}$ (stor) represents that part of the total stored hole charge that is neutralized by storage of electron charge. $\Delta Q_{EB}$ (stor) is not only present in the space charge neutral emitter but also in the space charge region itself (see fig. 2.17). The two components of the emitter-base capacitance are also shown in fig. 2.17.

# 3. EXPERIMENTS AND CALCULATIONS OF BANDGAP NARROWING IN SILICON

## 3.1. Introduction

*At low impurity concentrations* the impurities do not disturb the periodicity of the crystal and the forbidden energy gap $E_g$ is well-defined by sharp band-edges. The density of states functions for electrons and holes are proportional to $\sqrt{E}$ and independent of the impurity concentration. The pn-product in equilibrium is given by (see chapter 1):

$$p_0 n_0 = n_{io}^2(T) = N_c N_v \exp(-E_g(T) / kT) \; . \tag{3.1}$$

The impurity atoms create a discrete energy level located in the forbidden band at a distance from the nearest band-edge equal to the ionization energy. The wave function of an electron occupying this level is localized in the vicinity of the impurity atom.

*At high impurity concentrations* when the impurity atoms are no longer "isolated" the wave functions of their associated electrons are going to overlap. This results in a splitting of the impurity energy levels (an impurity band is formed) and due to the Coulomb interaction of the large number of electrons the potential energy of these electrons is reduced, resulting in a narrowing of the bandgap.
Because the impurity atoms are not periodically arranged in the crystal but are more likely to be randomly distributed, there will be local variations in impurity concentration on a microscopic scale. This results in potential fluctuations which introduce *tails* on the energy bands, not only the impurity band but also the conduction and valence bands. Theoretical description of these effects is very complicated and will be briefly discussed in 3.3.

A number of experiments are known in the literature which support the above described bandgap changes at high impurity concentrations:

(1) Pearson and Bardeen [53] found in silicon, that the impurity ionization energy decreases with increasing impurity concentration and vanishes at concentrations of about $10^{18}$ cm$^{-3}$ (see fig. 3.1). This is explained as the result of an impurity band which broadens with increasing impurity concentration and finally overlaps with the conduction or valence band resulting in one continuous energy band. Debye and Conwell [54] found the same behaviour in germanium.

(2) Optical absorption measurements on heavily doped germanium [55] and silicon [56, 57] show a reduction in the energy bandgap.

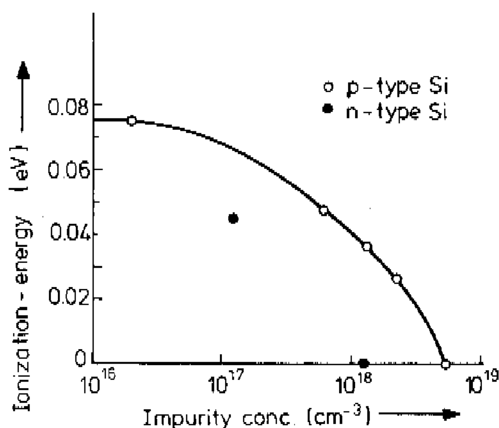



Fig. 3.1. Dependence of ionization energy of impurities upon concentration.

(3) Current gain measurements on bipolar transistors [58, 59]. It was stated that a difference in bandgap in the emitter and base regions could explain the measured positive temperature coefficient of the current gain and also the discrepancy between theoretically predicted and measured values for the current gain.

Other experiments on the recombination-radiation spectra of high concentrations of pn carrier plasmas, excited in pure silicon [60, 61] and low-doped germanium [62], are also explained by bandgap narrowing due to Coulomb interaction. In all these experiments the bandgap narrowing was empirically described by

$$\Delta E_g :: N^{1/3} . \tag{3.2}$$

This can roughly be explained [63] by the Coulomb energy saved by the electron-electron avoidance that results from the Pauli principle and that scales with the inverse of the average inter-electronic distance $r_s$, given by :

$$\frac{4}{3} \pi r_s^3 = \frac{1}{N} . \tag{3.3}$$

The reduction in energy is of the order

$$\Delta E_g \approx C \frac{q^2}{4\pi\epsilon r_s} \approx C\, 2 \cdot 10^{-8} N^{1/3} . \tag{3.4}$$

Keyes [63] shows that our measurements of the bandgap narrowing (see 3.2) are closely proportional to $N^{1/3}$ with C = 2.

Recently several authors (for a short review see 5.3) brought these bandgap changes at high impurity concentration back into discussion and stressed then possible large influence on important device characteristics such as $h_{FE}$ and $f_T$ in bipolar transistors. They calculated the dope-dependent pn-product and derived carrier transport equations incorporating extra terms for the position-dependent bandgap (see chap. 1). They found, however, fairly considerable quantitative differences between the different theoretical calculations and the optical absorption measurements.

Because it was clear that this fundamental phenomenon could be of great importance for the understanding of silicon devices in general, we felt it necessary to try to measure it in the transistor itself. This should also avoid difficulties in interpretation of possible differences between optical and electrical bandgap. These experiments will be described in 3.2 and will be supported by calculations discussed in 3.3. In section 3.4 a discussion of the results will be given.

### 3.2. **"Measurements of bandgap narrowing in Si-bipolar transistors"** (see 5.3)

Because many transistor characteristics depend strongly upon changes in the bandgap we tried to derive the influence of the impurity concentration on the bandgap from electrical measurements of a number of transistor structures.
The emitter is the most heavily doped region in bipolar transistors and therefore it seems obvious to investigate the $I_B$-$V_{EB}$ characteristic because this is primarily controlled by the properties of the emitter region. However, the base current is not only determined by recombination in the neutral emitter but also by recombination in the space charge region, at the surface or in the base and collector regions. Where most of the recombination takes place depends upon technology (lifetimes in the bulk, at the interfaces, in the space charge region, etc.) and further upon the bias conditions (applied voltages, temperature). Because the electrical and temperature behaviour of these base current components is different, the $I_B$-$V_{EB}$ characteristic is not a very attractive means of investigating the bandgap narrowing. It seems better to consider the $I_C$-$V_{EB}$ characteristic, which is one of the best defined transistor characteristics and is not affected by surface effects or recombination phenomena. This characteristic usually shows a perfect exponential relationship over many decades of current, and the non-ideality factor m deviates less than about 2‰ from the exact value 1.

$$I_C = A_E \frac{kT\,\mu_n\,n_i^2}{Q_B} \exp(q\,V_{EB}/kT) \qquad (3.5)$$

where $n_i^2$ is the pn-product in the base and $Q_B$ the total base charge.

The simplest method to determine $n_i^2$ is first to measure at a given temperature (e.g. T = 300K) the $I_C$-$V_{EB}$ characteristic for low injection and then to determine $Q_B$ from a measurement of the base sheet resistance underneath the emitter ($R_{\square EB}$) on a sample taken from the same slice as the transistor. The value of $n_i^2$ follows directly (see also 5.3 eq. (8) ). For the mobility in the base region mean values have to be used.

It is important that the temperature is measured and kept constant during the measurement. The next example gives an impression of how large the increase in $n_i^2$ can be.

For an epitaxial base transistor with $\rho_{epi} = 0.005\ \Omega cm$, which corresponds to an impurity concentration of $2.3 \cdot 10^{19}\ cm^{-3}$, the ratio measured is

$$\left(\frac{n_i}{n_{io}}\right)^2 = 70 \qquad (T = 303.1\ K) \tag{3.6}$$

which means a bandgap narrowing (see 5.3 eq. 7) of $\Delta V_{go} = 112$ mV.

A narrowing of the bandgap not only increases the value of $n_i^2$ but also changes its temperature dependence. This suggests that $\Delta V_{go}$ could be derived equally well from temperature measurements of the $I_C$-$V_{EB}$ characteristic. However, this is a much more complicated measurement than the method described above. The reason is that the change in temperature dependence is relatively small (compare $\Delta V_{go}$ with the total bandgap $E_g$) and moreover it is complicated by the temperature dependence of the mobility $\mu_n$ in eq. (3.5). The following procedure seems obvious.

Substitute the expression (3) of 5.3 for $n_i^2$ into eq. (3.5) and assume that $V_g(T) = V_{go} - \alpha T$ and $\mu_n :: T^{-\eta}$ :

$$I_C = C\, T^{\beta} \exp\left\{q\,(V_{EB} - V_{go})/kT\right\} \ . \tag{3.7}$$

This can be written as

$$V_{EB} = V_{go} - \frac{kT}{q}\ln\left(\frac{C}{I_C}\right) - \beta\frac{kT}{q}\ln(T)\ . \tag{3.8}$$

The emitter-base voltage is measured as a function of temperature with $I_C$ kept constant. The three parameters C, $\beta$ and $V_{go}$ can be derived in principle from a least square fit of these measurements using eq. (3.8). However, we did not succeed to derive a unique set of parameter values in this way. Moreover, the values changed appreciably upon a change in the temperature interval in which the measurements were made. This difficulty is due to the fact that the assumptions made for the temperature dependence of the mobility and of the bandgap are only valid approximately in a limited temperature interval which is difficult to predict in advance. Additional independent information is necessary in order to eliminate the approximation for the temperature dependence of the mobility. This can be derived from measurements of $R_{\square EB}$ assuming that the electron and hole mobilities in the base region have the same temperature dependence:

$$\mu_n(T) = \frac{\mu_n(T = 300)}{\mu_p(T = 300)}\ \mu_p(T)\ . \tag{3.9}$$

Now eq. (3.5) reduces to

$$I_C = A_E \frac{kT}{q\,Q_B^2} \frac{b}{R_{\square EB}}\, n_i^2 \; \exp\left(q\,V_{EB} / kT\right) \tag{3.10}$$

(b is the ratio of the electron and hole mobility at T = 300 K in the base).
The values of $n_i^2$ (T) follow directly from eq. (3.10) using the measured $V_{EB}$ values with $I_C$ is constant and the measurements of $R_{\square EB}$ (T). However, a logarithmic plot of $n_i^2$ as a function of $T^{-1}$ does not clearly show the influence of the impurity concentration on the temperature dependence. The change in temperature dependence is more clearly shown after differentiation of eq. (3.10) :

$$\left(V_g - T\frac{d\,V_g}{dT}\right) = \left(V_{EB} - T\frac{d\,V_{EB}}{dT}\right) - \frac{kT}{q}\left(4 - T\frac{d}{dT}\ln R_{\square EB}\right) , \tag{3.11}$$

where $n_i^2$ has been chosen proportional to $T^3 \exp(-q\,V_g / kT)$.

Two examples are given in fig. 4 of 5.3. for a transistor with a low base-dope concentration ($N = 4 \cdot 10^{16}$ $cm^{-3}$) showing no bandgap narrowing ($V_{go} \approx 1.200$ eV) and in fig. 6 of 5.3. for a heavily doped transistor ($N_B = 10^{19}$ $cm^{-3}$) showing a large narrowing of about 70 mV.

An overall survey of the results of both measuring methods is given in fig. 7 of 5.3. Because the latter method is difficult and requires a high degree of accuracy of the $V_{EB}$ and temperature values, the measurement set-up will briefly be discussed.

The transistor is placed above the surface of liquid nitrogen and can be heated by a small oven, built around the transistor holder. A feedback system keeps the adjusted temperature stable within 0.1 K. The temperature of the transistor is measured with a gauged chromel-alumel thermocouple with the other junction in melting ice. The absolute accuracy is assumed to be 0.5 °K. With an accurate current source two fixed collector currents between $10^{-9} - 10^{-2}$ A can be adjusted (acc. ≈ 1 ‰ ) [64]. The $V_{EB}$ is measured on a digital HP 3450 B-multimeter with an accuracy of 0.1 mV or better. At each temperature two $V_{EB}$ values are measured by switching between the two pre-adjusted collector currents. This makes it possible to check the temperature derived from the slope of the $I_C$-$V_{EB}$ characteristic with the temperature derived from the thermocouple voltage during each measurement. The difference for the transistors used in the bandgap investigation was always less than about 0.5 K. This means that the non-ideality factor m in the $I_C$-$V_{EB}$ relation deviates less than 2‰ from the ideal value 1. This check is very important to ensure the reliability of the derived changes in bandgap. It was found that high-frequency transistors in particular, which have thin base regions, may sometimes deviate appreciably from the ideal slope because the total base charge $Q_B$ is sensitive to changes in $V_{EB}$ (compare the Early effect for the collector base junction) [65, 66].

It should be remarked that the above mentioned accuracy between the two temperatures makes transistors very suitable for use as accurate absolute thermometers over a large temperature range [67, 68].

### 3.3. "The pn-product in silicon" (see 5.4)

This paper presented calculations of the pn-product ($n_i^2$) as a function of temperature for a number of impurity concentrations. The calculations were based on the expressions for the density of energy states functions resulting from the work of Kane, Morgan and Bonch-Bruyevich. Their theoretical considerations are very complicated and show that the random distribution of the impurities introduces local potential fluctuations which cause band tails for the impurity band as well as for the conduction and valence bands. This effect increases with higher impurity concentrations. An important parameter in these theories is the screening length $\lambda$. This is a measure of the distance over which the Coulomb potential of an ion is effective ($V = q^2 / 4 \pi \epsilon r \cdot \exp(-r/\lambda)$). Several authors (see 5.4) have used the results of these theories for calculating the pn-product as a function of impurity concentration and for different rates of compensation. In this paper these theories were used to investigate whether $n_i^2$ as a function of temperature and impurity concentration could be described by the following, already experimentally found, relation :

$$p_o n_o = n_i^2 (T, N) = CT^3 \exp\left\{-q V_g(N) / kT\right\} \quad . \tag{3.12}$$

After differentiation, an expression for ($V_g - T \, d V_g / dT$) follows which can be derived from the calculated $n_i^2$ values (see eq. 9 in 5.4). The result for different impurity concentrations is shown in fig. 3 of section 5.4. It appears that in a practical temperature range $V_g(N, T) \approx V_{go}(N) - \alpha T$. Moreover, the calculated values for $V_{go}(N)$ and the constant C agree well with the corresponding values derived from pn measurements (see fig. 3 and Table 1 in 5.4).

It should be noted that calculations made at impurity concentrations above $5 \cdot 10^{19}$ start to deviate from this relationship. One of the probable reasons is that Fermi statistics at these very high impurity concentration causes that C is no longer a constant.

### 3.4. Discussion

A lively discussion is going on in the literature [63, 69, 70] about the presence and magnitude of the bandgap narrowing at high impurity concentrations because it is important for the understanding and optimal design of semiconductor devices in general. The interpretation of the experiments is often complicated and leads to rather large differences. Several of these difficulties will be discussed below.

(1) In our measurements (see 3.2 and 5.3) the product ($\mu_n n_i^2$) is determined as a function of temperature for a number of base-dope concentrations. And although for high impurity concentrations ($N > 10^{17}$ cm$^{-3}$) there is no experimental evidence

that the temperature dependence of the minority and majority carrier mobilities is the same, we still have made that assumption (see eq. (3.9) ). This raises the question of the correctness of the derived $n_i^2$ values. In the papers reprinted in 5.3 and 5.5 it is argued that the given interpretation agrees with theoretical calculations and gives a quantitative explanation for the measured injection of minority carriers into heavily doped regions for a large number of device structures. Furthermore, highly improbable values for $\mu_n$ would result from the measurements of $(\mu_n n_i^2)$ if the classical values for $n_i^2$ were assumed.

(2) A large difference is found between the bandgap values derived from the $n_i^2$ measurements and those obtained from optical absorption measurements (see fig. 7 in 5.3 and [63] ). The reason of this discrepancy is not known. At high impurity concentrations the band-edges disappear and the physical concept of the bandgap is no longer well-defined.
As a definition of the bandgap we have used eq. (3.12) by analogy with the expression of $n_i^2$ at low impurity concentrations. However, this is an empirical definition and may in principle differ from the bandgap involved in optical absorption measurements.
On the other hand interpretation of the optical absorption measurements is also complicated because of the presence of free carrier absorption (this topic is discussed in the book by Fistul [71] ).

(3) Strain and lattice defects have been suggested as possible reasons for bandgap narrowing. At high impurity concentrations the misfit of the impurity atoms in the silicon lattice can change the bandgap. X-ray measurements [72] show that the change in the lattice parameter of silicon increases with impurity concentration but is always less than about $8 \cdot 10^{-3}$ Å, so that $\Delta a / a < 1.5 \cdot 10^{-3}$.
The estimated bandgap narrowing of this effect is

$$\Delta E_g = \frac{3\,\Delta a}{\chi\, a} \left(\frac{dE_g}{dP}\right) , \tag{3.13}$$

where the compressibility $\chi$ is equal to $0.98 \cdot 10^{-12}$ cm$^2$/dyne and the pressure dependence of the bandgap is $dE_g/dP = -1.5 \cdot 10^{-12}$ eV dyne$^{-1}$ cm$^2$ for silicon. This predicts $\Delta E_g < 7 \cdot 10^{-3}$ eV.It seems therefore that this effect can be neglected. Another argument for believing that strain is probably not the explanation for bandgap narrowing is that minority carrier injection into heavily doped regions is not very sensitive to the dopant atoms used (arsenic emitters should give much less strain than phosphorus emitters because the radius of arsenic atoms deviates much less than phosphorous atoms from silicon atoms).

(4) Partial ionization has been suggested [73] as a possible source of difficulties, particularly affecting the results of the temperature measurement. Calculations [74] show that due to partial ionization the concentration of mobile carriers would be 20 - 50% less than the total impurity concentration in the dope range of $10^{17}$ –

$2 \cdot 10^{18}$ cm$^{-3}$. For higher impurity concentrations complete ionization is present. From a practical point of view it seems justified to assume complete ionization for the whole impurity concentration range, the more so because this corresponds to the experimental mobility data which are based upon the same assumption. The calculations of the pn-product also took into account the total carrier concentrations without discriminating between carriers in the impurity band and in the conduction and valence band.

(5) Recently Martinelli [69] reported measurements of IV characteristics for silicon bipolar transistors as a function of temperature. He concluded that there was no evidence of bandgap narrowing. Because the base doping was about $10^{18}$ cm$^{-3}$ a narrowing of 42 mV would be predicted by our empirical formula (see 5.3). The reason for this disagreement is analysed in a short comment (see 5.6). After a revision of his analysis [75] Martinelli confirmed our interpretation in terms of bandgap narrowing.

# 4. EMITTER EFFICIENCY

## 4.1. Introduction

The current gain $h_{FE}$ of modern bipolar transistors is usually determined by the emitter efficiency $\eta$, which is defined for an NPN transistor as

$$\eta = J_n / (J_n + J_p) \tag{4.1}$$

where $J_n$ is the injected electron current density and $(J_n + J_p)$ is the total current density at the emitter base junction. When it is assumed that the impurity concentrations in emitter and base are constant ($N_E$ and $N_B$), then classical transistor theory states:

$$J_n = \frac{q\, n_{io}^2\, D_{nB}}{N_B\, W_B} \exp(q\, V_{EB} / kT) \tag{4.2}$$

$$J_p = \frac{q\, n_{io}^2\, D_{pE}}{N_E\, W_E} \exp(q\, V_{EB} / kT)\ . \tag{4.3}$$

The current gain is then given by:

$$h_{FE} = \frac{J_C}{J_B} \approx \frac{J_n}{J_p} = \frac{N_E\, W_E\, D_{nB}}{N_B\, W_B\, D_{pE}} = \frac{R_{\square EB}}{R_{\square E}}\ \frac{\mu_{pB}\, D_{nB}}{\mu_{nE}\, D_{pE}} \tag{4.4}$$

where $R_{\square EB}$ and $R_{\square E}$ are the base and emitter sheet resistances. Although this simple description has been used for many years in the design of transistors, it suffers from a number of serious difficulties:

- It predicts unrealistically high current gain values. Consider for instance a transistor with the following data: $N_E = 5 \cdot 10^{20}\,cm^{-3}$, $N_B = 10^{17}\,cm^{-3}$, $W_E = 2 \cdot 10^{-4}\,cm$, $W_B = 10^{-4}\,cm$, $D_{nB} = 20\,cm^2/sec$ and $D_{pE} = 1\,cm^2/sec$. Eq. (4.4) then gives $h_{FE} = 2 \cdot 10^5$. This has to be compared with a measured value of about 100.

- It predicts a continuous increase of $h_{FE}$ with higher emitter doping. In reality, however, the current gain saturates more or less at impurity concentrations above $10^{19}\,cm^{-3}$.

- Eq. (4.4) predicts a negative temperature coefficient of the current gain on the ground that the diffusion constant in the heavily doped emitter would increase while the diffusion constant in the lower doped base would decrease with temperature (see 1.4). However, measurements show a positive temperature coefficient of the $h_{FE}$.

Although these difficulties have been recognized and discussed for a long time, a quantitative explanation of the current gain and its temperature dependence for different doping profiles was seriously obstructed by the lack of reliable experimental data about *the influence of the impurity concentration on the bandgap and on the minority carrier lifetime.* These experimental data, becoming only recently available (see fig. 1.8 in section 1.5 and chap. 3), made it possible to remove much of the uncertainties.

This has been illustrated already by the agreement between the numerical calculations made with TRAP1 and the measurements (see section 2.2.2).

It was also shown that using these two effects an extra recombination peak in the neutral emitter region (see fig. 2.16) is predicted which is important for the interpretation of the voltage and temperature dependence of the base current. In order to verify the presence of this recombination peak, thin layers have been etched from the surface of the emitter [76] and the influence of the thinner emitter on the current gain repeatedly measured (see also section 5.7). This experiment has been simulated with TRAP1. It was expected that when the emitter contact, after etching, reaches the recombination peak, the emitter would become more and more

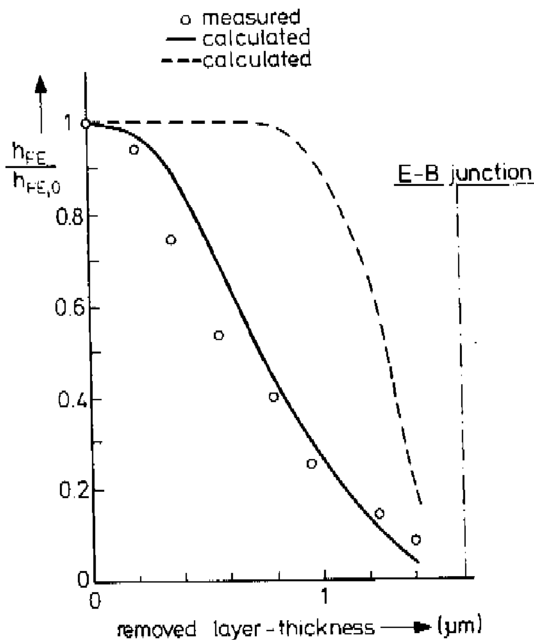



Fig. 4.1. Comparison between measured and calculated reduction of the maximum current gain as a function of the removed emitter layer. The dashed line is calculated with $n_i^2$ values derived from the optical absorption measurements of Vol'fson and Subashiev [57] and the full drawn line follows for $n_i^2$ values derived from our measurements discussed in chap. 3.

"transparent" and would affect the injected hole current. The result would be an increasing base current and therefore a decreasing current gain. In fig. 4.1. this simulation and the measured values are compared and agree rather well. When bandgap narrowing occurs only for $N > 2 \cdot 10^{19}$ (see the dashed line in fig. 4.1) this simulation shows that the fall-off starts much too late (more than 0.8 $\mu$m has to be etched away). In this case it was necessary to assume a much smaller value for the lifetime $\tau_0$ in order to calculate reasonable values for the base current corresponding

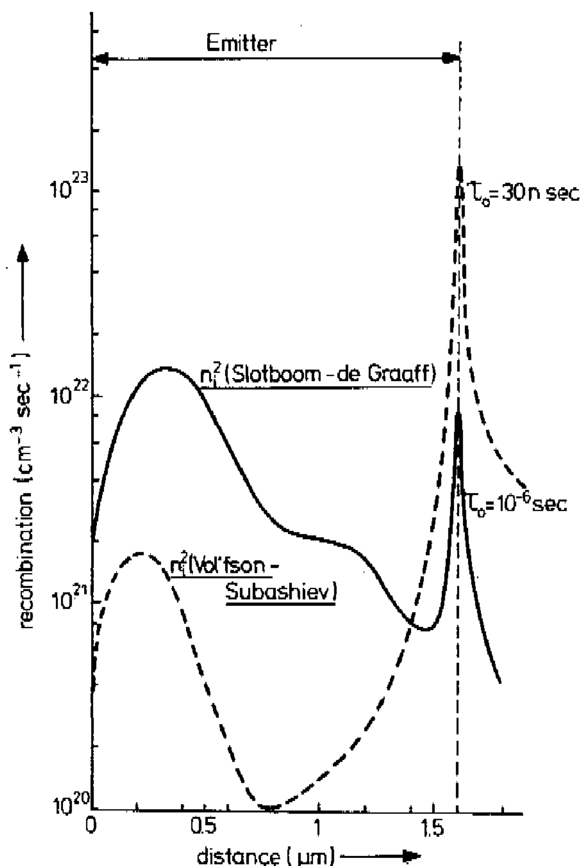



Fig. 4.2. Distribution of the recombination inside the emitter at $V_{EB}$ = 700 mV for two different $n_i^2$ models.

to $h_{FE} \approx 150$ (fig. 4.2). Moreover it is found (see fig. 4.3) that such a short lifetime introduces a large non-ideal ($m > 1$) base current component which is not in agreement with the measured $I_B$-$V_{EB}$ characteristic.

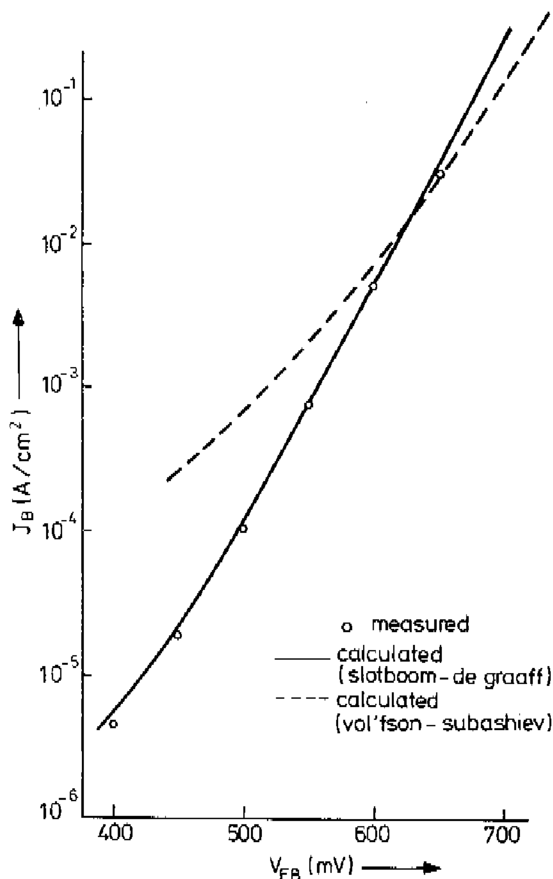



Fig. 4.3. $J_B$-$V_{EB}$ characteristic. Measurements are compared with calculations using two different $n_i^2$ models.

It may be concluded that this etching experiment proves that most of the recombination takes place in the neutral emitter at a distance of about 1.4 $\mu$m from the junction. This confirms the dominance of the Auger recombination peak instead of the usually assumed recombination maximum in the junction space-charge region.

The classical equations (4.2) and (4.3) will change when the high doping effects are taken into account and are usually described by

$$J_n = \frac{q\, n_{io}^2}{G_B} \exp(q\, V_{EB} / kT) \tag{4.5}$$

and

$$J_p = \frac{q\, n_{io}^2}{G_E} \exp(q\, V_{EB} / kT)\ , \tag{4.6}$$

where $G_B$ is called the Gummel number and $G_E$ is a figure of merit for the emitter region.

When bandgap narrowing in the base region is present $G_B$ is given by

$$G_B = \int_0^L \frac{1}{D_n}\, p\, (\frac{n_{io}}{n_i})^2\, dx\ . \tag{4.7}$$

Injection into the emitter is more complicated because, apart from the bandgap narrowing, recombination in the bulk and at the contact is also involved.
In the following sections analytical expressions for $G_E$ are given.

## 4.2. "Minority carrier injection into heavily doped silicon" (see 5.6)

For the injection of minority carriers (holes) into a heavily doped $n^+$ region it is assumed that the impurity concentration is constant ($N_E$). It is well known that the excess hole concentration is given by the diffusion equation:

$$\frac{d^2\,\hat{p}}{dx^2} = \frac{\hat{p}}{L^2} \tag{4.8}$$

where $L = \sqrt{D_p \tau}$ is the recombination-diffusion length and depends upon the impurity concentration via $D_p$ and $\tau$ (see 1.4 and 1.5). This classical equation is solved in the normal way with a finite surface recombination velocity s at the emitter contact and using the Boltzmann relationship for the hole concentration at the edge of the space charge region $x = W_E$ (see fig. 1 in 5.6). It follows that $G_E$, defined by eq. (4.6), is given by the next expression:

$$G_E = f\,\frac{N_E\, W_E}{D_{pE}}\ (\frac{n_{io}}{n_{iE}})^2 \tag{4.9}$$

and $f = F^{-1} L / W_E$, where F is a function of s and $W_E / L$ (see eq. 4 of 5.6).
Fig. 4.4 shows how $G_E$ depends upon s and $W_E / L$. When the emitter region is not transparent, that means $W_E / L \gg 1$, all of the injected minorities recombine in the bulk of the emitter and do not reach the surface. In that case the surface conditions

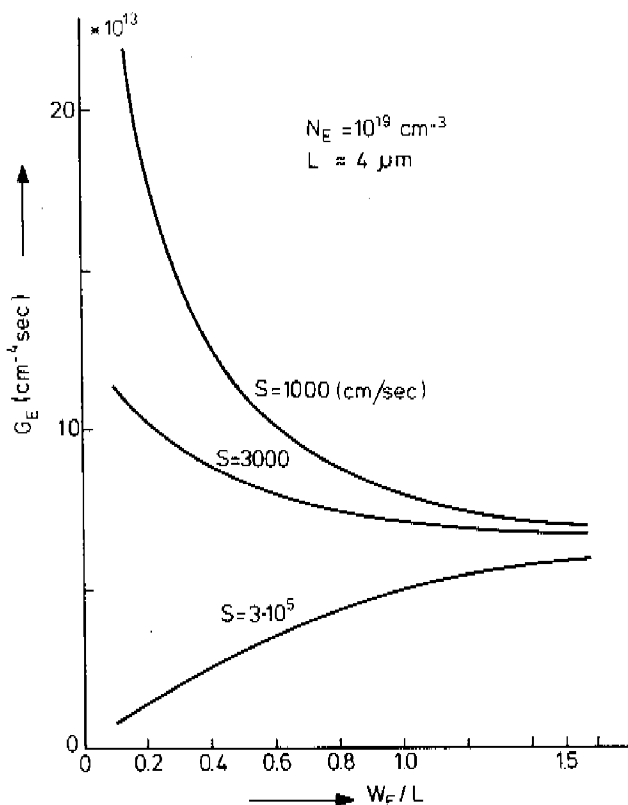



Fig. 4.4. Influence of the surface recombination velocity on the minority carrier injection into a heavily doped transparent region (calculated with eq. (4.9) ).

do not influence the injection. How $G_E$ depends on the impurity concentration and the lifetime is shown in 5.6, fig. 2, and agrees rather well with experiments. This simple model explains the fact that measurements of injection into heavily doped regions appear to be rather insensitive of the fabrication process (diffusion depth, dopant atoms or doping profile). This was already noticed by Burtscher, Dannhäuser and Krausse [77] and is clearly shown in fig. 4.5 which has been taken from their work. Using eqs. (4.7) and (4.9) the current gain is given by

$$h_{FE} = \frac{G_E}{G_B} = f \frac{R_{\square EB}}{R_{\square E}} \frac{\mu_{pB}\, D_{nB}}{\mu_{nE}\, D_{pE}} \exp\left(-q\, \Delta V_{gEB} / kT\right) \tag{4.10}$$

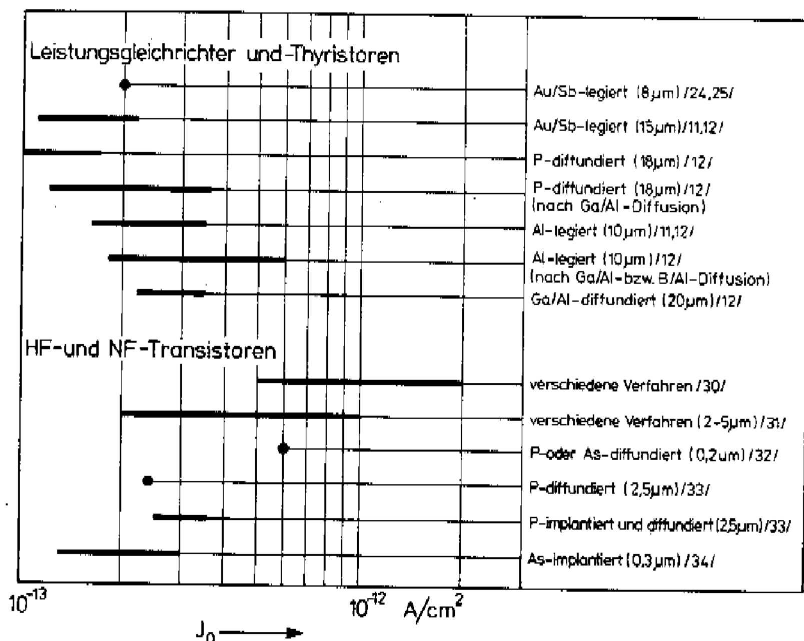



Fig. 4.5. A review of measured minority currents in heavily doped regions which were made by many different processes ($J_0 = q\, n_{io}^2 / G_E$, see eq. (4.6) ). The review is copied from ref. [77].

where $\Delta V_{gEB}$ is the difference in bandgap between the base and the emitter. This equation has to be compared with the classical formula eq. (4.4). The last term in eq. (4.10) causes the positive temperature coefficient of the $h_{FE}$. Values for the bandgap narrowing as a function of impurity concentration are given in chap. 3.

## 4.3. "The emitter efficiency of bipolar transistors: Theory and Experiments" (see 5.7)

In the previous section we have seen how $G_E$ depends upon impurity concentration, bandgap narrowing and recombination in a constant doped region. In this paper more realistic doping profiles are considered. A general expression for $G_E$ is derived and given as:

$$G_E = \frac{N(0)}{s} \left\{\frac{n_{io}}{n_i(0)}\right\}^2 g(0) + \int_0^{W_E} \frac{N(x)}{D_p(x)} \left\{\frac{n_{io}}{n_i(x)}\right\}^2 g(x)\, dx \qquad (4.11)$$

where g (x) is a weighting function

$$g(x) = J_p(x) / J_p(W_E) \tag{4.12}$$

and s is the recombination velocity at the emitter contact at $x = 0$.
It can be shown that for $N(x) = \text{const.}$ eq. (4.11) simplifies to the already described formula for $G_E$ (see eq. 4.9).

From one-dimensional calculations (see 2.2) for a large number of Gaussian doping profiles it appeared that g (x) can be characterized by only one parameter, $L_{p,\,eff}$. This effective recombination diffusion length is defined as the distance between the edge of the neutral emitter ($x = W_E$) and the point where the hole current has been halved. In the plots of the internal recombination distribution (see 2.2.2, fig. 2.1.6) this corresponds to the Auger recombination peak.
The calculated weighting functions g (x) could be fitted, after normalization with a position variable

$$u = (W_E - x) / L_{p,\,eff} \;, \tag{4.13}$$

by the following expression (see fig. 1 in 5.7) :

$$G(u) = \exp\left\{ -\left(\frac{u}{1.096}\right)^4 \right\} . \tag{4.14}$$

Fig. 4 in 5.7 shows how $L_{p,\,eff} / W_E$ depends upon the surface concentration N (0). For gaussian doping profiles however, it is not very sensitive to the shape of the doping profile. For deeply diffused emitters ($W_E / L_{p,\,eff} > 1.5$) the surface recombination is not important and g (x) is independent of the surface recombination velocity. When the emitter is transparent ($W_E / L_{p,\,eff} < 1.5$) the surface recombination changes the weighting function. This effect can also be modelled analytically, as described in 5.7, eq. (10), which approximates the numerical calculations rather well (see 5.7, fig. 3).

Several experimental results could be explained with the model just given:

(1) The etching experiment (see section 4.1). From the measured $G_E$ value and the doping profile the effective recombination diffusion length $L_{p,\,eff}$ is found from eq. (4.11) to be 1.4 μm. This agrees well with the position of the recombination peak calculated in 2.2.2.1. When so much emitter material is etched away that the remaining emitter region becomes transparent, the measured current-gain values indicated that the surface recombination velocity was about $3 \cdot 10^5$ cm/sec.

(2) Eq. (4.11) predicts that $G_E$ is proportional to L for a Gaussian doping profile. This has been verified by experiments on a number of LEC transistor structures (see 4.4) with increasing diffusion depths (see 5.7, fig. 8).

(3) Minority carrier injection in $I^2L$ (see 5.7 fig. 9) and LEC structures.
The injected minority carriers recombine at the surface, which is partly covered with an aluminium contact and partly with $SiO_2$. Investigation of a number of structures varying in the ratio of area covered by Al and $SiO_2$ makes it possible to derive $s \gg 10^5$ cm/sec for the surface recombination velocity at the Al contact and $s \approx 3000$ cm/sec at the Si-$SiO_2$ interface. Caution is necessary with these quantities because of possible deviations of the recombination current components from the ideal voltage dependence. The measured bandgap narrowing values (see chap. 3) were consistent with the minority currents measured for the $I^2L$-structures [78, 79].

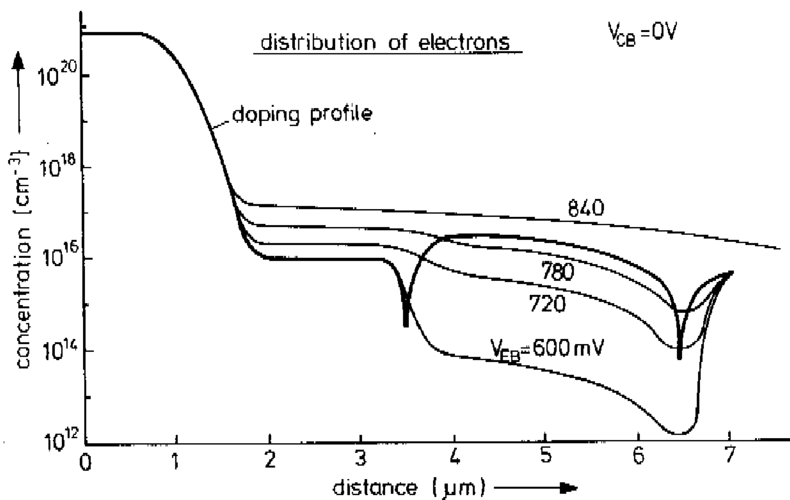
(4) Polysil emitter. This is a new transistor structure having an $n^+$ emitter consisting of doped polysilicon. The emitter-base junction lies precisely in the monocrystalline silicon. This transistor structure deviates considerably from all other known emitters in having a $G_E$ value about an order of magnitude higher. Although the detailed physical mechanism is not yet known, it can be described by an effective surface recombination velocity of about 5000 cm/sec for an assumed surface concentration of $10^{20}$ $cm^{-3}$.

### 4.4. "Some aspects of LEC transistor structures" (see 5.8).

Several years ago a Low-Emitter-Concentration (LEC) transistor structure was suggested. The emitter region consisted of a lightly doped $n^-$ region followed by a usual $n^+$ region, and can be considered as the inverse of a double-diffused transistor as is used in an $I^2L$ structure. It was suggested that the $nn^+$ barrier in the emitter would "reflect" the minority carriers, resulting in extremely high current gain values ($h_{FE} \approx 2000$).
Measurements and numerical calculations, however, show that the minority carrier current is determined by recombination in the $n^+$ region and is about equal to that in a corresponding conventional transistor without the $n^-$ layer.

Possible advantages of this transistor structure are probably of a more technological nature, e.g. the total base doping is independent of the emitter $n^+$ diffusion or implantation and can be made rather small, resulting in high current gain values.
In practical current ranges the electron current will cause a voltage drop in the $n^-$ region, thus influencing the distribution of the injected hole charge. This follows directly from eq. (16) and fig. 2 in section 5.8. One of the main effects of the $n^-$ layer is the extra charge storage in that region, particularly at higher currents (see fig. 4.6). This reduces the $f_T$ values appreciably compared with a corresponding conventional transistor (see 5.8, fig. 7).
When the area of the $n^+$ region becomes small in relation to the total emitter area, voltage drop in the $n^-$ region results in an inhomogeneous distribution of the emitter and base currents (see 5.8, fig. 9).

a.




b.

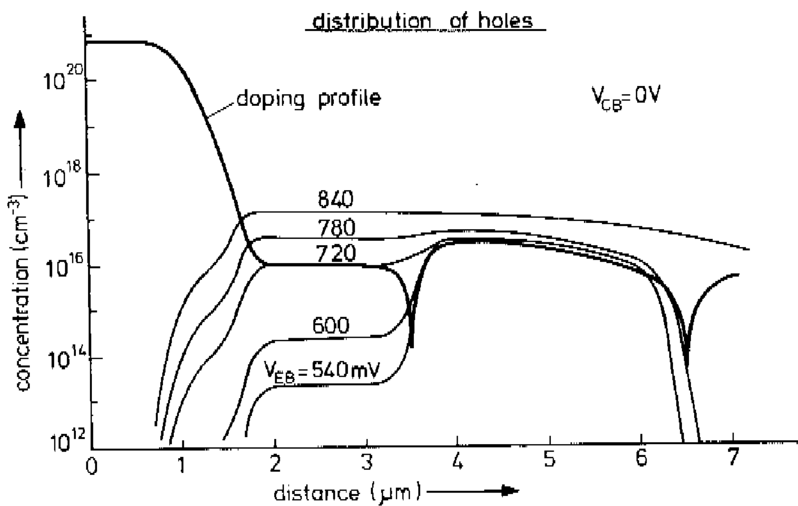



**Fig. 4.6.** Distribution of electrons (a) and holes (b) in a LEC transistor, calculated for several forward voltages.

The lightly doped emitter region offers the possibility of making other diffusions e.g. a $p^+$ ring around the $n^+$ area. The transistor characteristics can be influenced strongly by the voltage applied to this $p^+$ ring.

## REFERENCES


1. W. Shockley, Bell Syst. Techn. J., Vol. 28, p. 435 (1949).
2. H.K. Gummel, IEEE Transact. El. Dev., Vol.ED-16, p. 64 (1964).
3. W. van Roosbroeck, Bell Syst. Techn. J., Vol. 29, p. 560 (1950).
4. H.D. Barber, Solid-St. Electr., Vol. 10, p. 1039 (1967).
5. R.J. van Overstraeten, H.J. de Man, R.P. Mertens, IEEE Transact. El. Dev., Vol.ED-20, p. 290 (1973).
6. M.S. Mock, Solid-St. Electr., Vol. 16, p. 1251 (1973).
7. A.H. Marshak, K.M. van Vliet, to be published in Solid-St. Electr. (1977).
8. E. Tannenbaum, Solid-St. Electr., Vol. 12, p. 123 (1961).
9. J.R. Ehrstein (editor), Spreading Resistance Symposium, NBS special publication 400-10 (1974).
10. W.K. Hofker, H.W. Werner, D.P. Oosthoek, H.A.M. de Grefte, Radiation Effects, Vol. 17, p. 83 (1973).
11. R.W. Dutton, A.G. Gonzalez, R.D. Rung, D.A. Antoniadis, Proc. Third Internat. Symposium on Silicon Materials Science and Technology: Semicond. Silicon 1977, Electrochem. Soc., p. 910.
    P. Shah, ibid. p. 923.
12. R.E. Thomas, A.R. Boothroyd, Solid.St. Electr., Vol. 11, p. 365 (1968).
13. B.R. Chawla, H.K. Gummel, IEEE Transact. El. Dev., Vol.ED-18, p. 178 (1971).
14. R.L. Kronquist, J.P. Soula, M.E. Brilman, Solid-St. Electr., Vol. 16, p. 1159 (1973).
15. C. Jacoboni, C. Canali, G. Ottaviani, A. Alberigi Quaranta, Solid-St. Electr., Vol. 20, p. 77 (1977).
16. J.C. Irvin, Bell Syst. Techn. J., Vol. 41, p. 387 (1962); G. Baccarani, P. Ostoja, Solid-St. Electr., Vol. 18, p. 579 (1975); S. Wagner, J. Electrochem. Soc.: Solid-St. Science and Techn., Vol. 119, No. 11, p. 1570 (1972).
17. D.M. Caughey, R.E. Thomas, Proc. IEEE, Vol. 55, p. 2192 (1967).
18. J.F. Gibbons, IEEE Transact. El. Dev., Vol.ED-14, p. 37 (1967).
19. W.W. Gärtner, "Transistors, Principles, Design and Applications", Van Nostrand, Princeton (1960).
20. G.W. Ludwig, R.L. Watters, Phys. Rev., Vol. 101, Nr. 6 (1956).
21. D.A. Evans, P.T. Landsberg, Solid.St. Electr., Vol. 6, p. 169 (1963).
22. W. Shockley, W.T. Read, Phys. Rev., Vol. 87, p. 835 (1952).
23. J. Beck, R. Conradt, Solid-St. Commun., Vol. 13, p. 93 (1973).
24. J. Dziewior, W. Schmid, to be published in Appl. Phys. Lett., 1977.
25. P.A. Iles, S.I. Soclof, Proc. of the Photovoltaic Specialists' Conference (1975).
26. J. Krausse, Solid-St. Electr., Vol. 17, p. 427 (1974).
27. J.P. Woerdman, Philips Res. Suppl., No. 7 (1971).
28. P. Henrici, "Discrete variable methods in ordinary differential equations", John Wiley and Sons, Inc., New York (1962).
29. J.L. Moll, I.M. Ross, Proc. IRE, Vol. 44, p. 72 (1956).
30. M.S. Mock, Solid-St. Electr., Vol. 15, p. 1 (1972).

31. C.T. Kirk, IRE Transact. El. Dev., Vol.ED-9, p. 164 (1962).
32. H.C. de Graaff, "Electrical behaviour of lightly doped collectors in bipolar transistors", Thesis, Techn. Univ. Eindhoven (1975).
33. R.S. Varga, "Matrix Iterative Analysis", Prentice-Hall (1962).
34. C. Weber, Philips Res. Repts. Suppl., No. 6 (1967).
35. D.L. Scharfetter, H.K. Gummel, IEEE Transact. El. Dev., Vol.ED-16, p. 64 (1969).
36. H.C. de Graaff, J.W. Slotboom, Internat. El. Dev. Meeting, Washington (1972).
37. H.C. de Graaff, Philips Res. Repts. 24, 34-52 (1969).
38. H.H. Heimeier, IEEE Transact. El. Dev., Vol.ED-20, p. 708 (1973).
39. O. Manck, W.L. Engl, IEEE Transact. El. Dev., Vol.ED-22, p. 339 (1975).
40. S.P. Gaur, D.H. Navon, IEEE Transact. El. Dev., Vol.ED-23, p. 50 (1976).
41. C.R. Jesshope, E.J. Zaluska, H.A. Kemhadjian, Electr. Lett., Vol. 11, no. 14, p. 285 (1975).
42. M.S. Mock, Solid-St. Electr., Vol. 17, p. 819 (1974).
43. G.D. Hachtel, M.H. Mack, R.R. O'Brien, H.F. Quinn, Internat. El. Dev. Meeting, Washington (1976).
44. W.L. Engl, O. Manck, A.W. Wieder, Invited paper at Solid-State Dev. Research Conf., Grenoble (1975).
45. J.J. Barnes, Thesis, University of Michigan (1976).
46. J.J. Barnes, R.J. Lomax, IEEE Transact. El. Dev., Vol.ED-23, p. 1042 (1976).
47. J.G. Ruch, D.L. Scharfetter, Internat. El. Dev. Meeting, Washington (1973).
48. Pulse measurement equipment developed by A.P.M. van 't Hof (pulse-width 150 ns - 60 μs; freq. 1 kHz).
49. J.R. Hauser, IEEE Transact. El. Dev., Vol.ED-11, p. 238 (1964).
50. P.L. Hower, V.G.K. Reddi, IEEE Transact. El. Dev., Vol.ED-17, p. 320 (1970).
51. J. te Winkel, "Past and present of the Charge-Control concept in the characterization of the bipolar transistor", Advances in Electr. and Electr. Physics, Vol. 39, Acad. Press (1975).
52. H. de Man, R. Mertens, R. van Overstraeten, Electr. Lett., Vol. 9, p. 175 (1973).
53. G.L. Pearson, J. Bardeen, Phys. Rev., Vol. 75, p. 865 (1949).
54. P.P. Debye, E.M. Conwell, Phys. Rev., Vol. 93, p. 693 (1954).
55. C. Haas, Phys. Rev., Vol. 125, p. 1965 (1962).
56. M. Balkanski, A. Aziza, E. Amzallag, Phys. Stat. Sol., Vol. 31, p. 323 (1969).
57. A.A. Vol'fson, V.K. Subashiev, Sov. Phys. Sem., Vol. 1, p. 327 (1967).
58. W.L. Kauffman, A.A. Bergh, IEEE Transact. El. Dev., Vol. ED-15, p. 732 (1968).
59. D. Buhanan, IEEE Transact. El. Dev., Vol.ED-16, p. 117 (1969).
60. N.G. Nilsson, K.G. Svantesson, Solid-St. Commun., Vol. 11, p. 155 (1972).
61. V.S. Vavilov, E.L. Nolle, Sov. Phys. Sem., Vol. 2, p. 616 (1968).
62. V.M. Asnin, A.A. Rogachev, Sov. Phys. Solid-St., Vol. 5, p. 1257 (1963).
63. R.W. Keyes, Comments on Solid-St. Phys., Vol. 7, No. 6, p. 149 (1977).
64. Measurement equipment developed by R.J. van der Wal.
65. H.N. Gosh, F.H. de la Moneda, N.R. Dono, Solid-St. Electr., Vol. 10, p. 705 (1967).

66. J.R.A. Beale, J.A.G. Slatter, Solid-St. Electr., Vol. 19, p. 549 (1976).
67. T.C. Verster, Instr. Soc. of Amer., Vol. 4, Pittsburg (1972).
68. A.A. Felimban, D.J. Sandiford, Journ. of Physics E, Scient. Instr., Vol. 7, p. 341 (1974).
69. R.U. Martinelli, IEEE Transact. El. Dev., Vol.ED-23, p. 1218 (1976).
70. F.A. Lindholm, A. Neugroschel, C.T. Sah, M.P. Godlewski, H.W. Brandhorst, IEEE Transact. El. Dev., Vol.ED-24, p. 402 (1977).
71. V.I. Fistul', "Heavily Doped Semiconductors", Plenum Press, New York (1969).
72. B.G. Cohen, Solid-St. Electr., Vol. 10, p. 33 (1967).
73. F. Berz, private communication.
74. K.B. Wolfstirn, J. Phys. Chem. Solids, Vol. 16, p. 279 (1960).
75. R.U. Martinelli, IEEE Transact. on El. Dev., Vol.ED-24 (1977).
76. A procedure for etching away thin layers of emitter material has been successfully realized by A. Schmitz.
77. J. Burtscher, F. Dannhäuser, J. Krausse, Solid-St. Electr., Vol. 18, p. 35 (1975).
78. H.E.J. Wulms, IEEE Journ. of Solid-St. Circ., Vol.SC-12, p. 143 (1977).
79. H.H. Heimeier, N.H. Berger, IEEE Journ. of Solid-St. Circ., Vol.SC-12, p. 205 (1977).

# 5. REPRINTS OF PUBLICATIONS

## 5.1. "Iterative scheme for 1- and 2-dimensional dc transistor simulation"

J.W. Slotboom, Electr. Lett., Vol. 5, No. 26, Dec. 1969.

## ITERATIVE SCHEME FOR 1- AND 2-DIMENSIONAL D.C.-TRANSISTOR SIMULATION

A numerical iterative scheme is presented for the solution of the 1- and 2-dimensional semiconductor d.c. transport equations. This scheme is applied to an $n$-$p$-$n$ transistor structure. Input data are geometry, doping profile, boundary conditions and, optionally, mobility dependencies and generation-recombination law.

*Introduction:* With a view to accurate device analysis, and to be able to design accurate device models, one must regard the device as a single structure in which the basic semiconductor transport equations must be solved.

This has been realised already in 1-dimensional models for transistors and diodes. [1-3] However, realistic solutions require 2-dimensional calculations in order to include spreading effects and the influence of the surface on the electrical behaviour of the device. There exist approximate 2-dimensional transistor representations, where the structure is split up into a set of coupled 1-dimensional transistors [1-4] or into neutral and depletion regions. [5] These approximations are often not rigorous enough.

The normalised transport equations are

$$\nabla^2 \psi = n - p - N \quad (1)$$

$$J_p = -\gamma_p^{-1}(p\nabla\psi + \nabla p) \quad (2)$$

$$J_n = -\gamma_n^{-1}(n\nabla\psi - \nabla n) \quad (3)$$

$$\nabla . J_p = -R \quad (4)$$

$$\nabla . J_n = R \quad (5)$$

where

$\psi(x, y)$ = electrostatic potential
$n(x, y), p(x, y)$ = electron, hole density
$N(x, y)$ = net concentration of the ionised impurity atoms
$J_n(x, y), J_p(x, y)$ = electron, hole current density
$R(x, y)$ = generation-recombination rate
$\gamma_n^{-1}(x, y), \gamma_p^{-1}(x, y)$ = electron, hole mobility

We explicitly introduce two variables $\Phi_n$ and $\Phi_p$ :

$$\Phi_n = \exp(-\phi_n) \quad (6)$$

$$\Phi_p = \exp(\phi_p) \quad (7)$$

where $\phi_n$ and $\phi_p$ are the electron and hole quasi-Fermi potentials defined by

$$\phi_n = \psi - \ln(n) \quad (8)$$

$$\phi_p = \psi + \ln(p) \quad (9)$$

We substitute the variables $\Phi_n$ and $\Phi_p$ in the potential equation (eqn. 1) and in the two current-density equations (eqns. 2 and 3). Using the new current-density expressions in the continuity equations (eqns. 4 and 5), we find

$$\nabla^2 \psi = \Phi_n \exp(\psi) - \Phi_p \exp(-\psi) - N \quad (10)$$

$$\nabla . \{\gamma_p^{-1} \exp(-\psi) \nabla\Phi_p\} = R \quad (11)$$

$$\nabla . \{\gamma_n^{-1} \exp(\psi) \nabla\Phi_n\} = R \quad (12)$$

These are three coupled elliptic partial-differential equations for the three unknown variables $\psi$, $\Phi_p$ and $\Phi_n$.

*Problem:* The problem is to solve this set of equations for a device structure of which the following data are given: geometry, doping profile, mobilities as functions of impurity concentration and electric field, generation-recombination law and the right boundary conditions.

*Method of solution:* The iterative process, which is started with a trial solution for $\Phi_n$ and $\Phi_p$ is as follows:

($a$) We assume that $\Phi_n$ and $\Phi_p$ are known quantities and consider eqn. 10 as a non-linear elliptic differential equation in $\psi$ $(x, y)$. [7 – 8] This equation is linearised, as in Gummel's 1-dimensional approach. The resulting linear equation for the correction $\delta$ $(x, y)$ can be replaced by a set of difference equations, which can rapidly be solved by the method of successive overrelaxation (s.o.r.).[6] The difference equation at the point $(i, j)$ (see Fig. 1A) is

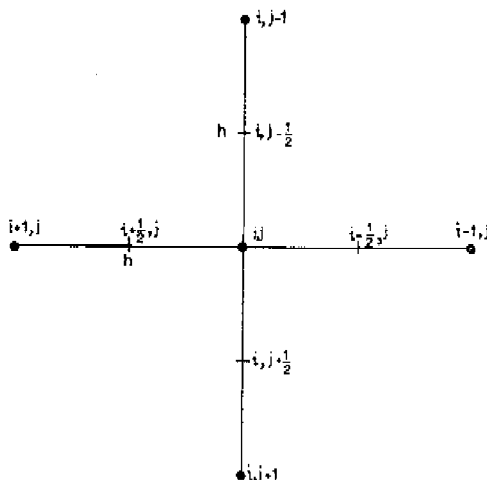



**Fig. 1A** ***Mesh points of a regular square net***

$$[4+h^2\{\Phi_{n_{i,j}} \exp(\psi_{i,j})+\Phi_{p_{i,j}} \exp(-\psi_{i,j})\}]\ \delta_{i,j}$$
$$=\delta_{i+1,j}+\delta_{i-1,j}+\delta_{i,j+1}+\delta_{i,j-1}-b_{i,j} \quad (13)$$

where $b$ is a known function of $\psi$, $\Phi_p$, $\Phi_n$ and N.

Convergence of the s.o.r. is always guaranteed by the diagonal dominance criterion. [6] With the calculated $\delta$ values, we correct the old $\psi$ values and so find new $\psi$ values.

($b$) Because $\gamma_p$, $\gamma_n$ and $R$ depend in a given way on the electrostatic potential, we can, at this point in the iteration cycle, compute new values for these three quantities.

(*c*) Now we assume that $\psi$, $\gamma_p$, $\gamma_n$ and $R$ are known quantities and consider eqns. 11 and 12. These linear equations are also replaced by two sets of difference equations for $\Phi_p$ and $\Phi_n$ and are solved by s.o.r. The difference equation for eqn. 11 at the point $(i, j)$ (see Fig. 1A) is (Reference 6, p. 184)

$$-(a_{i+\frac{1}{2},j}+a_{i-\frac{1}{2},j}+a_{i,j+\frac{1}{2}}+a_{i,j-\frac{1}{2}})\Phi_{p_{i,j}} + a_{i+\frac{1}{2},j}\Phi_{p_{i+1,j}} + a_{i-\frac{1}{2},j}\Phi_{p_{i-1,j}} + a_{i,j+\frac{1}{2}}\Phi_{p_{i,j+1}} + a_{i,j-\frac{1}{2}}\Phi_{p_{i,j-1}} = h^2 R_{i,j} \qquad (14)$$

where $a = \gamma_p^{-1} \exp(-\psi)$.

The value of $a$ at the point midway between two mesh points is approximated from the known values of $a$ at these two mesh points. The difference equation for eqn. 12 is similar to eqn. 14, but with $a = \gamma_n^{-1} \exp(\psi)$.

For both sets of difference equations, convergence is unconditionally guaranteed by the diagonal dominance criterion. With the computed $\Phi_p$ and $\Phi_n$ values we return to the beginning (*a*) for computing new $\psi$ values etc.

Our preliminary computer calculations demonstrate that this iterative scheme converges very wel.

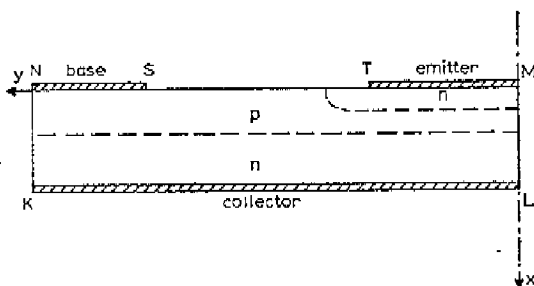



**Fig. 1B** ***2-dimensional geometry with metallurgical emitter and collector junctions of an n-p-n transistor.***

*Results:* Although this very general and rigorous solution method can be used not only for transistor calculations but for any 1- or 2-dimensional semiconductor structure, we will demonstrate the method for a 2-dimensional *n-p-n* germanium transistor. The geometry is shown in Fig. 1B. For the doping profile, we use the solution of the 2-dimensional diffusion equation.[9] For reasons of simplicity, we assume that the electron and hole mobilities are constant and that the generation-recombination rate equals zero.

The boundary conditions are

(i) At the emitter, base and collector contacts, we assume that the semiconductor is in thermodynamic equilibrium and that there exists charge neutrality (the usual assumptions for ideal ohmic contacts). Together with the applied voltages $V_{EB}$ and $V_{CB}$, we can then find the boundary values of $\psi$, $\Phi_p$ and $\Phi_n$ at the contacts.

(ii) Along ML, which is an axis of symmetry, we have

$$\frac{\partial\psi}{\partial y} = \frac{\partial\Phi_p}{\partial y} = \frac{\partial\Phi_n}{\partial y} = 0$$

(iii) Along NK, where the behaviour is approximately 1-dimensional, we have

$$\frac{\partial\psi}{\partial y} = \frac{\partial\Phi_p}{\partial y} = \frac{\partial\Phi_n}{\partial y} = 0$$

(iv) Along TS, we assume that the normal components of the current densities and of the electrostatic field are zero. Therefore we have

$$\frac{\partial\psi}{\partial x} = \frac{\partial\Phi_p}{\partial x} = \frac{\partial\Phi_n}{\partial x} = 0$$

With the given iterative scheme we solved this 2-dimensional transistor problem. In Fig. 2 the computed electron-density distribution is shown for $V_{EB} = -250$ mV and $V_{CB} = 0.5$ V.

The process converged very well, and within eight iteration cycles the computed new potential values differed not more than one percent from the old ones. The computation time at the EL-X8 machine was about 15 min. From Fig. 2 it can be seen that, in the neighbourhood of the base contact, the electron density is smaller than its equilibrium value due to the vicinity of the collector junction.

The iteration process can also be used for solving 1-dimensional structures. For instance, we computed with this method the electrostatic and quasi-Fermi potentials for the 1-dimensional transistor along ML ($y = 0$) in Fig. 1B. Comparing these results with those following Gummel's method, we found that the agreement was better than one percent.

The author wishes to thank C. Weber for helpful discussions and P.A.H. Hart for his stimulating interest.

J.W. SLOTBOOM

*27th November 1969*

*Philips Research Laboratories*
*NV Philips' Gloeilampenfabrieken*
*Eindhoven, Netherlands*

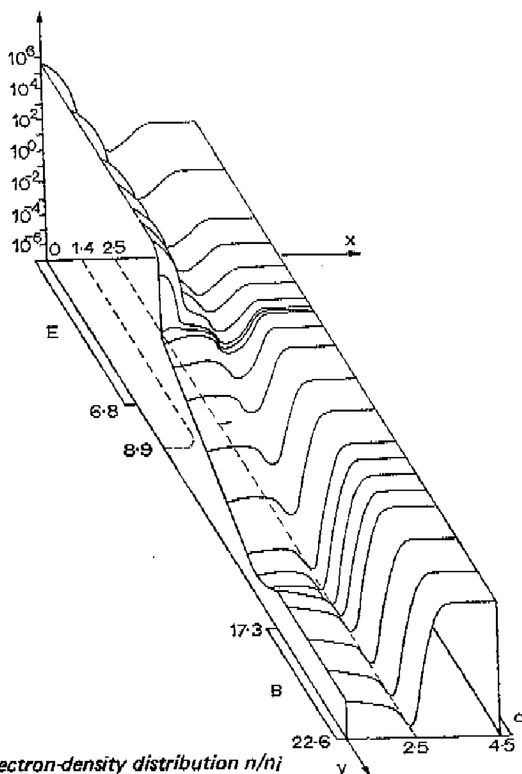



**Fig. 2** ***Normalised electron-density distribution n/ni***
$n_i = 2.5 \times 10^{13}$ $cm^{-3}$ ; $x$ and $y$ in micrometres

## References


1 GUMMEL, H.K.: 'A self-consistent iterative scheme for one-dimensional steady state transistor calculations', *IEEE Trans.*, 1964, ED-11, pp. 455–465

2 DE MARI, A.: 'An accurate numerical steady-state one-dimensional solution of the *p-n* junction', *Solid-State Electron.*, 1968, 11, pp. 33–58

3 CAUGHEY, D.M.: 'The computer simulation of gigahertz transistors', presented at the IEEE International Electronics Conference, 1967, Toronto

4 GOSH, H.N., DE LA MONEDA, F.H., and DONO, N.R.: 'Computer-aided transistor design, characterization, and optimization', *Solid-State Electron.*, 1967, 10, pp. 705–726

5 COLLINS, T.W.: 'Two-dimensional numerical analysis of integrated bipolar transistors', presented at the International Electron Devices Meeting, 1968, Washington

6 VARGA, R.S.: 'Matrix iterative analysis' (Prentice-Hall, 1962)

7 BELLMAN, R.E., and KALABA, R.E.: 'Quasilinearization and nonlinear boundary-value problems' (Elsevier, USA, 1965)

8 DUBOCK, P.: 'Numerical analysis of forward and reverse bias potential distribution in a 2-dimensional *p-n* junction with applications to capacitance calculations', *Electron. Lett.*, 1969, 5, pp. 236–238

9 KENNEDY, D.P., and O'BRIEN, R.R.: 'Analysis of the impurity atom distribution near the diffusion mask for a planar *pn*-junction', *IBM J. Res. Developm.*, 1965, 9, pp. 179–186

5.2. "Computer-aided two-dimensional analysis of bipolar transistors"
IEEE Transact. on El. Dev., Vol. ED-20, No. 8, Aug. 1973.

# Computer-Aided Two-Dimensional Analysis of Bipolar Transistors

JAN W. SLOTBOOM

*Abstract* — A method for solving numerically the two-dimensional (2D) semiconductor steady-state transport equations is described.
The principles of this method have been published earlier [1]. This paper discusses in detail the method and a number of considerable improvements. Poisson's equation and the two continuity equations are discretized on two networks of different rectangular meshes. The 2D continuity equations are approximated by a set of difference equations assuming that the hole and electron current density components along the meshlines are constant between two neighboring meshpoints in a way similar to that used by Gummel and Scharfetter [2] for the one-dimensional (1D) continuity equations. The resulting difference approximations have generally a much larger validity range than the conventional difference formulations where it is assumed that the change in electrostatic potential between two neighboring points is small compared with $kT/q$. Therefore, a much smaller number of meshpoints is necessary than for the conventional difference approximations. This reduces considerably the computation time and the required memory space. It will be shown that the matrix of the coefficients of this set of difference equations is always positive definite. This is an important property and gurarantees convergence and stability of the numerical solution of the continuity equations.

The way in which the difference approximations for the continuity equations are derived gives directly consistent expressions for the current densities that can be used for calculating the currents. In order to demonstrate the kind of solutions obtainable, steady-state results for a bipolar n-p-n silicon transistor are presented and discussed.

Manuscript received August 17, 1972; revised March 16, 1973.

The author is with the Research Laboratories, N.V. Philips' Gloeilampenfabrieken, Eindhoven, The Netherlands.

## I. INTRODUCTION

By means of numerical solution methods it is possible to solve the general semiconductor transport equations accurately without the conventional restrictions such as locally neutral or space-charge regions, constant mobilities, simplified doping profiles, etc.

In 1964 Gummel [3] described a very attractive and efficient method for solving the 1D steady-state carrier transport equations. Later many other authors published solution methods not only for steady-state but also for small-signal and transient operation of 1D diode and transistor structures [4] – [15]. With these methods it is possible to study how physical parameters such as doping profile, carrier mobilities, lifetimes, and geometry are related to the electrical behavior of the device and to get a clear insight into high-level effects that are of growing importance for device optimization and design of accurate circuit models.

Lateral variations in current densities and potentials and their relationship with base current cannot be taken into account in the 1D transistor structures. Quasi-2D models have been made to include these 2D effects [16], [17]. They consist of a number of 1D transistor models coupled with lateral base resistances. The basic assumption is that, apart from the base current, the transistor behavior is 1D. In fact, as will be shown by our rigorous 2D transistor calculations, this is not true when current spreading is present. Particularly for high current densities when base widening occurs (see, e.g., [18], [19]), current and voltage dependent current spreading takes place directly behind the emitter-base junction (see Section VII). It is also difficult to represent realistically the emitter sidewall injection in a quasi-2D model. Other approximate 2D transistor models have been made by splitting up the 2D transistor structure into neutral and space-charge regions [20], [21].

However, in order to analyze accurately transistors and other semiconductor structures under all kinds of operation conditions, the general basic carrier transport equations must be solved rigorously:

$$\nabla^2\psi = -q/\epsilon(p - n + N_D^+ - N_A^-) \qquad \text{(Poisson's equation)} \quad (1)$$

$$J_p = -q\mu_p p\nabla\psi - qD_p\nabla p$$
$$J_n = -q\mu_n n\nabla\psi + qD_n\nabla n \qquad \text{(Current density equations)} \quad (2)$$

$$\nabla\cdot J_p + q\partial p/\partial t = -qR$$
$$\nabla\cdot J_n - q\partial n/\partial t = qR \qquad \text{(Continuity equations).} \quad (3)$$

Kennedy and O'Brien [22] – [24] have presented detailed graphical results of 2D steady-state calculations for junction field-effect transistors (JFET's). Unfortunately, they give no details of the way in which these equations are solved. Kim and Yang [25] consider in their numerical analysis of JFET's the majority carriers only and assume that the quasi-Fermi potential is one dimensional. This makes direct integration of the current density equation possible, which simplifies appreciably the numerical solution.

Recently it has been shown [1] how the 1D steady-state iterative scheme given by Gummel [3] can be extended to the 2D problem. This method will be described briefly. Normalizing (1), (2), and (3) in the same way as de Mari [4] did, the steady-state equations become

$$\nabla^2\psi = n - p - N \tag{4}$$

$$J_p = -\gamma_p^{-1}(p\nabla\psi + \nabla p)$$

$$J_n = -\gamma_n^{-1}(n\nabla\psi - \nabla n) \tag{5}$$

$$\nabla\cdot J_p = -R$$

$$\nabla\cdot J_n = R. \tag{6}$$

Here $\gamma_p^{-1}$ and $\gamma_n^{-1}$ are the normalized carrier mobilities. In the expressions for the current densities the Einstein relationship between the diffusion coefficient and the mobility is used.

With two new variables $\phi_p$ and $\phi_n$, defined by

$$\phi_p = \exp(\varphi_p)$$

$$\phi_n = \exp(-\varphi_n) \tag{7}$$

where $\varphi_p$ and $\varphi_n$ are the hole and electron quasi-Fermi potentials

$$\varphi_p = \psi + \ln(p)$$

$$\varphi_n = \psi - \ln(n) \tag{8}$$

the basic equations (4) – (6) reduce to three coupled elliptic partial differential equations in $\psi$, $\phi_p$ and $\phi_n$:

$$\nabla^2\psi = \phi_n \exp(\psi) - \phi_p \exp(-\psi) - N \tag{9}$$

$$\nabla\cdot\{\gamma_p^{-1}\exp(-\psi)\nabla\phi_p\} = R \tag{10}$$

$$\nabla\cdot\{\gamma_n^{-1}\exp(\psi)\nabla\phi_n\} = R \tag{11}$$

with

$$J_p = -\gamma_p^{-1}\exp(-\psi)\nabla\phi_p \tag{12a}$$

$$J_n = \gamma_n^{-1}\exp(\psi)\nabla\phi_n. \tag{12b}$$

These three equations are discretized on a nonuniform mesh. Poisson's equation, (9), is considered as a nonlinear equation in $\psi$ and is linearized according to the Newton method (see Section III). The two continuity equations (10) and (11) are linear self-adjoint elliptic partial differential equations in $\phi_p$ and $\phi_n$. The sequence of solution of this system of equations is the same as in the 1D Gummel method: first the nonlinear Poisson equation is solved, assuming $\phi_p$ and $\phi_n$ are known, and next each of the continuity equations is solved using the just calculated electric potentials from the Poisson equation. This cycle is iterated until a sufficient accuracy is reached (see Fig. 1). The main difference between the 1D and the 2D solution method is the way each of these equations is solved numerically. In the 2D analysis the equations are solved by iterative solution methods, such as the method of successive overrelaxation (SOR) or successive line overrelaxation (SLOR) [26]. In the 1D solution process, on the other hand, the linearized Poisson equation can be solved directly by Gaussian elimination and the two continuity equations by integration. This difference in solution method introduces severe complications for the 2D solution process. The coefficient matrices of the continuity difference equations depend upon the electric potential $\psi$, which changes with each "outer" iteration. Therefore, it is necessary that independent of the values of the electric potential in each meshpoint the difference equations are numerically stable and that the iterative solution method of the individual equations is always convergent. Though the difference approximations described in [1] satisfy these conditions, we will derive in this paper other difference approximations that are more accurate and have also the mentioned properties.

With this method, taking into account the oxide and the oxide-semiconductor interface, Heydemann [27] and van Dorpe and Xuong [28] have analyzed a

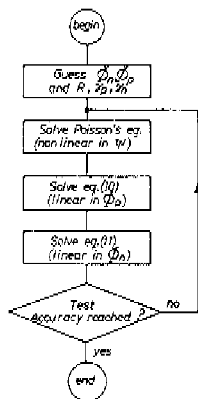



Fig. 1. Flow diagram of the iterative scheme.

metal-oxide-semiconductor (MOS) transistor. The last mentioned authors have refined the difference equations for the continuity equations in order to avoid computational difficulties for high applied voltages. Kilpatrick and Ryan [29] used the method for analyzing a lateral p-n-p transistor. Dubock [30] uses a special difference formulation for the continuity equations. Depending upon the sign of the electric field at each meshpoint a suitable difference approximation is chosen. The difference approximation thus changes in the different meshpoints, which seems unattractive from a numerical point of view. Reiser [31] has published a method for the solution of the time dependent 2D equations. Assuming that the space-charge term in (4) is known, this equation is solved by means of Hockney's method [32]. Although this is a very rapid and direct method of solution, it has some restrictions and cannot be used to solve the Newton-linearized Poisson equation. Therefore, the rapid rate of convergence of the Newton process is lost, which is especially important for low injection. Another restriction is that constant stepsizes in both directions are needed. This can introduce an excessive number of meshpoints in areas where they are not needed.

In a recent paper [33] van Dorpe, Borel, Merckel, and Saintot describe a 2D numerical analysis of the MOS transistor. The used difference approximations are the same as already given in [27]. At the end of their program they calculate the longitudinal component of the hole current density via an integral formulation because of the great inaccuracy met when the current density is calculated from numerically differentiation. Heydemann uses in his last paper [34] the strongly implicit method of Stone [35] for solving each of the three systems of difference equations instead of the more usual SOR or SLOR methods.

An important drawback of the above-mentioned methods is the large number of meshpoints involved for accurate calculations. This is caused by the used difference approximations and the way of solving. Rather large computation time and required memory space are the consequence of it. Usually the number of iterations for each of the three equations depends rather strongly upon the number of meshpoints, which increases the computation time even more.

We will show that the efficient difference approximation, which Gummel and Scharfetter used for their analysis of a silicon Read diode oscillator [2], can be extended to two dimensions. Assuming that the hole and electron current density components along the meshlines are constant between two neighboring meshpoints, it is possible to derive difference approximations for (10) and (11) that have a much larger validity range than those used in the above-mentioned methods, which are based on the assumption that the change in electrostatic potential between two meshpoints is small compared to $kT/q$. The large validity range of the new difference approximations requires less meshpoints for accurate calculations.

We will prove that the matrix built up from the coefficients of the difference equations is always positive definite [26]. This property is very important for the numerical solution because it guarantees numerical stability and convergence of, e.g., SOR and SLOR.

The use of different meshes that are optimal for each of the three equations reduces the total number of meshpoints again.

## II. DATA AND BOUNDARY CONDITIONS OF THE USED EXAMPLE

We shall illustrate our method with an example of a bipolar n-p-n silicon transistor. The geometrical structure is given in Fig. 2. The doping profile is calculated from the 2D diffusion equation [37]. The doping profile along the $X$ axis and under the base contact is shown in Fig. 3. The generation-recombination term is given by the Hall, Shockley, and Read model [38]

$$R = (pn - n_i^2)/\{\tau_{n0}(p + p_1) + \tau_{p0}(n + n_1)\} \tag{13}$$

where

$\tau_{n0}, \tau_{p0}$ electron and hole lifetime;

$p_1, n_1$ the hole and electron concentration that would exist if the Fermi level were at the trap level.

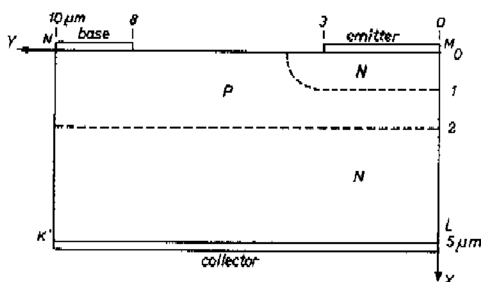



Fig. 2. Geometrical structure of n-p-n silicon transistor. Metallurgical junctions are indicated.

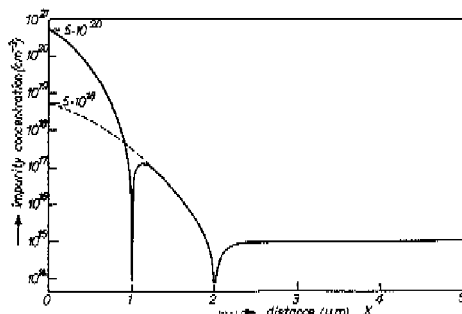



Fig. 3. Doping profile along *ML* and *NK* (see Fig. 2).

We will assume that the recombination centers are near the center of the bandgap:

$$n_1 = p_1 = n_i.$$

And further we have taken in our example

$$\tau_{p0} = \tau_{n0} = 0.5 \times 10^{-7} \text{ s}.$$

The mobility is a function of doping and electric field [3], [39]. The boundary conditions are the same as described in [1]. At the ohmic contacts the electrostatic and quasi-Fermi potentials are known from the assumed charge neutrality and equilibrium and from the given applied voltages. Along the other parts of the boundary the normal derivatives of the potentials are zero. For analyzing MOS transistors, more realistic boundary conditions at the interface with the oxide layer can be used [27]. [28]. The calculations and computer plots were made on a rather small Philips ELX8 computer with a memory capacity of about 23K words of 27 b.

## III. POISSON'S EQUATION

Poisson's equation, (9), is considered as a nonlinear equation in $\psi$ while $\phi_p$ and $\phi_n$ are known. This equation is linearized in the usual way [1], [3]:

$$\nabla^2\delta - a^2\delta = b \tag{14}$$

where

$\delta$ a perturbation on $\psi_0$;
$a^2 = \phi_n \exp(\psi_0) + \phi_p \exp(-\psi_0) = n + p$;
$b = -\nabla^2\psi_0 + \phi_n \exp(\psi_0) - \phi_p \exp(-\psi_0) - N$
$= -\nabla^2\psi_0 + n - p - N$;
$\psi_0$ the $\psi$ values of the former iteration.

Because $a^2$ is a function of $x$ and $y$ (14) cannot be solved by Hockney's method as mentioned before. This linear elliptic partial differential equation is discretized on a nonuniform mesh (see Fig. 4) and replaced by a set of finite difference equations by means of the box integration method [40]. This net of meshpoints is called the $E$-net.

The problem can be written in a matrix notation as

$$A\delta = b. \tag{15}$$

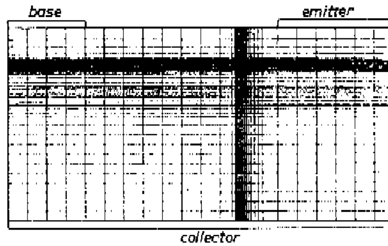



Fig. 4. Nonuniform mesh for discretizing Poisson's equation ($E$-net; 46 x 41 meshpoints).

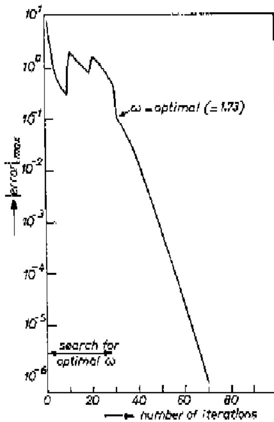



Fig. 5. Rate of convergence of the $\delta$-equation solved by SOR with $\omega$ as overrelaxation parameter.

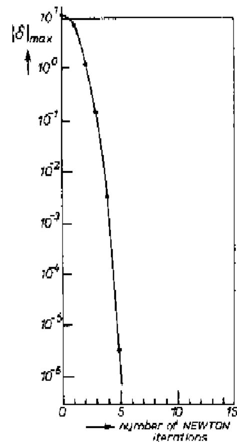



Fig. 6. Rate of convergence of the Newton process.

The matrix $A$ is a symmetric and positive definite matrix with diagonal dominance. Therefore, the usual iterative solution methods are convergent [26]. Even with this very nonuniform mesh the rate of convergence is sufficient (see fig. 5) because of the presence of the term $-a^2\delta$, which provides an extra diagonal dominance. The solution of the linearized Poisson equation is used to correct the $\psi_0$ values. This is one Newton iteration. By recalculating the coefficients $a^2$ and $b$, the process can be repeated, etc. Fig. 6 gives the quadratic rate of convergence of this Newton process. Usually it is sufficient to correct the $\psi_0$ values only once with the calculated $\delta$ values and then continue with the solution of the continuity equations.

## IV. CONTINUITY EQUATION

Consider the continuity equation for the holes, and omit the subscript $p$ for the sake of simplicity; the corresponding equation for the electrons can be treated in the same way. The functions $\gamma$, $\psi$, and $R$ are assumed to be known in the meshpoints of the $E$-net.

In Appendix A it is shown in detail how the Gummel-Scharfetter difference approximation for the current density, formulated in the variable $\phi$, is extended to two dimensions. The two continuity equations are discretized on a second network of meshpoints, the so-called $J$-net (see fig. 7). It is assumed that the current density component along the meshlines is constant between two neighboring meshpoints of the $J$-net, while the electric field component along the meshlines is assumed constant between two neighboring meshpoints of the $E$-net. The conventionally used difference approximations are only valid when the change in electric potential between two meshpoints is less than $kT/q$. In a reverse-biased base-collector junction this would require very small stepsizes [41].

According to these new approximations we use two different networks of meshpoints, the already mentioned $E$-net and $J$-net. The resulting system of difference equations can be written in matrix notation as (see Appendix A)

$$A\phi = b. \tag{16}$$

The coefficient matrix $A$ is symmetric and positive definite. This makes it possible to solve this problem by standard numerical techniques [26]. Because the $J$-net needs generally a much smaller number of meshpoints, which are more uniformly distributed than the meshpoints of the $E$-net, the rate of convergence of SOR and SLOR is much improved.

The "positive" properties of the matrix $A$ are also important for the difference approximations of the time dependent continuity equation (see Appendix C).

From the quasi-Fermi potentials in the corners of each cell of the $J$-net (points $A$, $B$, $C$, and $D$ in Fig. 8) we derive by means of bilinear interpolation the quasi-Fermi potentials in the meshpoints of the $E$-net. These are needed for the solution

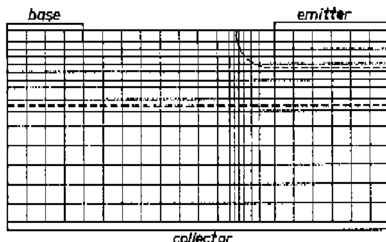



Fig. 7. Nonuniform mesh for discretizing the continuity equation ($J$-net; 16 x 25 meshpoints).

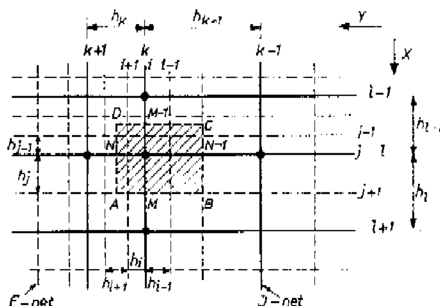


Fig. 8. The $E$-net is a refinement of the $J$-net. Meshpoints of the $E$-net are indicated by the integers $j$, $i$ and of the $J$-net by $l$, $k$. Variable stepsizes in the $x$ and $y$-directions are indicated for both nets. Points $M$, $N \cdot 1$, $N$, and $N-1$ are midway between two neighboring meshpoints.

Poisson's equation and for the calculation of the recombination in every meshpoint of the $E$-net.

In order to avoid very large values for $\phi_p$ and $\phi_n$ these variables are scaled down by means of a similarity transformation (see Appendixes A and B). Then a system of difference equations in the variable $p$ results [see Appendix A, (A12)]:

$$Tp = b. \tag{17}$$

$T$ is the transformed coefficient matrix $A$.

All the quantities in these difference equations are bounded in value and no problems with overflow can occur. The eigenvalue spectrum of the original coefficient matrix is not changed when a similarity transformation is applied. Consequently, the iterative solution process (e.g., SOR and SLOR) behaves in the same way for these difference equations.

## V. CURRENT, CURRENT DENSITY, AND CUTOFF FREQUENCY

At every midpoint between two meshpoints of the $J$-net the current density can be calculated from the same difference approximation for the current density as is used for the continuity equation:

$$J_M = -a_M \{ p_{l+1,k} \exp(\psi_{l+1,k} - \psi_{l,k}) - p_{l,k} \}. \tag{18}$$

This follows from transformation of (A9) (see Appendix A) to the variable $p$. When the quasi-Fermi potential is locally nearly constant this calculation introduces numerical inaccuracy because nearly equal numbers are subtracted. Therefore, it is difficult in this way to calculate the electron current density in the quasi-neutral emitter region or the hole current density underneath the base contact. Apart from

the difficulties in these areas the total current flowing through a line $x$ = constant (see Fig. 2) is easily calculated by integrating the current density along the line. The collector current is calculated in this way by integrating along a line $x$ = constant just before the collector contact. The base current is given by

$$I_B = -\iint_{\text{total transistor area}} R\,dx\,dy. \tag{19}$$

This follows from integration of the hole continuity equation on the assumptions that the base current consists of holes only and that holes are fed into the transistor via the base contact only.

From $I_C$ and $I_B$ the current amplification factor is calculated:

$$h_{FE} = I_C/I_B. \tag{20}$$

The cutoff frequency is calculated in accordance with the charge control principle from a perturbation of the charge concentration and the resulting change in collector current:

$$f_T = (1/2\pi)\Delta I_C/\Delta Q_p \tag{21}$$

where $\Delta Q_p$ is the change in total stored hole charge.

## VI. CONSIDERATIONS ABOUT CONVERGENCE AND CHOICE OF MESHPOINTS

In all our computer calculations the "outer" iteration (see Fig. 1) converged well except for very long device structures (e.g., larger than 100 $\mu$m) with heavy recombination in the whole device; the rate of convergence then can be extremely small which makes the method rather impractical under these conditions. This experience agrees with a recent study of this convergence problem [42].

In order to check the convergence behavior we test during each "outer" iteration the relative and absolute change in electric potential in each meshpoint. Also, the relative change in the carrier concentrations is checked in each meshpoint. In order to get an idea of the rate of convergence, we show in Fig. 9 the maximum absolute change in electric potential for two successive iterations as a function of the number of iterations. For low current levels the rate of convergence of the Newton-linearized Poisson equation dominates the overall convergence behavior and a nearly quadratic convergence results [compare the quadratic rate of convergence when Poisson's equation alone is solved by means of the Newton method (see Fig. 6)]. For higher current levels the mutual coupling between the equations becomes stronger and the convergence rate slows down (see Fig. 9). This rate of convergence can still be speeded up [43]. The convergence behavior is very similar to the 1D Gummel method.

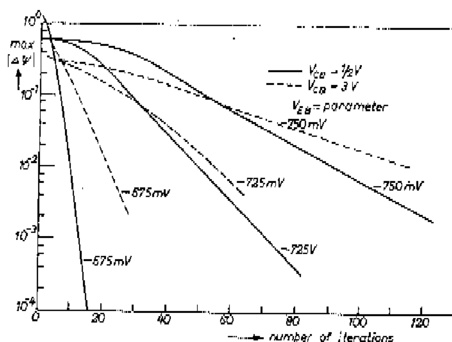



Fig. 9. Rate of convergence of the total system for different biases.

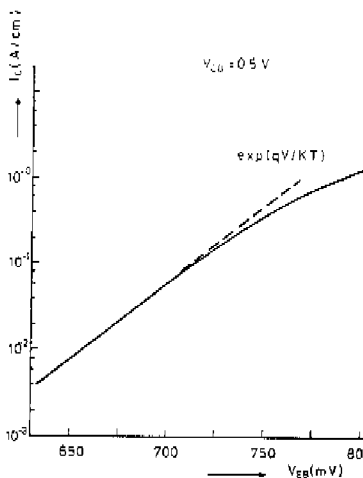



Fig. 10. Calculated $I_C V_{EB}$ characteristic for $V_{CB}$ = 0.5 V.

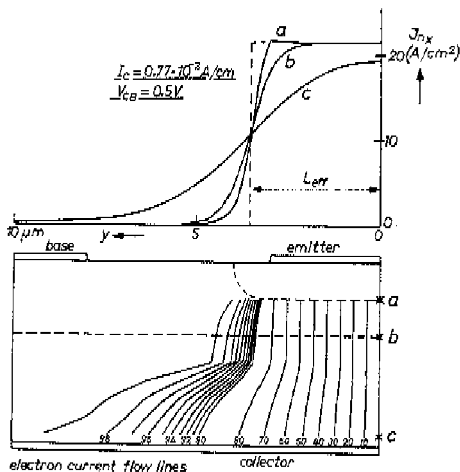



Fig. 11. Calculated electron current flow lines and current density distribution along the emitter junction (curve $a$), collector junction (curve $b$), and along the collector contact (curve $c$) are shown. The number under the flow line gives the percentage of the total current that flows between this line and the axis of symmetry ($x$ axis).

The total current flowing through lines $x$ = constant is, as a function of $x$, constant within 1 percent. This conservation of current is self-evident because the expression for calculating the current density is also used in the difference approximation for the continuity equations.

Because one wants to use a minimum number of meshpoints it is very important to choose carefully an optimal distribution of meshpoints. This depends also upon the specific phenomena and characteristics of the transistor to be studied. We have found that often errors, introduced because of locally too few meshpoints, are kept very localized and hardly influence phenomena somewhere else or the overall characteristics.

We refined the $J$-net from 400 to 729 meshpoints and found only small differences (less than 5 percent) in the carrier concentrations and the electric potential. But when we reduced the number of meshpoints to only 247, differences in these variables of the order of 20 percent and in some few points of 100 percent took place. However, the calculated total currents in the $I_C V_{EB}$ characteristic (see Fig. 10) deviate for these three $J$-nets less than 1 percent. The total base current in the last network of 247 meshpoints deviates in about 20 percent of the values calculated with the finer meshes, which were nearly equal. The number of meshpoints of the $E$-net was the same for these three $J$-nets, namely 1886.

## VII. DISCUSSION OF THE RESULTS

In a number of computer plots the internal as well as the external behavior of the n-p-n silicon transistor under normal and saturation conditions will be illustrated.

In Fig. 11 the electron current spreading under normal operation is shown. Current spreading is only present in the quasi-neutral collector region. Three-dimensional pictures of the electron density distribution in the transistor under low injection are shown in Fig. 12. When the current is increased, the space-charge concentration in the base-collector junction decreases and finally disappears. Spreading of electron flow lines takes place more and more right behind the emitter-base junction dependent upon current and voltage (see Fig. 13). Consequently the $x$ component of the electron current density is not constant in the $x$ direction as it is in a 1D model (apart from the influence of recombination) but deviates appreciably from that value (see Fig. 14). For low current densities the $x$ component of the electron current density along the line $y$ = 0 is nearly constant up to the point where the electrons leave the collector space-charge layer and spread into the ohmic collector region. The current density then decreases because of the current spreading. When $V_{CB}$ = 3 V, very little spreading is possible because the ohmic collector region is very thin. At higher current densities we see from this figure that the current density already decreases behind the emitter-base junction. In this situation the electric field at the collector junction is very small and cannot focus the electrons any longer. The lines of equipotential and of constant hole concen-

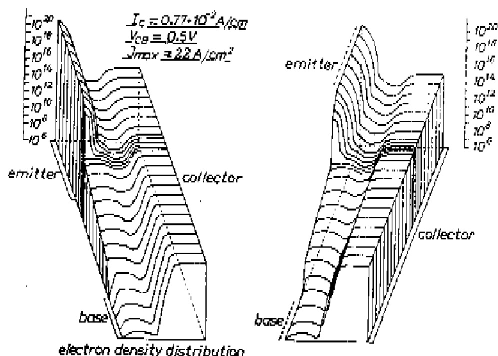



Fig. 12. Three-dimensionally represented electron density distribution seen from different angles. The concentration ($cm^{-3}$) is indicated along the vertical axis.

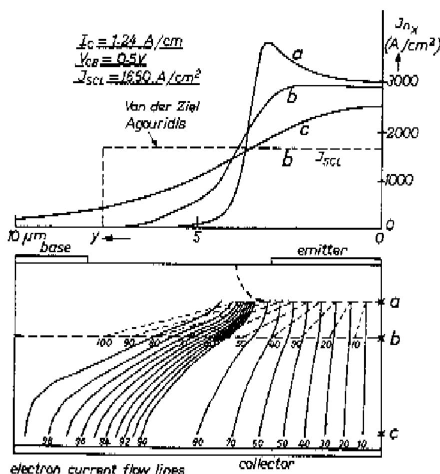



Fig. 13. Calculated electron current flow lines and current density distribution along the emitter junction (curve *a*), collector junction (curve *b*), and along the collector contact (curve *c*) are shown. The flow lines are labeled with the percentage of the total current flowing between this line and the $x$ axis. The corresponding results of the model of van der Ziel and Agouridis [44] are indicated by dashed lines. Curve *a* shows the current crowding along the emitter-base junction.

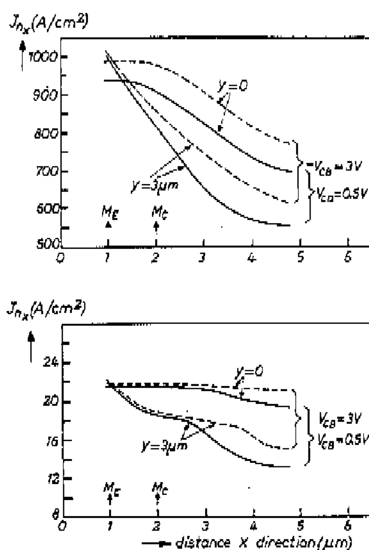



Fig. 14. The variation of the $x$ component of the electron current density as a function of $x$ is shown calculated along two lines, $y = 0$ μm and $y = 3$ μm (See Fig. 2). In the 1D model $J_{n_x}$ would be constant (when recombination is neglected). The metallurgical emitter and collector junctions are indicated by $M_E$ and $M_C$.

tration in Figs. 15 and 16 show how the well-known Kirk effect [18] works out in the lateral dimension.

In Fig. 17 it is shown how strongly the hole and electron concentrations vary along the base-collector junction in the lateral direction under several injection levels.

These rather strong lateral variations under the emitter area, together with the above-shown current and voltage dependent current spreading when the transistor is in saturation, are a severe difficulty in trying to model the transistor with a set of coupled 1D models.

In Fig. 13 current crowding along the emitter-base junction is shown. This crowding under saturation conditions is an effect that has not yet been explored. More numerical calculations are needed to be able to model crowding under these very complicated and nonlinear situations. In this figure we have also indicated how the current spreading would be according to the hypothesis of van der Ziel and Agouridis [44]. Their hypothesis states that there is a maximum current density, the so-called space-charge-limited current density $J_{SCL}$, when the current increases. This $J_{SCL}$ is defined as that current density at which the electric field at the base-collector junction has become zero. This current density is given by [45]

$$J_{SCL} = q v_{\mathrm{lim}} \left( N_D + 2\epsilon/q \, \frac{|V_{CB} + V_0|}{W_C^2} \right). \tag{22}$$

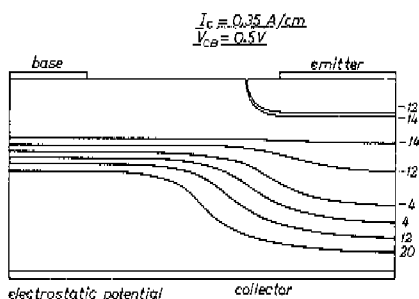



Fig. 15. Calculated lines of constant electrostatic potential. The niveau lines are labeled with the value of the potential normalized in $kT/q$ units.

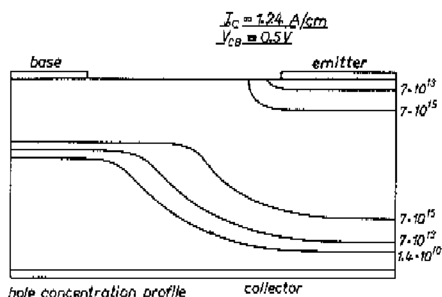



Fig. 16. Calculated lines of constant hole concentration. The niveau lines are labeled with the value of the hole concentration ($cm^{-3}$). The shape of the area where the hole charge is stored in the collector is clearly shown.

"Any further increase in collector current must result in a widening of the region in which collector current flows" [44]. Our 2D calculations, as represented in Fig. 13, show clearly that the current density is not limited by $J_{SCL}$. The current spreading in the base is much less than predicted by these authors, which is a result of their unrealistic assumption. More 2D numerical calculations are needed to construct a well-founded model for the current spreading under saturation conditions. In Fig. 18 the calculated cutoff frequency is given as a function of the current. The points of these characteristics that correspond with the presented and discussed plots of internal current density, potential, and carrier density distributions can easily be found. We have also indicated the corresponding results for the 1D transistor structure (that is, the transistor along the line $y = 0$ in Fig. 2). In order to compare the 1D calculations with the 2D results, the current density in the 1D structure is multiplied by an effective emitter width $L_{eff}$. This follows from the

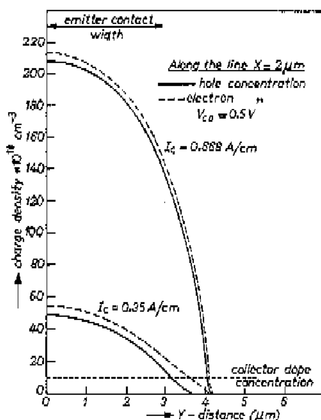



Fig. 17. Calculated hole and electron concentration along the base-collector junction ($x$ = 2 μm; see Fig. 2) as a function of $y$. The width of the emitter contact is indicated.

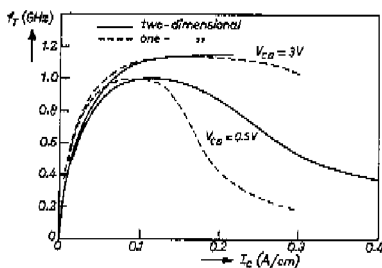



Fig. 18. 1D and 2D calculated cutoff frequency as a function of the current.

consideration that for low injection the integral of the 2D calculated current density along the emitter-base junction must be equal to the product of $L_{eff}$ and the 1D calculated current density (see Fig. 11). For low currents the $f_T$ values are for the 2D structure somewhat smaller than in the 1D structure because of the extra emitter sidewall capacitance. For high current densities the $f_T$ shows a later and slower falloff in the 2D structure than in the 1D structure, due to the current spreading in the 2D transistor. This effect of current spreading on the $f_T$ characteristic is also shown by Slatter [47] in his cylindrical-geometry bipolar transistor structure.

## VIII. CONCLUSION

An efficient and general numerical method for solving the carrier transport equations for 2D semiconductor device structures is presented and demonstrated for a bipolar n-p-n silicon transistor. The described difference approximations for the continuity equations allow rather coarse meshes while still being accurate. Because the coefficient matrices of the difference equations for the three basic partial differential equations (9), (10), and (11) are always positive definite, convergence and stability of the numerical solution of these three equations is guaranteed. The overall convergence was present in all our calculations and behaves very similar to the 1D Gummel method.

With the described method it is possible to analyze important nonlinear effects such as: emitter-base crowding under extreme conditions as high injection in the base and saturation of the base-collector junction; lateral current spreading and charge storage; and how these effects are related to the external electrical behavior.

## APPENDIX A

### DIFFERENCE EQUATIONS FOR THE CONTINUITY EQUATIONS

We will consider the continuity equation for the holes and omit the subscript $p$. The normalized hole continuity equation is

$$\nabla \cdot J = - R, \qquad \text{with } J = - \gamma^{-1} \exp(-\psi) \nabla \phi. \tag{A1}$$

The continuity equation is discretized on the so-called $J$-net. It is assumed that the current density changes slowly per cell of the $J$-net, while the electric field varies slowly per cell of the $E$-net. For simplicity we assume that the $E$-net covers the $J$-net (see Fig. 8).

Using the box integration method [40] and the Gaussian theorem it follows that

$$\begin{aligned}\iint_{ABCD} \nabla \cdot J \, dx \, dy &= \oint J_{\text{normal}} \, ds \\ &= - \iint_{ABCD} R \, dx \, dy.\end{aligned} \tag{A2}$$

Approximation of the integrals gives

$$\begin{aligned}\tfrac{1}{2}(h_k + h_{k-1})(J_M - J_{M-1}) + \tfrac{1}{2}(h_l + h_{l-1})(J_N - J_{N-1}) \\ = - \tfrac{1}{4}(h_k + h_{k-1})(h_l + h_{l-1}) R_{0H,l,k}\end{aligned} \tag{A3}$$

where $R_{eff,l,k}$ is the effective recombination in the rectangle $ABCD$ defined by

$$R_{eff,l,k} = \frac{\sum_{j,i} R_{j,i} h_j h_i}{\frac{1}{4}(h_k + h_{k-1})(h_l + h_{l-1})} . \tag{A4}$$

$R_{j,i}$ is the mean value of the recombination in the four corners of the cell $(j, i)$ (see Fig. 8). Only that part of each cell $(j, i)$ contributes to the summation in the numerator in as far as it lies within the rectangle $ABCD$. $J_M$, $J_{M-1}$, $J_N$, and $J_{N-1}$ are the current density components along the meshlines at points midway between two meshpoints (see Fig. 8) and are assumed to be constant between the two meshpoints. Considering the interval $x_l \leqslant x \leqslant x_{l+1}$ along the line $y = y_k$ it follows that

$$J_x = -\gamma^{-1} \exp(-\psi) d\phi/dx = \text{const } (= J_M). \tag{A5}$$

Instead of approximating this expression by numerical differentiation we first integrate and then approximate the integral. Then it follows that

$$J_M = -\frac{\phi_{l+1} - \phi_l}{\int_{x_l}^{x_{l+1}} \gamma(s) \exp\{\psi(s)\} ds} . \tag{A6}$$

Define a coefficient

$$a_M = \frac{1}{\int_{x_l}^{x_{l+1}} \gamma(s) \exp\{\psi(s) - \psi_{l,k}\} ds} . \tag{A7}$$

Assuming that the electric field between two meshpoints of the $E$-net is constant, we approximate this coefficient as follows:

$$a_M \approx \frac{1}{\sum_j \int_{x_j}^{x_{j+1}} \gamma(s) \exp\{\psi(s) - \psi_{l,k}\} ds}$$

$$= \frac{1}{\sum_j \gamma_{j+\frac{1}{2},i} h_j \exp(\psi_{j+\frac{1}{2},i} - \psi_{l,k}) sh(\Delta_{j+\frac{1}{2},i})} \tag{A8}$$

where

$$sh(x) = \sinh(x)/x$$

$$\Delta_{j+\frac{1}{2},i} = \tfrac{1}{2}(\psi_{j+1,i} - \psi_{j,i})$$

$$\psi_{j+\frac{1}{2},i} = \tfrac{1}{2}(\psi_{j+1,i} + \psi_{j,i}).$$

From (A6) and (A8) it follows that

$$J_M = -a_M \exp(-\psi_{l,k})(\phi_{l+1,k} - \phi_{l,k}), \qquad \text{with } a_M > 0. \tag{A9}$$

Notice that the integral in the denominator of $a_M$ is bounded in magnitude also when the electrostatic potential takes on large values, because we have $\exp(\psi - \psi_{l,k})$ instead of $\exp(\psi)$. In the same way, difference approximations can be derived for $J_{M-1}$, $J_N$ and $J_{N-1}$. Substitution of these approximations in (A3) gives

$$\begin{aligned}
&(a_M + a_{M-1} + ra_N + ra_{N-1}) \exp(-\psi_{l,k})\phi_{l,k} \\
&\quad - a_M \exp(-\psi_{l,k})\phi_{l+1,k} - a_{M-1} \exp(-\psi_{l,k})\phi_{l-1,k} \\
&\quad - ra_N \exp(-\psi_{l,k})\phi_{l,k+1} - ra_{N-1} \exp(-\psi_{l,k})\phi_{l,k-1} \\
&= -\tfrac{1}{2}(h_l + h_{l-1})R_{\mathrm{eff},l,k}
\end{aligned} \tag{A10}$$

with

$$r = (h_l + h_{l-1})/(h_k + h_{k-1}).$$

In matrix notation the set of difference equations can be written as

$$A\phi = b. \tag{A11}$$

Note that when the $E$-net and $J$-net coincide and if the stepsizes are small enough, this difference approximation is identical to the more conventional one [1]. The coefficient matrix $A$ is symmetric with positive diagonal entries and nonpositive off-diagonal entries. Further, this matrix is diagonally dominant and positive definite. This system of difference equations is therefore always stable, independent of the $\psi$ values during the iteration process and can be solved by standard numerical techniques [26].

Because the potentials $\phi_p$ and $\phi_n$ can take on large values, practical difficulties can arise, such as overflow of computer capacity [4], [28]. This can easily be avoided by applying a similarity transformation (see Appendix B) to the set of difference equations (A10)

$$(a_M + a_{M-1} + ra_N + ra_{N-1})p_{l,k} - a_M \exp(\psi_{l+1,k} - \psi_{l,k})p_{l+1,k} - a_{M-1}\exp(\psi_{l-1,k} - \psi_{l,k})p_{l-1,k} - ra_N \exp(\psi_{l,k+1} - \psi_{l,k})p_{l,k+1} - ra_{N-1}\exp(\psi_{l,k-1} - \psi_{l,k})p_{l,k-1} = -\tfrac{1}{2}(h_l + h_{l-1})R_{eff,l,k}. \tag{A12}$$

This final difference equation could have been derived more directly if we would have started with the variable $p$ instead of $\phi_p$. But then, however, it would not be easy to conclude from the resulting coefficient matrix, which then is not symmetric and not diagonally dominant, that the iterative solution methods are convergent. By the derivation as we have shown via the variable $\phi$ we have proven that this coefficient matrix is similar to a positive definite matrix.

## APPENDIX B

## SCALING BY MEANS OF A SIMILARITY TRANSFORMATION

The matrix $B$ is said to be similar to the matrix $A$ if a nonsingular matrix $C$ exists such that $B = C^{-1}AC$. Matrix $B$ is said to be obtained from matrix $A$ by a similarity transformation [46]. In Appendix A we have shown that it is possible to find a difference approximation that results in a diagonally dominant and symmetric positive definite matrix $A$:

$$A\phi = b, \qquad \text{with } A = (a_{i,j}). \tag{B1}$$

When $C$ is defined by the following diagonal matrix

$$C = \begin{bmatrix} \cdot & & & & & 0 & & \\ & \cdot & & & & 0 & & \\ & & \cdot & & & & & \\ & & & \alpha_{i-1} & & & & \\ & & & & \alpha_i & & & \\ & & & & & \alpha_{i+1} & & \\ & & & & & & \cdot & \\ & 0 & & & & & & \cdot \end{bmatrix} \tag{B2}$$

(B1) can be transformed:

$$((CAC^{-1})(C\phi) = (Cb). \tag{B3}$$

Working out these matrix multiplications we have at point $i$

$$(\alpha_i/\alpha_{i-1})a_{i,i-1}(\alpha_{i-1}\phi_{i-1}) + a_{i,i}(\alpha_i\phi_i) + (\alpha_i/\alpha_{i+1})a_{i,i+1}(\alpha_{i+1}\phi_{i+1}) = \alpha_i b_i. \tag{B4}$$

The new matrix $(CAC^{1})$ is generally not symmetric and diagonally dominant. However, the eigenvalue spectrum is invariant for a similarity transformation. The variable $\phi_i$ is transformed to a new variable $(\alpha_i\phi_i)$. The scaling factors $\alpha_i$ can be chosen in such a way that the new variable $(C\Phi)$ varies smoothly. When we use $\exp(-\psi_i)$ as scaling factors we have a difference formulation in the variable $p$. Because the eigenvalue spectrum of a matrix is invariant for a similarity transformation, the coefficient matrix of the difference equations in the variable $p$ behaves in the same way as the matrix of the difference equations in the variable $\phi$.

## APPENDIX C

## THE TIME-DEPENDENT CONTINUITY EQUATIONS

The time-dependent continuity equation for the holes is considered and the subscript $p$ is skipped for simplicity. The corresponding equation for the electrons can be treated in the same way. The time-dependent continuity equation is

$$\nabla \cdot J + \partial/\partial t\{\phi \exp(-\psi)\} = -R. \tag{C1}$$

Using the method of Crank-Nicolson we obtain the following difference equation for the unknown $\phi$ values at the time $t_{n+1}$:

$$\tfrac{1}{2}[A^{(n+1)}\phi^{(n+1)} + A^{(n)}\phi^{(n)}] + h/\Delta t[\phi^{(n+1)}\exp(-\psi^{(n+1)}) - \phi^{(n)}\exp(-\psi^{(n)})] = -{}^2r \tag{C2}$$

or

$$[A^{(n+1)} + (2h/\Delta t)\exp(-\psi^{(n+1)})I]\phi^{(n+1)} = -[A^{(n)} - (2h/\Delta t)\exp(-\psi^{(n)})I]\phi^{(n)} - {}^2r. \tag{C3}$$

Equation (C3) is of the type $A^*\Phi^{(n+1)} = b$. $I$ is the unit matrix. The matrix $A$ is the same as in the steady-state problem (see (A11)) with elements following from (A8) and (A9). The new coefficient matrix $A^*$ consists of the already known matrix $A$ with extra positive diagonal terms added. This extra diagonal dominance results in a larger rate of convergence. Because this matrix is also positive definite the usual overrelaxation methods can be used. The difference approximations used by Reiser [31] lead to a coefficient matrix with complex eigenvalues, and underrelaxation instead of overrelaxation has to be used, which usually results in a smaller rate of convergence [40].

## ACKNOWLEDGEMENT

The author wishes to thank C. Albrecht, H.C. de Graaff, P.A.H. Hart, J. Janse, and C. Weber for helpful discussions, and A.C.M. Kilsdonk for assistance with the programming of the plot procedures.

## REFERENCES

[1] J.W. Slotboom, "Iterative scheme for 1- and 2-dimensional dc-transistor simulation," *Electron. Lett.*, vol. 5, pp. 677-678, Dec. 1969.

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

[3] H.K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculations," *IEEE Trans. Electron Devices*, vol. ED-11, pp. 455-465, Oct. 1964.

[4] A. de Mari, "An accurate numerical steady-state one dimensional solution of the p-n junction," *Solid-State Electron.*, vol. 11, pp. 33-58, 1968.

[5] –, "An accurate numerical one-dimensional solution of the p-n junction under arbitrary transient conditions," *Solid-State Electron.*, vol. 11, pp. 1021-1053, 1968.

[6] D.M. Caughey, "The computer simulation of gigahertz transistors," presented at the IEEE Int. Electron. Conf., Toronto, Ont., Canada, 1967.

[7] –, "Simulation of UHF transistor small signal behavior to 10 GHz for circuit modeling," in *Proc. Cornell Conf. Computerized Electron.*, Aug. 1969, pp. 369-379.

[8] V. Arandjelovic, "General iterative scheme for one dimensional calculations of steady-state electrical properties of transistors," *Int. J. Electron.*, vol. 27, pp. 459-478, 1969.

[9] –, "Accurate numerical steady-state solutions for a diffused one-dimensional junction diode," *Solid-State Electron.*, vol. 13, pp. 865-971, 1970.

[10] B.V. Gokhale, "Numerical solutions for a one-dimensional silicon n-p-n transistor," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 594-602, Aug. 1970.

[11] M. Kurata, "A small-signal calculation for one-dimensional transistors," *IEEE Trans. Electron Devices*, vol. ED-18, pp. 200-210, Mar. 1971.

[12] G.D. Hachtel, R.C. Joy, and J.W. Cooley, "A new efficient one-dimensional analysis program for junction device modeling," *Proc. IEEE*, vol. 60, pp. 86-98, Jan. 1972.
[13] C. Albrecht and D. Dijkstra, to be published.
[14] S.C. Choo, "Numerical analysis of a forward-biased step-junction P-L-N diode," *IEEE Trans. Electron Devices*, vol. ED-18, pp. 574-586, Aug. 1971.
[15] E.D. Graham and J.R. Hauser, "Effects of base doping and width on the J-V characteristics of the n *i-n-p+* structure," *Solid-State Electron.*, vol. 15, pp. 303-310, 1972.
[16] H.N. Gosh, P.H. de la Moneda, and N.R. Dono, "Computer aided transistor design characterization and optimization," *Solid-State Electron.*, vol. 10, pp. 705-726, 1967.
[17] D.J. Roulston, S.G. Chamberlain, and J. Seghal, "High level asymptotic variation of transistor base resistance and current gain," *Electron. Lett.*, vol. 7, pp. 438-440, July 1971.
[18] C.T. Kirk, Jr., "A theory of transistor cutoff frequency ($f_T$) falloff at high current densities," *IRE Trans. Electron Devices*, vol. ED-9, pp. 164-174, Mar. 1962.
[19] H.C. de Graaff, *Solid-State Electron.*, vol. 16, pp. 587-600, 1973.
[20] T.W. Collins, "Two-dimensional numerical analysis of integrated bipolar transistors," presented at the Int. Electron Devices Meeting, Washington, D.C., 1968.
[21] P.U. Calzolari and S. Graffi, "Two-dimensional theory of the uniform base transistor at any injection level," *Alta Freq.*, vol. 38, pp. 126-134, Feb. 1969.
[22] D.P. Kennedy and R.R. O'Brien, "Two-dimensional mathematical analysis of a planar type junction field effect transistor," *IBM J. Res. Develop.*, vol. 13, pp. 662-674, Nov. 1969.
[23] –, "Computer-aided two dimensional analysis of the junction field-effect transistor," *IBM J. Res. Develop.*, vol. 14, pp. 95-116, Mar. 1970.
[24] –, "Two-dimensional analysis of J.F.E.T. structures containing a low-conductivity substrate," *Electron. Lett.*, vol. 7, pp. 714-716, Dec. 1971.
[25] C.K. Kim and E.S. Yang, "An analysis of current saturation mechanism of junction field-effect transistors," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 120-127, Feb. 1970.
[26] R.S. Varga, *Matrix Iterative Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1962.
[27] M. Heydemann, "Méthode numérique d'étude des structures MOST," *Electron. Lett.*, vol. 6, pp. 735-737, Nov. 1970.
[28] D. van Dorpe and N.H. Xuong, "Mathematical 2-dimensional model of semiconductor devices," *Electron. Lett.*, vol. 7, pp. 47-50, Jan. 1971.
[29] J.A. Kilpatrick and W.D. Ryan, "Two-dimensional analysis of lateral-base transistors," *Electron. Lett.*, vol. 7, pp. 226-227, May 1971.
[30] P. Dubock, "D.c. numerical model for arbitrarily biased bipolar transistors in two dimensions," *Electron. Lett.*, vol. 6, pp. 53-55, Feb. 1970.
[31] M. Reiser, "Difference methods for the solution of the time-dependent semiconductor flow equations," *Electron. Lett.*, vol. 7, pp. 353-355, June 1971.
[32] R.W. Hockney, "A fast direct solution of Poisson's equation using Fourier analysis," *J. Ass. Comput. Mach.*, vol. 12, pp. 95-113, 1965.
[33] D. van Dorpe, J. Borel, G. Merckel, and P. Saintot, "An accurate two-dimensional numerical analysis of the MOS transistor," *Solid-State Electron.*, vol. 15, pp. 547-557, 1972.
[34] M. Heydemann, "Solution numérique bidimensionelle des équations générales de transport dans les semiconducteurs en régime permanent," *l'Onde Elec.*, vol. 52, pp. 185-191, Apr. 1972.
[35] H.L. Stone, "Iterative solution of implicit approximations of multidimensional partial differential equations," *SIAM J. Numer. Anal.*, vol. 5, pp. 530-558, Sept. 1968.
[36] H.K. Gummel, "Computer device modeling," presented at the European Sem. Develop. Res. Conf., Munich, Germany, 1969.

[37] D.P. Kennedy and R.R. O'Brien, "Analysis of the impurity atom distribution near the diffusion mask for a planar pn junction," *IBM J. Res. Develop.*, vol. 9, pp. 179-186, 1965.

[38] W. Shockley and W.T. Read, "Statistics of the recombination of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835-842, Sept. 1952.

[39] D.M. Caughey and R.E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE* (lett.), vol. 55, pp. 2192-2193, Dec. 1967.

[40] E.L. Wachspress, *Iterative Solution of Elliptic Systems.* Englewood Cliffs, N.J.: Prentice-Hall, 1966.

[41] H.K. Gummel, private communication.

[42] M.S. Mock, "On the convergence of Gummel's numerical algorithm," *Solid-State Electron.*, vol. 15, pp. 1-4, 1972.

[43] J.W. Slotboom and A.C.M. Kilsdonk, to be published.

[44] A. van der Ziel and D. Agouridis, "The cutoff frequency falloff in UHF transistors at high currents," *Proc. IEEE*, vol. 54, pp. 411-412, Mar. 1966.

[45] J.L. Moll, *Physics of Semiconductors.* New York: McGraw-Hill, 1964, p. 155.

[46] V.N. Faddeeva, *Computational Methods of Linear Algebra.* New York: Dover, 1958.

[47] J.A.G. Slatter, "Fundamental modeling of cylindrical geometry bipolar transistors," *Electron. Lett.*, vol. 8, pp. 222-223, May 1972.



*Reprinted from* IEEE TRANSACTIONS
ON *ELECTRON DEVICES*
Volume ED-20, Number 8, August, 1973
pp. 669-679


COPYRIGHT © 1973—THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.
PRINTED IN THE U.S.A.

## 5.3. "Measurements of bandgap-narrowing in Si bipolar transistors"

Solid-State Electr., Vol. 19, p. 857, 1976.

# MEASUREMENTS OF BANDGAP NARROWING IN Si BIPOLAR TRANSISTORS

J. W. SLOTBOOM and H. C. DE GRAAFF
Philips Research Laboratories, Eindhoven, The Netherlands

(*Received* 29 *January* 1976; *in revised form* 22 *March* 1976)

**Abstract**—Theory predicts appreciable bandgap narrowing in silicon for impurity concentrations greater than about $10^{17}$ cm$^{-3}$. This effect influences strongly the electrical behaviour of silicon devices, particularly the minority carrier charge storage and the minority carrier current flow in heavily doped regions. The few experimental data known are from optical absorption measurements on uniformly doped silicon samples. New experiments in order to determine the bandgap in silicon are described here. The bipolar transistor itself is used as the vehicle for measuring the bandgap in the base. Results giving the bandgap narrowing ($\Delta V_{g0}$) as a function of the impurity concentration ($N$) in the base (in the range of $4.10^{17}$–$2.5\,10^{19}$ cm$^{-3}$) are discussed. The experimental values of $\Delta V_{g0}$ as a function of $N$ can be fitted by:

$$\Delta V_{g0} = V_1\left(\ln\frac{N}{N_0} + \sqrt{\left(\ln^2\frac{N}{N_0} + C\right)}\right)$$

where $V_1$, $N_0$ and $C$ are constants.

It is also shown how the effective intrinsic carrier concentration ($n_{ie}$) is related with the bandgap narrowing ($\Delta V_{g0}$).

## 1. INTRODUCTION

The phenomenon of bandgap narrowing in heavily doped germanium and silicon has been studied theoretically and experimentally because of the interest in tunnel diodes [1–4].

Theoretical work by Kane[4], Morgan[5] and Bonch-Bruyevich[6] shows that at high impurity concentrations the density of energy states no longer has a parabolic energy distribution and becomes dependent on the impurity concentration. The bandgap is effectively reduced due to the broadening of the impurity band and the formation of band tails on the edges of the conduction and valence band.

Vol'fson and Subashiev[7] in 1967 investigated experimentally the fundamental absorption edge of silicon and found only changes in the bandgap for impurity concentrations above $10^{19}$ cm$^{-3}$; the greatest narrowing measured was 0.068 eV for n-type silicon with $N = 9.10^{19}$ cm$^{-3}$.

Referring to these experiments Kauffman and Bergh[8] in 1968 used the difference in effective energy gap of the emitter and base region in bipolar transistors to explain the discrepancy between theoretically predicted injection efficiency and experimentally observed values and, like Buhanan[9] in 1969, they used this difference in bandgap to explain the temperature dependence of the current gain as well. De Man[10] in 1971 calculated the influence of the position-dependent bandgap in the emitter region on the injected minority carrier current. For the impurity-dependent bandgap he also used the experimental values of Vol'fson and Subashiev. On the basis of the theories of Kane, Morgan and Bonch-Bruyevich, Kleppinger and Lindholm[11] developed a general relationship between the impurity concentration and the density of states for the whole range of impurity concentrations of practical importance. These general formulations of the density of states were used by Van Overstraeten, de Man and Mertens[12] in 1973 in calculating the manner in which the effective intrinsic carrier concentration ($n_{ie}$), defined by

$$n_{ie}^2 = p_0 n_0 \qquad (1)$$

(where $p_0$ and $n_0$ are the equilibrium carrier concentrations), depends on the impurity concentration; they gave generalized transport equations, taking into account the influence of the position-dependent bandgap. In another paper[13] they took these results as the basis for calculations of emitter efficiency of bipolar transistors.

Also in 1973 Mock[14] presented very similar work, but he found quantitatively different $n_{ie}$-values, particularly for the higher dope concentrations above $4.10^{19}$ cm$^{-3}$. One of the main reasons for this was his use of a more general formulation for the screening length. The results of Heasell[15] in 1975 agree with the results of Mock. The results of these calculations are shown in Fig. 1, together with the experimental values reported by Vol'fson and Subashiev. In later work Van Overstraeten *et al.*[16–18] and Mock[19] have shown how bandgap narrowing influences both the electrical and the optical behaviour of silicon devices. However because the theory contains some not well-known parameters, e.g. the screening length, all this work yields quantitatively uncertain results for the theoretically calculated $n_{ie}$-values. This is illustrated in Fig. 2, where current gain values are shown which we have calculated by solving numerically the transport equations, taking high-doping effects into account in several different ways. This stresses the necessity of determining accurately the effective intrinsic concentration as a function of the impurity concentration by means of experiments. In this paper we will discuss in detail experiments used to measure the effective intrinsic

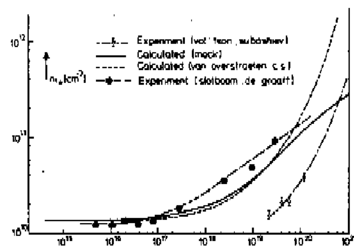



Fig. 1. The effective intrinsic concentration $n_{ie}$ as a function of the impurity concentration $N$, according to different theories. Experimental values from optical absorption (☿)[7] and from this paper (●) are shown.

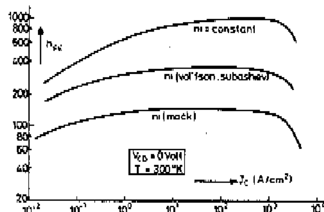



Fig. 2. Calculated current gain ($h_{FE}$) vs collector current density ($J_c$), showing the influence of the intrinsic concentration values on device characteristics.

concentration $n_{ie}$ as a function of impurity concentration and temperature and its relationship with the bandgap narrowing $\Delta V_{g0}$. We also give an empirical expression which fits our experimental values. Preliminary results of this experimental investigation have already been presented elsewhere[20, 21].

## 2. THEORETICAL MODEL AND MEASURING METHODS

In the experiments we focussed our attention on the well-known $I_C - V_{EB}$ characteristic of NPN bipolar transistors. During all our measurements care was taken to keep the injection in the base very low. Assuming that the minority carriers in the base obey the Boltzmann distribution law, it can be shown that the $I_C - V_{EB}$ characteristic is given by:

$$I_C = I_0 \exp(qV_{EB}/kT) \tag{2}$$

where

$$I_0 = A_E \frac{kT\mu_n n_{ie}^2}{Q_B}.$$

In this expression: $A_E$ = emitter area, $\mu_n$ = electron mobility in the $p$-type base, $Q_B$ = total number of holes/cm$^2$, $n_{ie}^2 = pn$-product in equilibrium in the base. (When the impurity concentration in the base is not constant, $\mu_n$ and $n_{ie}^2$ are mean values over the base region weighted by the hole concentration). Equation (2) makes it clear that, apart from the product ($\mu_n n_{ie}^2$), all the other quantities are known from geometry and straightforward measurements.

For low impurity concentrations the intrinsic carrier concentration is given by:

$$n_{ie}^2 = n_{i0}^2 = \text{const.}\ T^3 \exp(-qV_g/kT). \tag{3}$$

The intrinsic concentration $n_{i0}$ is dependent only on temperature. The bandgap $V_g$ is also a function of temperature. From optical experiments of MacFarlane[22] (see Fig. 3) we know that above a certain temperature ($T > T_0$) the bandgap can be approximated by a linear function of $T$, $V_g = V_{g0} - \alpha T$.

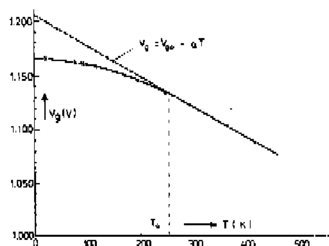



Fig. 3. The bandgap ($V_g$) vs temperature ($T$) in very pure silicon[22]. $T_0 = 250$ K and $V_{g0} = 1.206$ V.

Substitution in eqn (3) gives:

$$n_{i0}^2 = CT^3 \exp(-qV_{g0}/kT). \tag{4}$$

Putley and Mitchell[23] have found experimentally that the constants are $C = 9.6 \times 10^{32}$ and $V_{g0} = 1.206$ V.

For high impurity concentrations we have calculated numerically the effective intrinsic concentration ($n_{ie}$) in a similar way as in [12, 14, 15], not only as a function of impurity concentration, but also as a function of temperature[24]. These calculations show that not only for low but for all impurity concentrations $n_{ie}$ can be described by:

$$n_{ie}^2 = CT^3 \exp(-qV_{g0}(N)/kT) \tag{5}$$

for $T > T_0$.

The extrapolated bandgap value, $V_{g0}(N)$, is a function of the impurity concentration ($N$), but $C$ is approximately independent of the impurity concentration. If we define the bandgap narrowing $\Delta V_{g0}$ by

$$\Delta V_{g0}(N) = 1.206 - V_{g0}(N), \tag{6}$$

the effective intrinsic carrier concentration can be written as:

$$n_{ie}^2(N, T) = n_{i0}^2(T) \exp(q\Delta V_{g0}(N)/kT). \tag{7}$$

We will now describe two methods for determining $\Delta V_{g0}$ which are based on eqn (2).

*Method* 1

From eqn (2) it follows that

$$n_{ie}^2 = I_0 \frac{Q_B}{A_E k T \mu_n}. \quad (8)$$

At room temperature the $I_C - V_{EB}$ characteristic and the sheet resistance underneath the emitter ($R_{\square base}$) are measured. From the results $I_0$ and $Q_B$ can easily be derived. For the electron mobility the majority carrier value in $n$-type material with the same impurity concentration as in the base is taken (see Section 4). In this way $n_{ie}^2$ is derived and $\Delta V_{g0}$ is found from eqn (7) for a number of transistors with varying base dope concentrations.

*Method* 2

If we substitute the general form for the intrinsic concentration (eqn 3) in eqn (2) we obtain:

$$I_C = \text{const. } T^a \mu_n(T) \exp(q(V_{EB}(T) - V_g(T))/kT). \quad (9)$$

Here it is assumed that $Q_B$ is constant (see Section 4). Differentiation with respect to the temperature $T$ while $I_C$ is kept constant gives:

$$V_g - T\frac{dV_g}{dT} = \left(V_{EB} - T\frac{dV_{EB}}{dT}\right) - \frac{kT}{q}\left(4 + T\frac{d}{dT}(\ln \mu_n)\right). \quad (10)$$

With $I_C$ constant, $V_{EB}$ is measured very accurately for a series of temperatures in the range 150–420 K, increasing in small temperature steps. The temperature is measured with a calibrated thermocouple and a check is made on the accuracy of the ideal exponential slope of the $I_C - V_{EB}$ characteristic. It is found that over the whole temperature range the temperature derived from this slope always differs less than 0.5 K as compared to the temperature measured with the thermocouple. From the measured $V_{EB}$ values the term $(V_{EB} - T(dV_{EB}/dT))$ of eqn (10) is derived as a function of temperature. In the same way we measure the sheet resistance of the base ($R_{\square base}$) as a function of temperature. This gives us the temperature dependence of the hole mobility in this base. Knowing[28] that the temperature dependence of the mobility of electrons and holes, when they are majority carriers, is very similar and using the same assumption for the minority carrier mobility as in method 1 (see also the discussion in Section 4), we assume:

$$\frac{d}{dT}(\ln \mu_n) = \frac{d}{dT}(\ln \mu_p). \quad (11)$$

In this way the term $kT/q(4 + T(d/dT)(\ln \mu_n))$ in eqn (10) follows from the temperature measurement of the $R_{\square base}$. By subtracting the two terms in eqn (10) which resulted from the transistor and the $R_{\square base}$ measurements, we find how $(V_g - T(dV_g/dT))$ changes with temperature. In agreement with theoretical calculations[24] it appears that this function has the constant value $V_{g0}$ above a certain temperature value $T_0$. The bandgap narrowing, $\Delta V_{g0}$, is found with the aid of eqn (6).

## 3. EXPERIMENTAL RESULTS

The measuring procedures and the analysis of the results will be illustrated for two different cases: a transistor with a lightly doped base and one with a heavily doped base. We consider first a transistor with a base doping of $N = 4.10^{16}$ cm$^{-3}$. Using method 1 and substitution of $n_{ie}^2$ in eqn (7) we find $\Delta V_{g0} = 0$.

In Fig. 4 the results of method 2 are shown. The term $(V_{EB} - T\Delta V_{EB}/\Delta T)$ is corrected by the term resulting from the temperature dependence of the electron mobility. The variation of the hole mobility in the base with temperature, constructed from the measurements on the $R_{\square base}$ is shown in Fig. 5. The temperature dependence

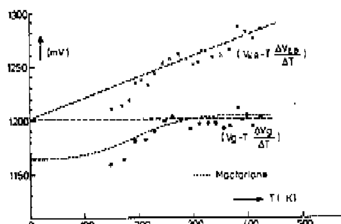



Fig. 4. The quantities $(V_{EB} - T(\Delta V_{EB}/\Delta T))$ and $(V_g - T(\Delta V_g/\Delta T))$ vs temperature. Base doping concentration $N = 4.10^{16}$ cm$^{-3}$.

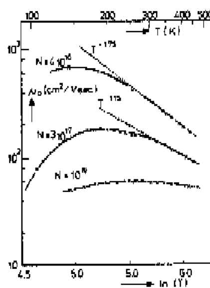



Fig. 5. Majority carrier mobility ($\mu_p$) in the base vs temperature ($T$), for various base dopings. The curves were obtained from $R_{\square base}$ measurements.

of the term $(V_g - T(\Delta V_g/\Delta T))$ agrees very well with the values for this term derived from MacFarlane's measurements of the temperature dependence of the bandgap (see Fig. 3). For temperatures higher than about 250 K this term has a constant value $V_{g0}$ of about 1200 mV due to the linear temperature dependence. In conclusion we can say that the silicon bandgap measured in the base region of a bipolar transistor with a base dope of $4.10^{16}$ cm$^{-3}$ is not narrowed and agrees very well with MacFarlane's measurements.

Next we consider a transistor with a large impurity concentration in the base ($N = 10^{19}$ cm$^{-3}$).

Table 1.

| $N$ (base) (cm$^{-3}$) | $I_0$ (A) | $Q_B$ (cm$^{-2}$) | $A_E$ (cm$^{2}$) | $T$ (K) | $\mu_n$ (cm$^2$/V sec) | [illegible] | [illegible] | $\Delta V_{g0}$ (mV) |
|---|---|---|---|---|---|---|---|---|
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0 |
| $1.10^{19}$ | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 74 |

In Table 1 the measured data and the results of method 1 are given. The bandgap narrowing $\Delta V_{g0}$ is 74 mV. The results of method 2 are presented in Fig. 6. The term $(V_{EB} - T(\Delta V_{EB}/\Delta T))$ is seen to be a linear function of the temperature, with a slope of $4\,kT/q$. In Fig. 5 we see that the mobility is almost independent of temperature. According to eqn (10) this results in a nearly constant value for $(V_g - T(dV_g/dT))$ over the whole temperature interval measured with $V_{g0} = 1130$ mV. We see that both methods agree and give a bandgap narrowing $\Delta V_{g0}$ of approximately 75 mV.

Finally, Fig. 7 gives a survey of similar measurements on a number of transistors having base dope concentrations from $4\times10^{15}$ to $2.5\times10^{19}$ cm$^{-3}$. The experimental values of Vol'fson and Subashiev[7], also indicated in this figure, are seen to differ appreciably from ours (for a discussion see Section 4). Our experimental values are also indicated in Fig. 1 and are in good agreement with theoretical $n_{ie}$ values. Apart from the uncertainty in the mobility for minority carriers (see Section 4), the estimated accuracy in the determination of $\Delta V_{g0}$ is $\pm 10$ mV.

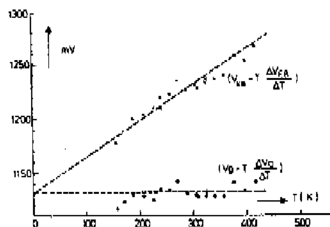



Fig. 6. The quantities $(V_{EB} - T(\Delta V_{EB}/\Delta T))$ and $(V_g - T(\Delta V_g/\Delta T))$ vs temperature. Base doping concentration $N = 1.10^{19}$ cm$^{-3}$.

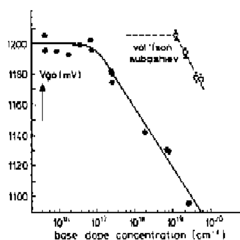



Fig. 7. The measured extrapolated bandgap value $V_{g0}$ as a function of the impurity concentration $N$. The fully drawn line represents $V_{g0}$ according to eqn (18) (● measured according to method 2 and ⊗ measured according to method 1).

## 4. DISCUSSION

As mentioned, the experimental investigation was concentrated on the $I_C - V_{EB}$ characteristic (eqn (2)) and in fact we measured the product $(\mu_n n_{ie}^2)$ as a function of temperature. Method 1 measures $(\mu_n n_{ie}^2)$ at room temperature and method 2 its variation with temperature.

From eqn (2) it follows that for $I_C$ constant:

$$\frac{d}{dT}(\ln\,(T\mu_n n_{ie}^2)) = \frac{q}{kT^2}\left(V_{EB} - T\frac{dV_{EB}}{dT}\right). \qquad (12)$$

Several authors[8, 9, 31, 32] have described the temperature dependence of $I_0$, which is identical to that for $(T\mu_n n_{ie}^2)$, in terms of an activation energy $E_a$:

$$T\mu_n n_{ie}^2 = \text{const.}\exp(-E_a/kT) \qquad (13)$$

However, this would mean that $(V_{EB} - T(dV_{EB}/dT))$ is constant and this is not in agreement with our experiments (see e.g. Fig. 4 and Fig. 6). Our experiments indicate that above the temperature $T_0$ this term is a linear function of temperature:

$$V_{EB} - T\frac{dV_{EB}}{dT} = V_{g0} + \beta\frac{kT}{q}, \quad (T > T_0). \qquad (14)$$

Consequently our experiments together with eqn (12) indicate that the product $(\mu_n n_{ie}^2)$ must be described by

$$\mu_n n_{ie}^2 = \text{const.}\,T^{\beta-1}\exp(-qV_{g0}/kT), \qquad (15)$$

where $\beta$ and $V_{g0}$ follow from the measured term $(V_{EB} - T(dV_{EB}/dT))$. In Table 2 these parameters are given for three transistors with different base dope concentrations.

In order to be able to determine $n_{ie}^2$ itself as a function of temperature it is necessary to have more data about the electron mobility in the base. However, in contrast to majority carrier mobilities, very little is known about the minority mobility. The majority carrier mobility is usually derived from resistivity measurements assuming that the carrier concentration is equal to the chemically determined concentration of impurity atoms[25–27]. The temperature dependence of the majority carrier mobility for holes and electrons are approximately equal[28] and can be described by:

$$\mu \propto T^{-n} \qquad (16)$$

Table 2.

| N (Base) (cm$^{-3}$) | $\beta$ | $V_{g0}$ (mV) | $\eta$ | C | [illegible] |
|---|---|---|---|---|---|
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] |

where $\eta$ is constant above a certain temperature (see also Fig. 5).

For low impurity concentrations ($N < 10^{17}\,cm^{-3}$) it is known from experiments[29, 30], that the electron mobility and its temperature dependence are the same in $n$-type and $p$-type silicon. Although for high impurity concentrations there is no such experimental evidence we assume that this form of relationship is still true. Substitution of eqn (16) in eqn (15) gives:

$$n_{ie}^2 = CT^{(\beta-1+\eta)}\exp(-qV_{g0}/kT). \qquad (17)$$

Table 2 also gives values for $C$ and $\eta$, where $\eta$ was derived from the measured temperature dependence of the $R_{\square Base}$ and $C$ was obtained from an $I_b$ measurement at room temperature. It can be seen that at all impurity levels $C \approx 9\times10^{32}$ (the value found by Putley and Mitchell[23] for intrinsic silicon) and that $\beta - 1 + \eta \approx 3$.

This means that the well-known Putley and Mitchell formula for $n_{ie}^2$ can be used for any impurity concentration ($N$), provided that $V_{g0}$ is taken as a function of $N$.

Theoretical calculations[24] confirm this remarkable result and therefore strongly indicate that the assumption concerning the equality of the electron mobility and its temperature dependence in $n$- and $p$-type silicon is justified.

Using the majority carrier mobility values as discussed above, $Q_B$ is the net sum of chemical impurity atoms in the base region and therefore independent of temperature. In all our experiments transistors with rather thick bases (varying from 1 to 5 $\mu$m) were used and the influence of $V_{EB}$ on $Q_B$ can be neglected. For these reasons $Q_B$ is taken as a constant in our experiments. The results of our measurements differ appreciably from the bandgap values obtained from optical absorption experiments (Vol'fson and Subashiev[7], see Figs. 1 and 7). This discrepancy is probably due to the fact that optical absorption involves the energy gap between two free levels, which can be larger than the bandgap that determines the $pn$-product. Regarding the origins of bandgap narrowing it has been suggested[31] that the bandgap is greatly reduced by mechanical stress due to the misfit of the impurity atoms. However, from X-ray measurements[33] of changes in lattice constant in heavily doped silicon samples a $\Delta V_{g0}$ of only 5–7 mV is predicted[34] in the worst case.

Long range impurity density fluctuations (over hundreds of atomic distances) can cause potential fluctuations which increase the averaged $pn$-product without a change in bandgap. However for this hypothesis no quantitative justification is known at present. It seems therefore, as if the formation of an impurity band and bandtails on the edges of the conduction and valence band based on the theories of Kane, Morgan and Bonch–Bruyevich is the most reasonable explanation for the measured bandgap narrowing.

## CONCLUSION

Bandgap narrowing in silicon has been determined in the base region of bipolar transistors. At impurity concentrations greater than about $10^{17}\,cm^{-3}$ this effect becomes important. The experimentally derived bandgap narrowing as a function of impurity concentration can be fitted by the following empirical formula:

$$\Delta V_{g0}(N) = V_1(F + \sqrt{(F^2 + C)})[mV] \qquad (18)$$

where $V_1 = 9[mV]$

$F = \ln(N/N_0)$

$N$ = impurity concentration [cm$^{-3}$]

$N_0 = 10^{17}$[cm$^{-3}$]

$C = 0.5$

It is shown that the temperature dependence of the effective intrinsic carrier concentration is described by the equation:

$$n_{ie}^2(N, T) = n_{i0}^2(T)\exp(q\Delta V_{g0}(N)/kT) \qquad (19)$$

where $T > T_0 \approx 250$ K.

*Acknowledgements*—The authors are grateful to A. Schmitz, H. G. R. Maas, P. J. W. Jochems and J. W. A. Scholte for making the devices which were used in our experiments and wish to thank also A. K. Jansen and D. Polder for helpful discussions.

## REFERENCES

1. J. I. Pankove, *Phys. Rev. Lett.* 4, 20 (1960).
2. M. Cardona and M. S. Sommers, *Phys. Rev.* 122, 1382 (1961).
3. C. Haas, *Phys. Rev.* 125, 1965 (1962).
4. E. O. Kane, *Phys. Rev.* 131, 79 (1963).
5. T. N. Morgan, *Phys. Rev.* 139, A343 (1965).
6. V. L. Bonch-Bruyevich, *The Electronic Theory of Heavily Doped Semiconductors*. Elsevier, Amsterdam (1966).
7. A. A. Vol'fson and V. K. Subashiev, *Sov. Physics-Semiconductors* 1, 327 (1967).
8. W. L. Kauffman and A. A. Bergh, *IEEE Transact. on El. Dev.* ED-15, 732 (1968).
9. D. Buhanan, *IEEE Transact. on El. Dev.* ED-16, 117 (1969).
10. H. J. J. De Man, *IEEE Transact. on El. Dev.* ED-18, 833 (1971).
11. D. D. Kleppinger and F. A. Lindholm, *Solid-St. Electron.* 14, 199–206 (1971).
12. R. J. Van Overstraeten, H. J. J. De Man and R. P. Mertens, *IEEE Transact. on El. Dev.* ED-20, 290 (1973).
13. R. P. Mertens, H. J. De Man and R. J. Van Overstraeten, *IEEE Transact. on El. Dev.* ED-20, 772 (1973).

14. M. S. Mock, *Solid-St. Electron.* **16**, 1251–1259 (1973).
15. E. L. Heasell, *Internat. J. Electronics* **38**, 127–135 (1975).
16. H. De Man, R. Mertens and R. Van Overstraeten, *Electron. Lett.* **9**, 172 (1973).
17. R. P. Mertens, R. Van Overstraeten, H. De Man, F. A. Lindholm and D. P. Kennedy, *Intern. Electr. Dev. Meeting*, Washington (1974).
18. R. J. Van Overstraeten and R. Mertens, *European Solid-State Device Res. Conf.*, Grenoble (1975).
19. M. S. Mock, *Solid-St. Electron.* **17**, 819 (1974).
20. J. W. Slotboom and H. C. de Graaff, *European Solid-State Device Res. Conf.*, Grenoble (1975).
21. J. W. Slotboom and H. C. de Graaff, *Intern. Electr. Dev. Meeting*, Washington (1975).
22. G. G. MacFarlane, J. P. McLean, J. E. Quarrington and V. Roberts, *Phys. Rev.* **111**, 1245 (1958).
23. E. H. Putley and W. H. Mitchell, *Proc. Phys. Soc. London* **A72**, 193 (1958).
24. J. W. Slotboom, *Solid-St. Electron.* Submitted for publication.
25. J. C. Irvin, *Bell. Syst. Techn. J.* **41**, 387 (1962).
26. G. Baccarani and P. Ostoja, *Solid-St. Electron.* **18**, 579 (1975).
27. G. W. Ludwig and R. L. Watters, *Phys. Rev.* **101**, 1699 (1956).
28. W. W. Gärtner, *Transistors, Principles, Design and Applications*, Van Nostrand, Princeton (1960).
29. G. W. Ludwig and R. L. Watters, *Phys. Rev.* **101**, 1699 (1956).
30. E. M. Conwell, *Proc. IRE* **46**, 1281 (1958).
31. P. J. Kannam, *IEEE Trans.* **ED-20**, 845 (1973).
32. R. B. Fair, *IEEE Trans.* **ED-20**, 642 (1973).
33. B. G. Cohen, *Solid-St. Electron.* **10**, 33 (1967).
34. V. I. Fistul, *Heavily Doped Semiconductors*, p. 33–34. Plenum Press, New York (1969).

## 5.4. "The pn-product in silicon"

Solid-State Electr., Vol. 20, p. 279, 1977.

# THE $pn$–PRODUCT IN SILICON

J. W. SLOTBOOM
Philips Research Laboratories, Eindhoven, The Netherlands

(*Received* 8 *June* 1976; *in revised form* 9 *August* 1976)

**Abstract**—Optical absorption measurements[1] and theoretical calculations[2–9] have shown that the bandgap in silicon is not only temperature-dependent but is also influenced by the impurity concentration at higher values. Recent electrical measurements[10] of the $pn$-product in the base region of bipolar transistors made it possible to derive the bandgap narrowing quantitatively as a function of the impurity concentration.

In this paper it will be shown that theoretical calculation of the $pn$-product as a function of temperature and impurity concentration can be approximated by the following relationship:

$$pn = n_{ie}^2(N, T) = CT^3 \exp(-qV_{go}(N)/kT)$$

for temperatures between about 280 and 450°K.

Moreover the calculated values for $C$ and $V_{go}(N)$ show surprisingly good quantitative agreement with the values derived from the above mentioned $pn$–measurements[10].

### 1. INTRODUCTION

At low impurity concentrations the $pn$–product is independent of the impurity concentration and is theoretically given by[11]:

$$pn = n_{io}^2(T) = CT^3 \exp(-qV_{go}/kT) \qquad (1)$$

where $V_{go}$ is the towards zero degree Kelvin extrapolated bandgap. This formula has been verified experimentally by Putley and Mitchell[12]. In agreement with the optical bandgap measurements of MacFarlane[13] they found $V_{go} = 1.206$ eV and for the constant $C$ the value of $9.61 \times 10^{32}(\mathrm{cm}^{-6}\,\mathrm{K}^{-3})$ which can be explained theoretically[11].

At higher impurity concentrations the $pn$–product becomes dope-dependent. The influence of the impurity concentration has recently been derived[10] from measurements of the $pn$–product in the base region of a number of bipolar NPN transistors with base dope concentrations varying from $10^{17}$–$2 \times 10^{19}$ cm$^{-3}$. These measurements show that above a temperature of about 250 K the $pn$–product is described by the well-known formula (1) provided that $V_{go}$ is taken as a function of the impurity concentration ($N$) only. For impurity concentrations above about $10^{17}$ cm$^{-3}$ $V_{go}(N)$ starts to deviate from 1.206 eV.

In this paper theoretical calculations of the $pn$–product as a function of temperature (280–450°K) and impurity concentration ($10^{15}$–$3 \times 10^{19}$ cm$^{-3}$) will be discussed. It will be shown that these calculations confirm the above mentioned experimental results and the calculated values for $C$ and $V_{go}(N)$ agree well with the measured values.

Further it will be demonstrated that, in analogy to the bandgap definition at low impurity concentrations, $V_{go}(N)$ can be considered as the impurity concentration-dependent bandgap extrapolated towards zero degree Kelvin.

The accuracy of the derived bandgap narrowing in [10], $\Delta V_{go}$, was estimated as ± 10 mV, apart from the uncertainty about the assumption that the electron mobility in the $p$-type base is the same as if the base were $n$-type.

In the calculations we will consider uncompensated $p$-type silicon. Most of the measured transistor structures had uncompensated boron doped base regions (epitaxial bases). In the double diffused transistor structures the phosphorous concentration in the base was always less than about 20% of the boron concentration.

### 2. THEORETICAL CALCULATIONS

The calculations are based on the density of energy states functions resulting from the work of Kane[2], Morgan[3] and Bonch–Bruyevich[4]. We shall only mention the equations used, without discussion. For more details about the physics and possible uncertainties in the theory we refer to the work of these authors themselves and to a number of authors who have used these equations for similar calculations[6–9]. In this paper we use more particularly the formulation as given by Mock[8]. In the equations rationalized MKS-units will be used.

The impurity concentration-dependent density of states functions are formulated as perturbations on the density of states functions for intrinsic silicon. The energy is chosen zero in the middle of the intrinsic bandgap and is defined positive towards the conduction band for electrons and towards the valence band for holes. Apart from a small correction term (see Section 3), the temperature dependent intrinsic bandgap, $E_g(T)$, is assumed to be given by the values measured by MacFarlane[13].

The impurity band density of states is given by Morgan[3] as:

$$\rho_i(E) = 2N(2\pi\sigma_i^2)^{-1/2} \exp\left(-\frac{(E - E_A)^2}{2\sigma_i^2}\right) \qquad (2a)$$

where $E_A$ is the acceptor energy level and $\sigma_i$ is the

effective standard deviation of the impurity band, given by:

$$\sigma_t = 1.03\sigma \exp\left(-(11.3806\pi N)^{-1/2}\lambda^{-3/2}\right) \tag{2b}$$

$$\sigma = q^2\left(\frac{\lambda N}{8\pi\epsilon^2}\right)^{1/2}. \tag{2c}$$

The screening length $\lambda$ is a critical parameter in the theory and is given by the general formula[14]:

$$\lambda^{-2} = \frac{q^2}{\epsilon}\left(\left|\frac{\partial n}{\partial F}\right| + \left|\frac{\partial p}{\partial F}\right|\right). \tag{2d}$$

Here $F$ is the Fermi level. For non-degenerate material, when Boltzmann statistics can be applied, this formula reduces to the well-known Debye length ($\lambda_D \propto N^{-1/2}$) while for strongly degenerate material it can be shown[17] that $\lambda \propto N^{-1/6}$ assuming Fermi–Dirac statistics and parabolic energy distribution of the density of states.

According to the theory of Kane and Bonch–Bruyevich the density of states function for the valence band is:

$$\rho_V(E) = m_h^{*3/2}(2^{1/2}\sigma)^{1/2}\pi^{-2}\hbar^{-3}\, y\left(\frac{E - \frac{E_g(T)}{2}}{\sigma\sqrt{2}}\right) \tag{3a}$$

where $m_h^*$ is the hole effective mass[11] and $\sigma$ is the standard deviation given in eqn (2c).

$$y(x) = \pi^{-1/2}\int_{-\infty}^{x}\sqrt{(x-u)}\exp(-u^2)\,\mathrm{d}u. \tag{3b}$$

This integral can be approximated[15] as follows:

$$y(x) = x^{1/2}\left(1 - \frac{1}{16x^2}\right) \quad \text{for} \quad x \geq 0.601 \tag{3c}$$

representing the parabolic energy distribution in the band itself, while in the band tails the following approximation holds:

$$y(x) = \tfrac{1}{2}\pi^{-1/2}\exp(-x^2)\{1.225 - 0.906(1-\exp(2x))\} \quad \text{for} \quad x \leq 0.601. \tag{3d}$$

The density of states function for the conduction band (minority carriers) can be described by the same equations with replacing $m_h^*$ by the electron effective mass $m_e^*$. However, Bonch–Bruyevich[4] has pointed out that in this case the deeper lying energy levels, given by eqn (3d), are not present because of the repelling force between the electrons and acceptor ions[16]. He suggests that only the first-order term in the series expansion of eqn (3d) should be taken into account:

$$y(x) = \tfrac{1}{2}\pi^{-1/2}(1.225 + 1.812x) \quad \text{for} \quad -0.676 \leq x < 0.601. \tag{3e}$$

The density of energy states of the conduction band is then described by eqns (3a)–(3e). At low impurity concentrations the density of states functions reduce to the usual approximation of a discrete acceptor energy level and parabolic energy distribution for the valence and conduction band. For higher impurity concentrations the impurity band broadens and band tails appear at the conduction and valence band edges. The total hole and electron concentrations are:

$$p = \int_{-\infty}^{+\infty}\frac{\max(\rho_I(E), \rho_V(E))}{1+\exp\left(\frac{E-F}{kT}\right)}\,\mathrm{d}E \tag{4}$$

$$n = \int_{-\infty}^{+\infty}\frac{\rho_C(E)}{1+\exp\left(\frac{E+F}{kT}\right)}\,\mathrm{d}E. \tag{5}$$

Further, charge neutrality is assumed:

$$p - n - N = 0. \tag{6}$$

The sets of eqns (2)–(6) are solved by means of a Newton-iteration scheme until a consistent solution is reached.

### 3. RESULTS

The intrinsic bandgap $E_g(T)$, used in the equations given in Section 2, is chosen in such a way that the bandgap for $N = 10^{15}\,\mathrm{cm}^{-3}$ as a function of temperature equals MacFarlane's measured bandgap values. From the calculated standard deviation $\sigma$ in Fig. 2 it follows that even for this low impurity concentration of $10^{15}\,\mathrm{cm}^{-3}$ the edges of the energy bands are not exactly abrupt. We have defined these edges as the energies above which 90% of the carriers are located. This definition is not very critical and has been chosen such that the calculated $pn$–products at low impurity concentrations are equal to the measured $n_{ie}^2$-values[12]. In Fig. 1 the band edges for $N = 10^{15}\,\mathrm{cm}^{-3}$ are indicated and also the bandgap value $E_g(T = 300) = 1.121$ eV according to MacFarlane's measurements. How the bandgap is defined at higher impurity concentrations will be discussed later on. In Fig. 2 several quantities characterising the width of the band tails ($\sigma, \sigma_t$) and the screening length ($\lambda$) are shown as function of the impurity concentration. It appears that the screening length is nearly equal to the Debye length, valid for non-degenerate material.

In the following figures calculated and measured results[10] will be compared. In analogy with the case of low doping we want to describe the calculated $n_{ie}$-values as follows:

$$n_{ie}^2(N,T) = KT^3\exp(-qV_g(N,T)/kT). \tag{7}$$

In [10] it has been shown that $V_g(N,T)$ can be approximated for $T > 250°$K by:

$$V_g(N,T) = V_{g0}(N) - \alpha T. \tag{8}$$

In order to check if this experimental result is also confirmed by the calculations, we differentiate eqn (7):

$$V_g - T\frac{\mathrm{d}V_g}{\mathrm{d}T} = \frac{kT}{q}\left(\frac{\mathrm{d}\ln n_{ie}^2}{\mathrm{d}\ln T} - 3\right). \tag{9}$$

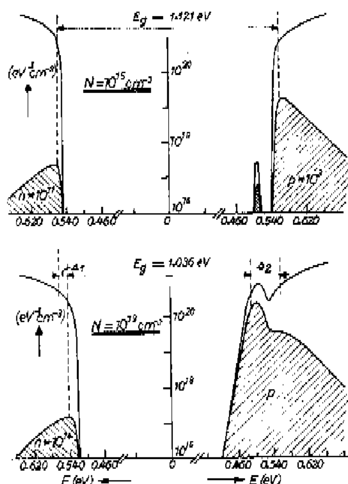



Fig. 1. Density of energy states for $p$-type silicon doped with $10^{15}$ and $10^{19}$ impurity atoms/cm$^3$. The distribution of the carriers as a function of energy at $T = 300$ K is also shown. At $N = 10^{19}$ cm$^{-3}$ the bandgap is smaller due to broadening of the impurity band and the forming of band tails.

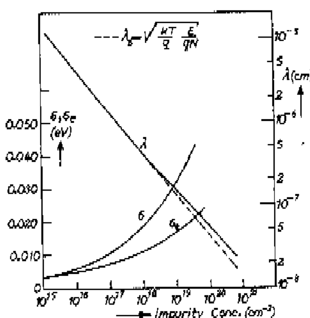



Fig. 2. Calculated standard deviations ($\sigma, \sigma_e$) and screening length ($\lambda$) as a function of the impurity concentration at $T = 300$ K.

The calculation of $n_{ie}$ (see Section 2) does not involve any fitting parameters, but only those known from the energy gap, effective masses, impurity energy level etc.

Upon substitution of the calculated $n_{ie}^2$-values on the right-hand side of eqn (9), it follows how $V_g - T(dV_g/dT)$ behaves as a function of temperature. This is shown in Fig. 3 for a number of impurity concentrations. From the flat part of these curves for temperatures above about 280°K it follows that $V_g(N, T)$ indeed is a linear function of temperature. $V_{go}(N)$ is given by the value of the flat part of these curves. This value is used as a definition of the impurity concentration-dependent bandgap. It follows that in this temperature range eqn (7) can be written as:

$$n_{ie}^2(N, T) = CT^3 \exp(-qV_{go}(N)/kT). \qquad (10)$$

The bandgap narrowing is:

$$\Delta V_{go}(N) = 1.206 - V_{go}(N). \qquad (11)$$

Figure 3 also indicates $(V_g - T(dV_g/dT))$ derived from measurements[10]. It was found that the constant $C$ in eqn (10) is not exactly constant but depends weakly on the impurity concentration. This may be due to $\alpha$ (see eqn 8) not being a constant but a function of $N$.

In Table 1 is shown that $C$ changes by about a factor of two when $N$ varies over about five orders of magnitude. The value of $C$ agrees well with the corresponding constant in the experimental formula of Putley and Mitchell (see Section 1). It follows from these calculations that as a first order approximation the relationship between $n_{ie}(N, T)$ and $n_{io}(T)$ is given by:

$$n_{ie}^2(N, T) = n_{io}^2(T) \exp(q\Delta V_{go}(N)/kT). \qquad (12)$$

It has been shown in [10] that the measured bandgap narrowing can be fit by the following formula:

$$\Delta V_{go,meas.}(N) = 9 \times 10^{-3}\left\{\ln\left(\frac{N}{10^{17}}\right) + \sqrt{\left(\ln\left(\frac{N}{10^{17}}\right)\right)^2 + \frac{1}{2}}\right\} \times [\text{eV}]. \qquad (13)$$

In Fig. 4 and Table 1 this is compared with the calculated values taken from Fig. 3.

Having shown that the $n_{ie}^2$ calculations as function of $N$ and $T$ can be approximated by eqn (10) where $V_{go}(N)$ is considered as a doping dependent bandgap, we will check up what part of the total narrowing comes from respectively the valence and conduction band. In analogy with the expressions for low impurity concentrations we try to describe the carrier concentrations as follows:

$$N \approx p = C_1 T^{\beta_1} \exp\left(-\frac{1}{kT}\left(\frac{E_g}{2} - \Delta_1 - F\right)\right) \qquad (14)$$

$$n = C_2 T^{\beta_2} \exp\left(-\frac{1}{kT}\left(\frac{E_g}{2} - \Delta_2 + F\right)\right) \qquad (15)$$

where $\Delta_1$ and $\Delta_2$ are the doping dependent narrowing of the bandgap on the side of the valence and conduction band. In order to determine $\Delta_1$ and $\Delta_2$ eqns (14) and (15) are differentiated:

$$\left(\frac{E_g}{2} - F\right) - T\frac{d}{dT}\left(\frac{E_g}{2} - F\right) = \Delta_1 - \beta_1 kT \qquad (16)$$

$$\left(\frac{E_g}{2} + F\right) - T\frac{d}{dT}\left(\frac{E_g}{2} + F\right) - kT\frac{d\ln n}{d\ln T} = \Delta_2 - \beta_2 kT. \qquad (17)$$

The terms on the left-hand side are derived from the calculations (see Section 2) and can be approximated as linear functions of $kT$ (see Fig. 5). From these straight lines the parameters $\Delta_1$, $\beta_1$ and $\Delta_2$, $\beta_2$ are derived and

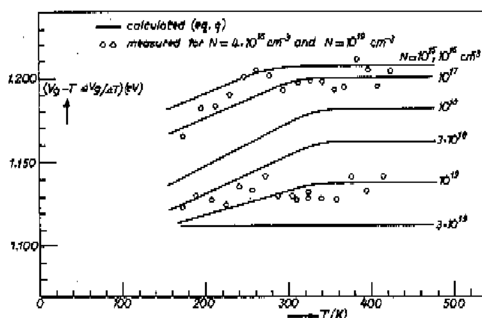



Fig. 3. The calculated and measured quantity $(V_g - T\Delta V_g/\Delta T)$ as a function of temperature for different impurity concentrations.

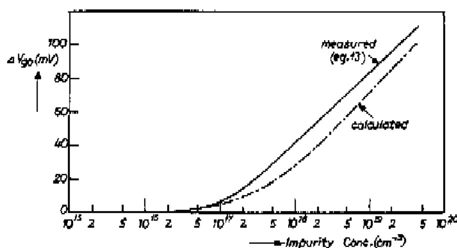



Fig. 4. Comparison between calculated (see Fig. 3) and measured [10] bandgap narrowing as a function of impurity concentration.

Table 1.

| $N$ ($cm^{-3}$) | $\Delta V_{go,meas}$ (mV) | $\Delta V_{go,calc}$ (mV) | $C \times 10^{32}$ (eqn 10) | $\Delta_1$ (mV) | $\Delta_2$ (mV) | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|---|---|---|
| $10^{15}$ | 0 | 0 | 9 | 1 | 0 | 1.5 | 1.5 |
| $10^{16}$ | 0 | 0 | 9 | 4 | 0 | 1.6 | 1.5 |
| $10^{17}$ | 6 | 7 | 9 | 13.5 | 0 | 1.8 | 1.5 |
| $10^{18}$ | 42 | 26 | 6 | 44 | 4 | 2.1 | 1.5 |
| $3 \times 10^{18}$ | 62 | 45 | 5 | 61 | 10 | 2.2 | 1.6 |
| $10^{19}$ | 83 | 69 | 5 | 68 | 17 | 1.8 | 1.7 |
| $3 \times 10^{19}$ | 103 | 95 | 5 | 70 | 17 | 1.5 | 1.5 |

given in Table 1. It appears that apart from the doping dependent narrowing $\Delta_1$ and $\Delta_2$ also $\beta_1$ and $\beta_2$ are not exactly constant and differ somewhat from the low doping value 1.5. Comparing the product of eqns (14) and (15) with eqn (10) $(\beta_1 + \beta_2)$ can deviate somewhat from the value 3 which in eqn (10) was enforced upon the power of $T$. This influences not only the constant factor but is also the reason for the discrepancy between the sum $(\Delta_1 + \Delta_2)$ and the total bandgap narrowing $\Delta V_{go}$ calculated from eqns (10) and (11). In Fig. 1 the values of $\Delta_1$ and $\Delta_2$ for $N = 10^{19}\ cm^{-3}$ are indicated.

#### 4. CONCLUSIONS

It has been shown that calculations of the $pn$-product for temperatures from 280 to about 450°K and for

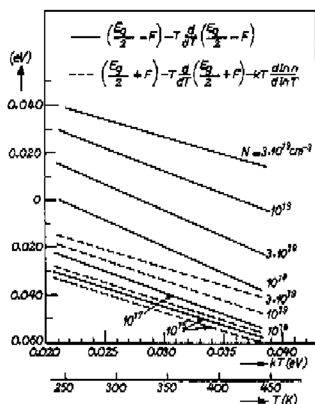



Fig. 5. Calculated quantities $\left(\frac{E_g}{2} - F\right) - T\Delta\left(\frac{E_g}{2} - F\right)/\Delta T$ and $\left(\frac{E_g}{2} + F\right) - T\Delta\left(\frac{E_g}{2} + F\right)/\Delta T - kT\Delta(\ln n)/\Delta(\ln T)$ as a function of $kT$ (see eqns (14)–(17)).

impurity concentrations varying from $10^{17}$ to about $3\times10^{19}\,\text{cm}^{-3}$ agree well with the measurements and support the already experimentally found [10] relationship:

$$pn = n_{ie}^2(N,T) = n_{io}^2(T)\exp\left(q\Delta V_{go}(N)/kT\right) \quad (17)$$

where $n_{io}^2(T)$ and $\Delta V_{go}(N)$ are given by eqns (1) and (13) which have been derived from experiments.

According to the calculations the total carrier concentrations can be described by eqns (14) and (15).

*Acknowledgement*—Many helpful discussions with H. C. de Graaff are gratefully acknowledged.

REFERENCES

1. A. A. Vol'fson and V. K. Subashiev, *Sov. Physics-Semiconductors* **1**, 327 (1967); M. Balkanski, A. Aziza and E. Amzallag, *Phys. Stat. Sol.* 31, 323 (1969).
2. E. O. Kane, *Phys. Rev.* **131**, 79 (1963).
3. T. N. Morgan, *Phys. Rev.* **139**, A343 (1965).
4. V. L. Bonch-Bruyevich, *The Electronic Theory of Heavily Doped Semiconductors*. Elsevier, Amsterdam (1966).
5. V. I. Fistul', *Heavily Doped Semiconductors*. Plenum Press. New York (1969).
6. D. D. Kleppinger and F. A. Lindholm. *Solid-St. Electron.* **14**, 199 (1971).
7. R. J. Van Overstraeten, H. J. J. De Man and R. P. Mertens, *IEEE Trans. Electron. Dev.* **ED-20**, 290 (1973).
8. M. S. Mock, *Solid-St. Electron.* **16**, 1251 (1973).
9. E. L. Heasell, *Internat. J. Electronics* **38**, 127 (1975).
10. J. W. Slotboom and H. C. de Graaff, *Solid-St. Electron.* **19**, 857 (1976).
11. H. D. Barber, *Solid-St. Electron.* **10**, 1039 (1967).
12. E. H. Putley and W. H. Mitchell, *Proc. Phys. Soc., London* **A72**, 193 (1958).
13. G. G. MacFarlane, J. P. McLean, J. E. Quarrington and V. Roberts, *Phys. Rev.* **111**, 1245 (1958).
14. F. Stern, *Phys. Rev.* 3, B3559 (1971).
15. R. P. Mertens, Thesis, Kath. Univ. Leuven (1972).
16. This problem was recognized in a discussion with R. P. Mertens.
17. R. B. Dingle, *Phil. Mag.* **46**, 831 (1955).

## 5.5. Bandgap Narrowing in Silicon Bipolar Transistors

Reprinted from IEEE Transactions ON *ELECTRON DEVICES*, Volume ED-24, August, 1977.

J.W. SLOTBOOM and H.C. DE GRAAFF

Abstract – **Martinelli [1] recently reported on measurements of the $I$-$V$ characteristics of silicon bipolar transistors as a function of temperature. His conclusion was that there was no evidence of bandgap narrowing in the transistors.**

**Our experiments [2] on N-P-N transistors indicate that the bandgap does narrow for impurity concentrations above $N = 10^{17}$ cm$^{-3}$. The reason for this discrepancy follows from Martinelli's assumption that the temperature dependence of the minority carrier mobility in the p-type base is given by $T^{-2.6}$, independently of the impurity concentration, which is not justified by our measurements.**

Optical absorption measurements [3], [4] have shown that the bandgap of silicon changes for high impurity concentrations. Using these measurements Kauffman and Bergh [5] and Buhanan [6] interpreted their measurements of the temperature dependence of the $I$-$V$ characteristics and the current gain in bipolar transistors by assuming a different bandgap in the base and in the heavily doped emitter. Because of the importance of this effect for the understanding and optimal design of semiconductor devices, there has been a discussion about the presence and magnitude of this phenomenon. In a recent publication [1] dealing with the temperature dependence of the ($I_B$-$V_{EB}$) and ($I_C$-$V_{EB}$) characteristics of silicon bipolar transistors Martinelli came to the following conclusions:

1) the ($I_B$-$V_{EB}$) characteristics are nonideal and therefore cannot be used to prove the presence of bandgap narrowing.
2) the ($I_C$-$V_{EB}$) characteristics are well described by the classical model without bandgap narrowing in the base region ($V_{go}$ = 1.20 eV).

In view of these conclusions it seems appropriate to present a short survey of arguments demonstrating why in our opinion, bandgap narrowing indeed is present. These arguments are taken from some recent publications [2] describing experiments on a number of N-P-N transistors, varying in base-doping concentration from $4 \times 10^{15}$ to $2 \times 10^{19}$ cm$^{-3}$. It appeared that bandgap narrowing ($\Delta V_{go}$) is present for impurity concentrations above about $N = 10^{17}$ cm$^{-3}$ and given by

$$\Delta V_{go}(N) = 9 \cdot \left\{ \ln \frac{N}{10^{17}} + \sqrt{\left(\ln\left(\frac{N}{10^{17}}\right)\right)^2 + 0.5} \right\} \text{ [mV]}. \tag{1}$$

Concerning the above mentioned conclusions [1], we completely agree with the first one and in fact it was for that reason that our experiments were concentrated on the ($I_C$-$V_{EB}$) instead of the nonideal ($I_B$-$V_{EB}$) characteristics. We cannot agree with his second conclusion. For transistors with base doping concentration of about $N = 10^{18}$ cm$^{-3}$, as were used by Martinelli, a bandgap narrowing of 42 mV

would follow from (1). The reason for the disagreement lies in Martinelli's assumption that the temperature dependence of the electron mobility is given by $T^{-2.6}$, independently of the base doping concentration. This assumption means that in the expression

$$I_C = CT^m \exp\left(-q\left(V_{g0} - V_{EB}\right)/kT\right) \tag{2}$$

$m$ has a constant value 1.4 for N-P-N transistors [1]. It will be shown, however, that our experiments do not agree with this assumption. Differentiation of (2) with respect to the temperature $T$ while $I_C$ is kept constant gives:

$$V_{EB} - T\frac{dV_{EB}}{dT} = V_{g0} + m\frac{kT}{q}\,. \tag{3}$$

We measured $V_{EB}$ as a function of temperature for a number of N-P-N transistors with different impurity concentrations in the base while $I_C$ was constant. In Fig. 1 the term $(V_{EB} - T\,\Delta\,V_{EB}\,/\,\Delta\,T)$, which was directly derived from these measurements, is shown as a function of temperature for two transistors with different base doping concentrations. For comparison a line with $V_{g0}$ = 1.205 eV and $m$ = 1.4 is included. It is clear that this line does not fit our experiments very well, these

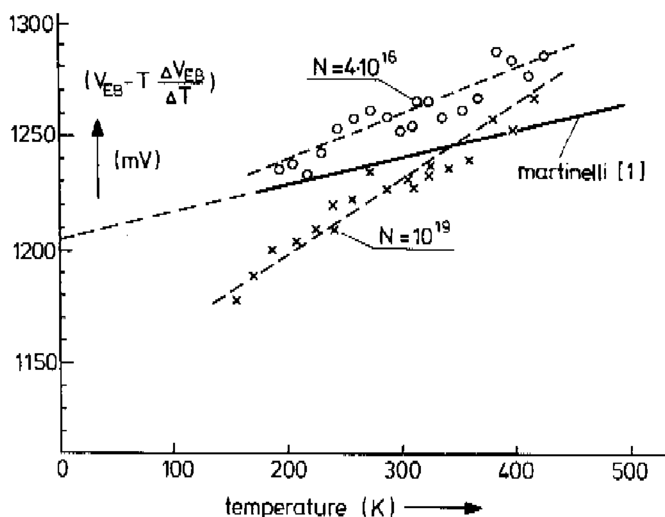



Fig. 1. Measurements of $V_{EB} - T\,\Delta\,V_{EB}\,/\,\Delta\,T$ as a function of temperature for two transistors with different base doping concentrations [2], compared with the behaviour suggested by Martinelli [1];

experiments indicating that $m$ should be a function of the impurity concentration rather than being a constant. These transistor temperature measurements are not enough to derive the value of $m$ and the bandgap narrowing accurately and we combined them with similar temperature measurements of the base sheet resistance ($R_{\square Base}$) underneath the emitter, which was taken from the same slice as the transistor [2]. The $m$-values derived in this way are shown in Fig. 2 (the $\beta$-values from [2c] table 2 are the same as the $m$-values). We see that for $N = 10^{18}$ cm$^{-3}$ $m$ is about 3. When this value for $m$ is applied in Martinelli's [1, fig. 6], which gives a relationship between $m$ and the bandgap for his transistors, a narrowing of about 45 mV appears. This accords well with the bandgap narrowing of 42 mV which we predict according to (1).

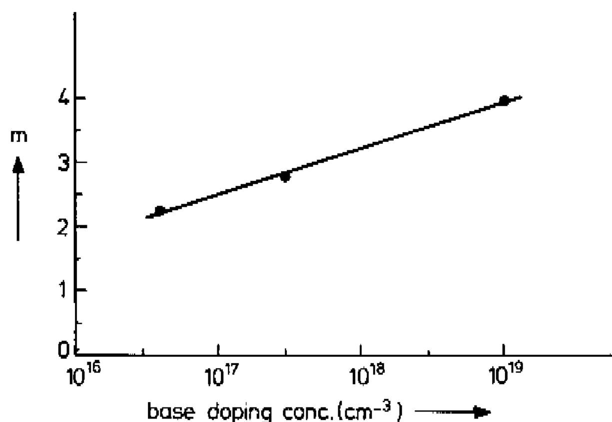



Fig. 2. The impurity concentration dependence of $m$ (see eq. (1)) derived from measurements [2].

It is important to notice that the bandgap narrowing was derived not only from the temperature dependence as described above, but also from the magnitude of the $I_0$ and $R_{\square Base}$ measured at room temperature [2]. Bandgap narrowing values obtained by both methods agree and are fitted by (1). It was pointed out in [2c] that in all these transistor measurements it is in fact the product $(\mu_n n_{ie}^2)$ which is measured and that the behavior of the pn-product, $n_{ie}^2$, can only be derived from these measurements if the minority carrier mobility is known. Unfortunately, there are no experimental data on minority carrier mobility for high impurity concentrations. For concentrations below $10^{17}$ cm$^{-3}$ there is no difference between drift mobility and conductivity mobility [7], and although there is no experimental evidence for high impurity concentrations we assumed that the electron mobilities

in n- and p-type silicon are similar as a function of impurity concentration and temperature. A number of arguments support this assumption:

1) theoretical calculations [8] of the pn-product as a function of temperature and impurity concentration,
2) if the measurements were interpreted in terms of minority carrier mobility and it was assumed that no bandgap narrowing at all occurred, the resultant mobility behaviour would seem highly improbable, being given by

$$\mu_{n,min.}(N,T) = \mu_{n,maj.}(N,T) \exp\left(q \, \Delta V_{go}(N) / kT\right) \tag{4}$$

as shown in Fig. 3,
3) using bandgap narrowing values according to (1) in calculations for the magnitude and temperature dependence of injection of minority carriers into heavily doped regions, such as $n^+$ of $p^+$ emitters, buried layers, isolation regions etc. agrees quantitatively well with measurements [9], [10].

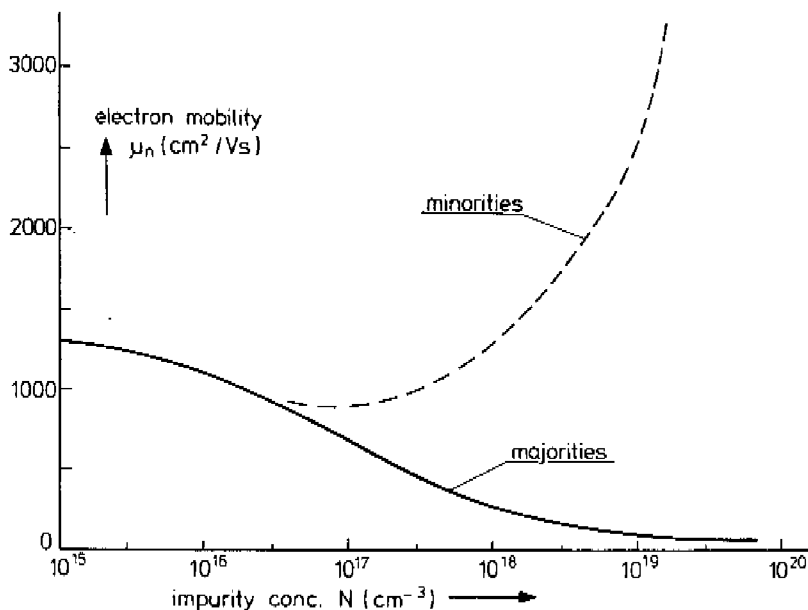



Fig. 3. Comparison of the majority carrier mobility with the minority carrier mobility as a function of impurity concentration, assuming that there is no bandgap narrowing at all (see eq. 4).

In conclusion it can be said that Martinelli's assumption that the parameter $m$ in (2) should have a constant value of 1.4 (for bipolar N-P-N transistors) is not in agreement with our experiments, which show that $m$ has a much higher value and varies with the impurity concentration in the base region.

It has been shown that our interpretation in terms of bandgap narrowing, which explains our own measurements, is equally applicable to Martinelli's measurements and indicates that for base doping concentrations of about $N = 10^{18}$ cm$^{-3}$ a bandgap narrowing of about 42 mV occurs. Several arguments supporting our assumption concerning the minority carrier mobility have been discussed.

## REFERENCES


[1] R.U. Martinelli, "The temperature dependence of the dc base and collector currents in silicon bipolar transistors," *IEEE Trans. Electron devices*, vol. ED-23. p. 1218, 1976.

[2] a) J.W. Slotboom and H.C. de Graaff, "Experiments on high-doping effects in bipolar transistors", presented at European Solid State Device Research Conf., Grenoble, France, 1975.

b) – –, "Experimental determination of the bandgap in the base region of bipolar transistors", presented at Int. Electron Devices Meeting, Washington DC, 1975.

c) – –, "Measurements of bandgap narrowing in Si bipolar transistors", *Solid-State Electron.*, vol. 19, p. 857, 1976.

[3] A.A. Vol'fson and V.K. Subashiev, "Fundamental absorption edge of silicon heavily doped with donor or acceptor impurities", *Soc. Physics – Semiconductors,* vol. 1, p. 327, 1967.

[4] M. Balkanski, A. Aziza, and E. Amzallag, "Infrared absorption in heavily doped n-type Silicon", *Phys. Stat. Sol.*, vol. 31, p. 323, 1969.

[5] W.L. Kauffman and A.A. Bergh, "The temperature dependence of ideal gain in double diffused silicon transistors", *IEEE Trans. Electron Devices*, vol. ED-15, p. 732, 1968.

[6] D. Buhanan, "Investigation of current gain temperature dependence in silicon transistors", *IEEE Trans. Electron Devices,* vol. ED-16, p. 117, 1969.

[7] E.M. Conwell, "Properties of silicon and germanium: II", *Proc. IRE,* vol. 46, p. 1281, 1958.

[8] J.W. Slotboom, "The pn-product in silicon", *Solid-State Electron.*, vol. 20, p. 279, 1977.

[9] H.C. de Graaff, J.W. Slotboom, and A. Schmitz, "The emitter efficiency of bipolar transistors: theory and experiments", *Solid-State Electron.*, vol. 20, p. 515, 1977.

[10] J.W. Slotboom, "Minority carrier injection into heavily doped silicon", *Solid-State Electron.*, Vol. 20, p. 167, 1977.

## 5.6. "Minority carrier injection into heavily doped silicon"
Solid-State Electr., Vol. 20, p. 167, 1977.

## COMMUNICATION

### MINORITY CARRIER INJECTION INTO HEAVILY DOPED SILICON

J.W. Slotboom
Philips Research Laboratories
Eindhoven, The Netherlands

Introduction

Many authors have studied the problem of injection of minority carriers into a heavily doped region, because of its great importance for the understanding of silicon devices (emitter efficiency in bipolar transistors, $I^2L$, thyristors, solar cells etc.). Particular attention has been paid to several effects e.g. the influence of a gradient doping profile (built-in electric field) [1], bandgap narrowing [2], Auger recombination [3] etc.

In this letter it will be shown that the most simple model with a constant impurity concentration, predicts the current density of the injected minority carriers in good agreement with measurements. The crucial point is that in the classical expression for the current density the doping dependence of both the effective bandgap [4] and the minority carrier lifetime [5] have to be taken into account. This simple model also explains the experimentally known fact [3] that above an impurity concentration level of about $10^{19}cm^{-3}$ the current density is fairly insensitive to the other process parameters (diffusion depth, dopant atoms, doping profile etc.). This effect will also be illustrated by accurate computer calculations showing that the gradient of the doping profile has only a relatively small influence on the injected minority current density as long as the region is not transparent.

Simple model (N = constant)

In the classical analysis of the injection of minority carriers (e.g. holes) into a heavily doped region ($N^+$-region) with a constant impurity concentration N (see fig. 1), the excess hole concentration is given by the diffusion equation :

$$\frac{d^2\hat{p}}{dx^2} = \frac{\hat{p}}{L^2} \qquad L = \sqrt{D_p \tau} \quad (1)$$

The doping dependence of the carrier lifetime $\tau$ is

$$\frac{1}{\tau} = \frac{1}{\tau_0} + 2.25\ 10^{-19}\ N^{1.36} \quad (2)$$

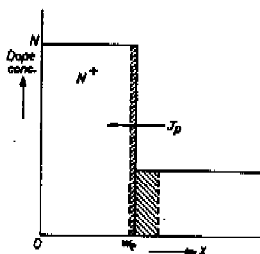



Fig. 1

where $\tau_0$ is the Hall-Shockley-Read lifetime and the second term follows from a fit of the measured minority carrier lifetimes by Beck and Conradt [5]. Eq.(1) is solved in the usual way with a finite surface recombination velocity, s, at the contact in x = 0. The hole current density injected at $x = w_e$ (the boundary of the quasi-neutral region) is given by :

$$J_p = J_o \exp(qV/kT) \quad , \quad J_o = F \cdot \frac{q D_p n_{ie}^2}{N L} \quad (3)$$

The factor F is given by :

$$F = \frac{\sinh(w_e/L) + (sL/D_p)\cosh(w_e/L)}{\cosh(w_e/L) + (sL/D_p)\sinh(w_e/L)} \quad (4)$$

and represents the influence of the surface recombination. If the region is not transparent ($w_e \gg L$) the surface condition has no influence and F = 1. Due to the doping dependence of the bandgap, $n_{ie}^2$ is a function of N. Experimentally and theoretically [4,6] it has been found that :

$$n_{ie}^2(T,N) = n_{io}^2(T)\exp\left(q\Delta V_{go}/kT\right) \quad (5)$$

where $n_{io}^2(T)$ is the usual pn-product for intrinsic and lowly doped silicon [7] and $\Delta V_{go}(N)$ the effective bandgap narrowing [4,6].

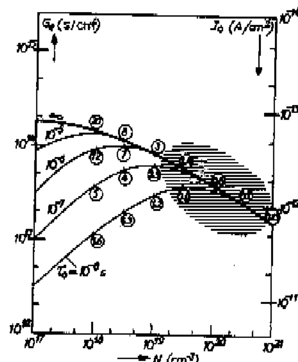



Fig. 2

Minority carrier injection into $N^+$-region as function of the doping level (see: simple model eq.(3)). The recombination diffusion length L (eqs. (1,2)) is indicated in µm in the circles along the curves.

By analogy with the Gummel number for the base region of a transistor, a figure of merit, $G_e$, can be defined characterizing the injection in heavily doped regions,

$$J_o = \frac{q\, n_{io}^2}{G_e} \quad (6)$$

In contrast with the classical analysis the influence of the impurity concentration on $n_{ie}^2$ and L in the expressions for $J_o$ and $G_e$ (see eq. (3) and (6)) have been taken into account in the manner given above. In fig. 2 the results are given assuming $w_e \gg L$. Experimentally found values for the minority current densities [3,9] are usually

always situated in the dashed area.

Numerical analysis ($N = N(x)$)

In order to analyse more realistic doping profiles, accurate numerical calculations have been done taking into account the doping dependence of the bandgap and of the minority carrier lifetime in the way described above. The minority carrier current density is [8] :

$$J_p = q\mu_p p\left(E + \frac{kT}{q}\frac{1}{n_{ie}}\frac{dn_{ie}}{dx}\right) - qD_p\frac{dp}{dx} = J_{FIELD} - J_{DIFF.} \quad (7)$$

The influence of a nonconstant impurity profile will be illustrated by the results calculated for a series of doping profiles varying in gradient (see fig. 3). The thickness of the doped region has been chosen such that the surface recombination is not important (i.e. nearly all of the injected minority carriers have disappeared by recombination in the bulk of the $N^+$ region before the contact at $x = 0$ is reached). From the continuity equation it follows then that the injected current density is the integral of the recombination

$$J_p = q\int_0^{w_e} p/\tau \ dx \quad (8)$$

From the calculations (see fig. 3) it appears that a pronounced maximum of the recombination occurs in the bulk of the quasi-neutral region at a point where the hole concentration is rather low. This is

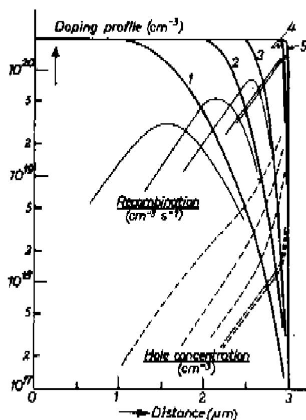



Fig. 3

Influence of the doping profile on the distribution of the recombination ($p/\tau$) and the holes; the scale for the recombination is 10 x larger as indicated and for the hole concentration $10^6$ x smaller. V = 600 mV.

caused by two essential phenomena acting together; doping-dependent bandgap narrowing increases the hole concentration and decreases its steep fall-off, and further the minority carrier lifetime decreases strongly with an increasing doping concentration. When the gradient of the doping profile is increased, the recombination peak shifts towards the junction and increases but narrows at the same time. It is clear that an abrupt profile can be made much shallower than a more slowly changing profile without the region becoming transparent. Although the internal distribution of carriers, electric field, recombination etc. depends strongly

upon the shape of the doping profile, the total injected minority current density at $x = w_e$ is only slightly changed (see table 1). The electric field at $x = w_e$ is partly compensated (20-90 percent) by the changing bandgap (see eq.(7)) and $J_{FIELD}$ and $J_{DIFF}$ were always nearly equal and several orders of magnitude larger than the resulting injected current density $J_p$.

Conclusion

The relative insensitivity of $J_o$ and $G_e$ to the process parameters and the doping profile, following from experiments and numerical calculations, is the basic reason that the simple Shockley model with N = constant predicts remarkably well the injected minority carrier current density provided that $n_{ie}^2$ and $\tau$ as function of N are taken into account.
However, the internal distributions are indeed strongly profile dependent. For a transparent region with a doping profile which strongly deviates from an abrupt profile the influence of the surface recombination according to the simple model, has to be interpreted cautiously.
A more detailed and accurate analytical model, valid for realistic doping profiles and also taking into account surface recombination correctly, will be published elsewhere [9].
The author is grateful to H.C. de Graaff and P.A.H. Hart for helpful discussions.

Table I

| Profile (fig.3) | Ge($s/cm^4$) | $J_p$($A/cm^2$) (V=600 mV, T=300K) |
|---|---|---|
| 1 | 4.6 $10^{13}$ | 0.6 $10^{-2}$ |
| 2 | 3.9 $10^{13}$ | 0.7 $10^{-2}$ |
| 3 | 3.4 $10^{13}$ | 0.8 $10^{-2}$ |
| 4 | 2.9 $10^{13}$ | 0.9 $10^{-2}$ |
| 5 | 2.9 $10^{13}$ | 0.9 $10^{-2}$ |

References

1. Kennedy, D.P., Murley, P.C., IRE Trans. ED, 1962, 9, pp. 136-142.
2. Mertens, R.P., de Man, H.J.J., Van Overstraeten, R.J., IEEE Transact. on El.Dev., 1973, Vol. ED-20, pp.772-778.
3. Burtscher, J., Dannhäuser, F., Krausse, J., Solid-State Electron., 1975, vol. 18, pp. 35-63.
4. Slotboom, J.W., de Graaff, H.C., Solid-State Electron., 1976, Vol. 19, pp. 857-862.
5. Beck, J.D., Conradt, R., 1973 Vol. 13, pp. 93-95.
6. Slotboom, J.W., to be published in Solid-State Electron.
7. Putley, E.H. Mitchell, W.H., Proc. Phys. Soc., 1958, London A72, pp. 193-200.
8. Van Overstraeten, R.J., de Man, H.J.J., Mertens, R.P., IEEE Transact. on El.Dev. 1973, Vol. ED-20, pp. 290-298.
9. De Graaff, H.C., Slotboom, J.W., Schmitz, A., to be published in Solid-State Electron.

5.7. "The emitter efficiency of bipolar transistors: Theory and Experiments". Solid-State Electr., Vol. 20, 1977.

# THE EMITTER EFFICIENCY OF BIPOLAR TRANSISTORS

## THEORY AND EXPERIMENTS

H. C. de GRAAFF, J. W. SLOTBOOM, and A. SCHMITZ
Philips Research Laboratories, Eindhoven, The Netherlands

(Received 14 September 1976; in revised form 13 December 1976)

**Abstract**—A figure of merit ($G_e$) for the emitter is defined, which takes account of bandgap narrowing caused by high impurity concentrations, a doping-dependent lifetime, the built-in electric field and the recombination velocity at the emitter contact. A simple formula is given for $G_e$, based on computer simulations, and tested by several experiments.

### 1. INTRODUCTION

When the current gain ($h_{FE}$) of a bipolar transistor is measured as a function of the collector current ($I_c$), one can see that $h_{FE}$ increases with $I_c$ at low current levels, then reaches a maximum, and decreases at high current levels. This decrease in $h_{FE}$ is due to high injection in the base and/or to collector saturation[1, 2]. The low level current gain is dominated by mechanisms such as surface recombination and recombination in the emitter-base space charge region. In this current range the base current ($I_b$) usually increases with $\exp(qV_{be}/mkT)$ where the non-ideality factor $m$ is between 1 and 2[3].

By choosing emitter geometries with a large area to periphery ratio side wall injection can be made negligibly small.

The maximum value ($h_{FEmax}$) of the current gain is mainly determined by the emitter efficiency. For an npn transistor this efficiency is defined as the ratio of the current densities of electrons and holes

$$\eta = J_n/(J_n + J_p).$$

For the electron current density ($J_n$) we can write

$$J_n = J_{no}\exp(qV_{be}/kT)$$

with $J_{no} = qn_{ib}^2/G_b$. $G_b$ is called the (base) Gummel number[4]. In the current range around $h_{FEmax}$ we can write for the base current density ($J_p$)[5]

$$J_p = J_{po}\exp(qV_{be}/mkt), \text{ with } m \approx 1.$$

$J_p$ is a hole current due to recombination in the bulk of the $n^+$ emitter and at the interface between the contact and the silicon. If the $m$-values are larger than 1.1–1.2 this indicates that other recombination components are also significant, and that $h_{FEmax}$ is not mainly determined by the emitter efficiency.

By analogy with the Gummel number for the base, we can introduce a figure of merit ($G_e$) for the emitter, by putting

$$J_{po} = \frac{qn_{io}^2}{G_e}. \quad (1)$$

In the literature $G_e$ is sometimes denoted by $(Q_e/D_e)_{eff}$. If the emitter impurity concentration ($N_D$) is constant and the surface recombination velocity is infinite $G_e$ takes the form[6]

$$G_e = \frac{L_p N_D}{D_p}\tanh\left(\frac{W_e}{L_p}\right) \quad (2)$$

with $L_p = \sqrt{D_p \tau_p}$ = diffusion recombination length
$D_p$ = diffusion constant for holes
$\tau_p$ = lifetime for holes
$W_e$ = width of the neutral emitter region.

In real emitters $N_D$ is not constant, but decreases from the surface towards the metallurgical junction. This gives rise to the following effects: (a) The lifetime is doping-dependent and thus varies with position. Auger processes are probably responsible for this[7, 8]. (b) The diffusion constant $D_p$ is also doping-dependent. Together with a non-constant $\tau_p$ it makes $L_p$ a function of position. Because it is desirable for modelling purposes to have a constant diffusion recombination length, a new definition of this quantity will be given. (c) At high doping concentrations, as encountered in emitters, the bandgap of silicon decreases. This decrease makes the intrinsic concentration ($n_i$) higher at higher doping concentrations[9] which, in turn, increases the injected minority carrier concentration. (d) There is a built-in electric field in the emitter, counteracting the diffusion of holes from the base to the contact. Moreover, the built-in electric field is modified by the position dependent $n_i$.

Kennedy and Murley[10] also treated the case of a non-constant doping profile, but with a constant lifetime and no bandgap narrowing. Choo[11] made the same simplifications but neglected moreover the emitter contact recombination.

Sheng has developed a model[8] which takes bandgap narrowing and Auger recombination into account, but it omits the influence of the electric field. Computer simulations have shown, however, that the field and diffusion components are of the same order of magnitude.

This paper describes a simplified, analytical model for the calculation of the figure of merit ($G_e$), based on numerical results of computer solutions of the one-

dimensional transport equations[12], extended with bandgap narrowing and Auger recombination. The model incorporates all the above-mentioned effects plus the influence of a finite surface recombination velocity at the emitter contact. With the help of this analytical model several experiments are analyzed, such as the influence of the emitter diffusion depth of the $G_e$ of LEC transistors[5], the influence of the Al contact in double-diffused and I²L structures and the effectiveness of the mono and poly parts in polysil emitters[13].

## 2. THE EMITTER MODEL

For the figure of merit $G_e$ in eqn (1) we can deduce (see Appendix)

$$G_e = \frac{N(o)}{s}\left\{\frac{n_{io}}{n_i(o)}\right\}^2 g(o) + \int_o^{W_e} \frac{N(x)}{D_p(x)}\left\{\frac{n_{io}}{n_i(x)}\right\}^2 g(x)\,dx \quad (3)$$

where $s$ = surface recombination velocity (surface band bending incorporated) at the emitter contact.
$W_e$ = width of the neutral emitter region.
$g(x)$ = weighting function for the doping concentration.

The emitter contact is situated at the surface ($x = 0$). Physically speaking $g(x)$ is the ratio of the hole current at a point $x$ and the hole current entering the neutral emitter zone at $x = W_e$

$$g(x) = \frac{J_p(x)}{J_p(W_e)}. \quad (4)$$

The advantage of writing $G_e$ in the form of eqn (3) is that only the weighting function $g(x)$ has to be modelled, because the other quantities are known functions of doping concentration. Thus the doping profile $N(x)$ being given, $D_p(x)$ can be calculated[14]. The intrinsic concentration follows from[15]

$$n_i^2(x) = n_{io}^2 \exp\{q\Delta V_{go}(x)/kT\} \quad (5)$$

where[16]

$$n_{io}^2 = 9.61 \times 10^{32} T^3 \exp\left(-1.205\frac{q}{kT}\right). \quad (6)$$

The bandgap narrowing ($\Delta V_{go}$) is a function of the doping concentration[17]

$$\Delta V_{go}(x) = 9 \times 10^{-3}\left\{\ln\left(\frac{N(x)}{10^{17}}\right) + \sqrt{\left(\ln\frac{N(x)}{10^{17}}\right)^2 + \frac{1}{2}}\right\}. \quad (7)$$

To evaluate $G_e$ by means of eqn (3) we still need to know the weighting function $g(x)$. In general this function depends on the contact recombination velocity ($s$) and, because of the Auger recombination, bandgap narrowing and built-in electric field, on the doping profile $N(x)$. It turned out that $g(x)$ can be modelled in a simple, normalized form, which is nearly independent of the doping profile. This simple form is obtained by fitting the numerical results of a one-dimensional computer simulation[12]. In this simulation the recombination is given by

$$R = \frac{pn - n_i^2}{\tau_o(p + n + 2n_i)} + C_A\, p\, n^{1.36} \quad (8)$$

where

$$C_A = 2.25 \times 10^{-19}\,\text{cm}^4/\text{s}.$$

The first term in eqn (8) represents the SRH recombination and influences mainly the recombination in the space charge region; it is made of minor importance by choosing $\tau_o = 10^{-4}$s. The Auger term in eqn (8) is a fit of the experimental results of Beck and Conradt[18] for the lifetime as a function of impurity concentration.

If a new diffusion-recombination length ($L_{peff}$) is defined as the distance between the edge ($x = W_e$) of the neutral emitter and the point where the hole current has been halved ($J_p(W_e - L_{peff}) = \frac{1}{2}J_p(W_e)$), we can introduce a normalized position variable

$$u = \frac{W_e - x}{L_{peff}} \text{ and put } g(x) \equiv G(u).$$

We learned from the computer simulations that the function $G(u)$ is nearly independent of the doping profile and can be approximated by

$$G(u) \approx \exp\left\{-\left(\frac{u}{1.096}\right)^4\right\}. \quad (9)$$

Figure 1 gives the numerical results for four Gaussian

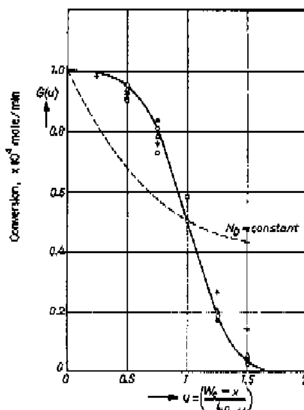



Fig. 1. The weighting function $G(u)$ for infinite surface recombination velocity. Fully drawn line: according to eqn (9). Dashed line: constant doping concentration and $W_e = 1.5\sqrt{D_p\tau_p}$. Computer simulations: ○○, gaussian profile, $N(o) = 1.10^{21}$ cm$^{-3}$; △△, gaussian profile, $N(o) = 1.10^{20}$ cm$^{-3}$; + +, gaussian profile, $N(o) = 3.10^{19}$ cm$^{-3}$; □□, gaussian profile, $N(o) = 1.10^{19}$ cm$^{-3}$; ●●, non-gaussian profile, $N(o) = 8.10^{20}$ cm$^{-3}$.

profiles with different values of the surface concentration $N(0)$ and for one non-gaussian profile. The surface recombination velocity ($s$) is assumed to be infinite. The fully drawn line gives $G(u)$ according to eqn (9). The weighting function for an emitter with $N_D$ = constant and $W_e = 1.5\sqrt{D_p\tau_p}$ is also shown in Fig. 1.

Going from the contact at $x = 0$ to the edge of the neutral emitter at $x = W_e$, the variable $u$ varies between $W_e/L_{peff}$ and 0. In the case of deep emitter diffusions, especially when $W_e/L_{peff} > 1.5$ and $G(W_e/L_{peff}) \approx 0$ (see Fig. 1) the value of the surface recombination velocity is of no importance because holes do not reach the contact. However, in the case of shallow emitters ($W_e/L_{peff} < 1$) the value of $s$ is important and in fact changes the function $g(x)$. Thus for $s$ infinite, the weighting function is given by $G(u)$. For finite values of $s$ the weighting function changes and becomes mathematically more complicated.

Therefore we approximate $g(x)$ in this case by two straight lines through the points P1 and P2, as indicated in Fig. 2. The value of $g(x)$ at the contact is modelled as

$$g(o) = g_{P1} = \frac{\alpha}{\frac{\alpha}{G(W_e/L_{peff})} + \sqrt{\frac{1}{G^2(W_e/L_{peff})} - 1}} \qquad (10a)$$

with

$$\alpha = \frac{sL_{peff}}{D_p(o)}.$$

Equation (10a) shows that for $s = \infty$ one has $g(o) = G(W_e/L_{p\,eff})$ and for $s = 0$ one has $g(o) = 0$. The point P2 is chosen in the middle of the neutral emitter region ($u = \frac{1}{2}W_e/L_{peff}$) and $g_{P2}$ is taken as

$$g_{P2} = \tfrac{1}{8}\,g(o) + \tfrac{7}{8}\,G(\tfrac{1}{2}\,W_e/L_{peff}). \qquad (10b)$$

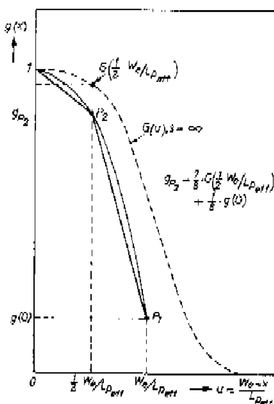



Fig. 2. The straight lines approximation for $g(x)$ when $s$ is finite. The coordinates of the points $P1$ and $P2$ are chosen as $P1$: $(W_e/L_{peff})$ and $g(o)$. P2: $\frac{1}{2}(W_e/L_{peff})$ and $(\frac{7}{8}G(\frac{1}{2}(W_e/L_{peff}) + \frac{1}{8}g(o))$.

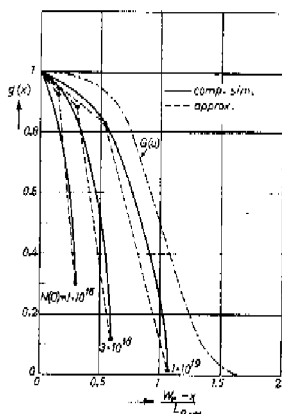



Fig. 3. Computer simulations for $g(x)$ with $s = 100$ cm/s. The dashed lines give the approximation indicated in Fig. 2.

In Fig. 3 the straight lines approximation is compared with numerically calculated weighting functions, taking $s$ = 100 cm/s.

Another important parameter of the weighting function is the newly defined diffusion-recombination length $L_{peff}$; it is not very sensitive to the shape of the doping profile, but it does depend on the surface concentration $N(o)$. Taking gaussian profiles with a characteristic length $L (= 4\,\mu m)$

$$N(x) = N(o)\exp\left\{-\left(\frac{x}{L}\right)^2\right\} - N_A, \qquad (11)$$

we obtained, by means of computer simulations, $L_{peff}/L$ as a function of $N(o)$. This is shown in Fig. 4. The neutral emitter width ($W_e$) is almost equal to the metallurgical junction depth ($X_{je}$). As a rule of thumb we put

$$\frac{W_e}{L} = 0.98\frac{X_{je}}{L} = 0.98\sqrt{\ln\frac{N(o)}{N_A}}. \qquad (12)$$

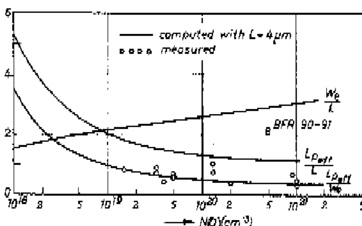



Fig. 4. The normalised width $W_e/L$ of the neutral emitter region and the normalised diffusion recombination length $L_{peff}/L$ as a function of the surface concentration $N(o)$ for gaussian profiles. The quantity $L_{peff}/W_e$ is also shown; the dots are experimental values.

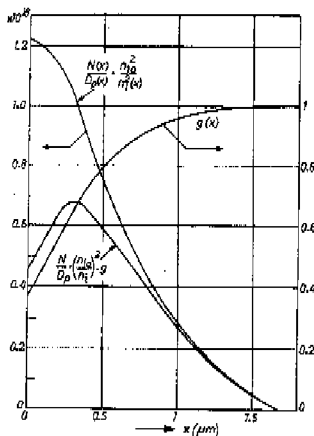



Fig. 5. The non-gaussian, reduced profile $[N(x)/D_p(x)][n_{io}^2/n_i^2(x)]$, the weighting function $g(x)$ and the integrand of eqn (3) for a double-diffused transistor with $W_e = 1.7\,\mu m$, $L_{peff} = 1.5\,\mu m$ and $G_e = 5.9\times 10^{13}\ s\ cm^{-4}$.

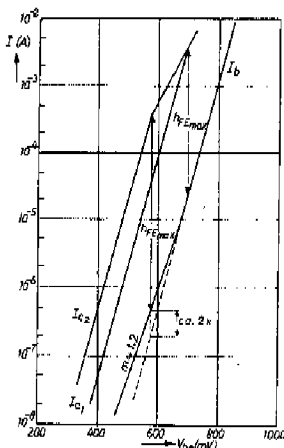



Fig. 6. $I_c$ and $I_b$ vs $V_{be}$ for two fictitious transistors with identical emitters but base dopings differing by a factor of ten. The most lightly doped base ($I_c = I_{c1}$) gives the lower $G_e$ value.

This is also shown in Fig. 4, together with the ratio $L_{peff}/W_e$. The latter quantity which is supposed to be useful for other than gaussian profiles too, is also determined experimentally (see Section 3.1). Figure 5 gives an example of an emitter with a figure of merit $G_e = 5.9\times 10^{13}\ s\ cm^{-4}$. The doping profile is not gaussian, $s$ is infinite and $N(o) \sim 5\times 10^{20}\ cm^{-3}$.

## 3. EXPERIMENTS

The figure of merit is determined experimentally from the $h_{FEmax}$ and the base Gummel number

$$G_e = h_{FEmax}\times G_b. \tag{13}$$

$G_b$ is obtained from the $(I_c, V_{be})$ characteristic. This method for the determination of $G_e$ is based on the assumption that for the base current the non-ideality factor $m$ is equal to 1, at least in the current range around $h_{FEmax}$. Especially with lightly doped bases this is not always true and errors in $G_e$ will occur. To illustrate this, two fictitious transistors are compared (see Fig. 6); they have identical emitters and therefore identical base currents, with $m = 1.2$ in the lower current range. Their base impurity concentrations and Gummel numbers differ however by a factor of ten. We obtain for the transistor with the more lightly doped base a $G_e$ which is about two times lower than the other $G_e$, where in fact they should be equal. An alternative method is to measure the complete $(I_b, V_{be})$ characteristic.

### 3.1 *Experimental determination of* $L_{peff}$

The $G_e$ was measured for a great many transistors of different types, with emitter surface concentrations ranging from $N(o) = 1.5\times 10^{19}$ to $1\times 10^{21}\ cm^{-3}$ and emitter diffusion depths between $X_{je} = 0.22$ and $16\,\mu m$. From the values of $G_e$, $N(o)$ and $W_e$ we obtained with the help of eqns (3) and (9) the diffusion-recombination length ($L_{peff}$), assuming that the profile is more or less gaussian and that the recombination velocity is infinite. The results are shown in Fig. 4; the agreement with the theoretical model is fairly satisfactory, except for the microwave transistors BFR 90-91. These have very shallow *As*-doped implanted emitters with a more or less constant impurity concentration. For such emitter the classical weighting function (see Fig. 1) is better suited than $G(u)$.

### 3.2 *Etching away the emitter material*

We start with eight identical slices. The emitter junction depth $X_{je} = 1.7\,\mu$, the doping profile is given in Fig. 7(a) and $G_b = 1.4\times 10^{11}\ s\ cm^{-4}$. Then for each slice a different amount of emitter surface material is etched away so that we end up with eight different junction depths, ranging from $X_{je} = 1.7$ to $0.3\,\mu m$. For each value of $X_{je}$ the $G_e$ is determined and the results are given in Fig. 7(b). Application of eqn (3) gives the best results for the higher $X_{je}$-values if $L_{peff} = 1.4\,\mu m$; the value of $s$ is of minor importance here. For low $X_{je}$-values $L_{peff}$ has little influence because the emitter becomes gradually transparant for holes ($g(x) \approx 1$). In this region $G_e$ is very sensitive to the value of $s$. The best fit is obtained for $s = 3\times 10^5\ cm/s$.

### 3.3 *Influence on* $G_e$ *of the diffusion depth of the* $n^+$ *region in LEC transistors*

Several LEC transistors[5] were investigated. The $n^+$ emitter doping profile was approximately gaussian with $N(o) = 1\times 10^{21}\ cm^{-3}$. The diffusion depth of the $n^+$ region

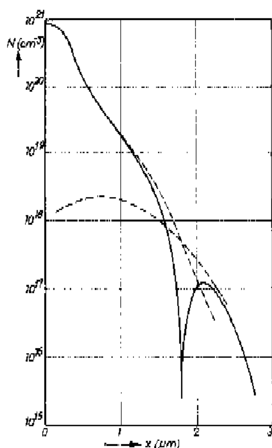



Fig. 7(a). Doping profile of the transistors in the etching experiment before any emitter material is etched away. The base Gummel number $G_b = 1.4 \cdot 10^{13}$ s cm$^{-4}$.

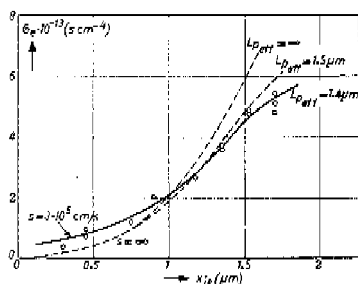



Fig. 7(b). $G_e$ as a function of junction depth. The best fit to the experimental points is obtained with $L_{peff} = 1.4\,\mu m$ and $s = 3 \times 10^5$ cm/s.

was varied between 1.35 and 3.37 $\mu$m. The ratio $W_e/L = 3.3$ being kept constant, the characteristic length $L$ (see eqn 11) varied between 0.4 and 1 $\mu$m.

If we assume $s$ to be sufficiently large, eqn (3) gives

$$G_e = \int_0^{W_e} \frac{N}{D_p}\left(\frac{n_{ie}}{n_i}\right) g\,\mathrm{d}x = L \int_0^{W_e/L} \frac{N}{D_p}\left(\frac{n_{ie}}{n_i}\right)^2 g\,\mathrm{d}\left(\frac{x}{L}\right)$$
$$= \text{constant} \times L. \quad (14)$$

Equation (14) predicts that $G_e$ increases linearly with $L$. Figure 8 shows the results of measurements of $G_e$ versus $L$. The best straight line fit is obtained with $L_{peff} = 1.5$ L.

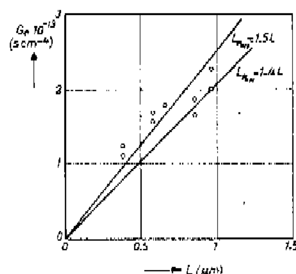



Fig. 8. Dependence of $G_e$ on the characteristic length ($L$) of the gaussian doping profile. The measured points are from LEC transistors.

3.4 *Minority carrier injection in $I^2L$ and LEC structures*

Wulms[19] used a structure as sketched in Fig. 9 to determine the electron current densities underneath the Al contact and oxide in a $p$-type base. He did so by varying the areas of the Al contact and the oxide and found at $V_{be} = 600\,mV$ $J_{Al} = 3.9 \times 10^{-2}$ A/cm$^2$ and $J_{ox} \approx 3.9 \times 10^{-3}$ A/cm$^2$. The impurity concentration in the base was $2 \times 10^{18}$ cm$^{-3}$, $D_n = 5$ cm$^2$/s, the basewidth was $W = 2\,\mu$m. At this doping level $L_{peff} = 17\,\mu m$, so the base is rather transparent for electrons ($g(x) = 1$). Because eqn (3) is generally valid for minority carrier injection into heavily doped regions we can apply it here to the electron current in the base and get

$$G_b = \frac{qn_{io}^2 \exp(xV_{be})}{J} = N_A\left(\frac{n_{io}}{n_i}\right)^2 \frac{1}{D_n}\left(W + \frac{D_n}{s}\right). \quad (15)$$

If we apply eqn (15) to $J_{ox}$, we find $s_{ox} \approx 5000$ cm/s.

If we substitute the value of $J_{Al}$ in eqn (15) we find $W + D_n/s = W = 2\,\mu$m, from which we can only conclude that $D_n/s \ll W$ or $s \gg 2.5 \times 10^4$ cm/s.

We carried out similar experiments on npn LEC transistors. In that case hole currents in an $n$-type emitter are involved. The impurity concentration in the $n$-type emitter was much lower than the $p$-type base of the $I^2L$ structure ($N_D = 1 \times 10^{16}$ cm$^{-3}$). Nevertheless we also found $s_{ox} = 3 \times 10^3$ to $1 \times 10^4$ cm/s and $s_{Al} > 1 \times 10^5$ cm/s. We see that the values for $s_{Al}$ found here are not in contradiction with the value given in Section 3.2, but the value for $s_{ox}$ seems to be too high for $n$-type material with

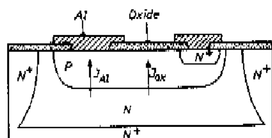



Fig. 9. Basic $I^2L$ structure for the determination of the current densities $J_{ox}$ and $J_{Al}$.

a < 111 > orientation. This can be attributed to the fact that recombination in the space charge region is included (see also Section 4).

### 3.5 *The Polysil emitter*

The polysil emitter[13] consists of a very thin ($X_\mu \sim 0.1\,\mu$m) monocrystalline region, covered with a polysilicon layer. If we estimate the surface concentration of the impurities in the monocrystalline part to be $1\times10^{20}$ cm$^{-3}$, then the integral in eqn (3) is

$$\int_0^{W_e} \frac{N(x)}{D_p(x)}\frac{n_{io}^2}{n_i^2(x)}\,g(x)\,dx = 3\times10^{12}\,s\,\text{cm}^{-4}.$$

The measured value of $G_e$ for polysil emitters is however $1$–$4\times10^{14}$ $s$cm$^{-4}$, so the contribution of the monocrystalline region to $G_e$ is negligible and recombination takes place in the poly-silicon and/or at the interface region. If this recombination is characterized by a recombination velocity for the mono-poly silicon interface, this velocity is about 5000 cm/$s$, as can be deduced with the help of eqn (3).

## 4. DISCUSSION AND CONCLUSIONS

In the foregoing a figure of merit ($G_e$) for emitters is defined and a simplified model is developed for calculating this $G_e$. The model incorporates bandgap narrowing due to high doping concentrations, a doping-dependent lifetime as occurs in Auger recombination processes, a built-in electric field and a finite recombination velocity ($s$) at the interface of the Al contact and the silicon. The essential part of the model, the weighting function $g(x)$ is modelled analytically, but it is based on one-dimensional computer simulations. The feature of $g(x)$ is that the function is characterized by only one quantity, the diffusion recombination length ($L_{p,e}$). Because the doping profile is not constant a modified definition of $L_{p,e}$, was necessary. By normalizing the variable $x$ to $L_{peff}$, the function $g(x/L_{peff})$ can be made almost independent of the doping profile, although for shallow emitters ($W_e/L_{peff} \lesssim 0.5$) $g(x)$ does depend on $s$. For $W_e/L_{peff} \gtrsim 1.5$ the value of $s$ is unimportant because the minority carriers do not reach the contact. For $W/L_{peff} \lesssim 0.5$ the emitter becomes transparent and $s$ acquires a dominant influence (see Figs. 1 and 2). The highest values of $G_e$ for conventional emitters are reached if $W/L_{peff} \gtrsim 1.5$ and the diffusion depth is large ($W_e = 15\,\mu$m); the doping concentration must be lowered in that case in order to enlarge $L_{peff}$ also.

It has been shown that the formula for $G_e$ (eqn 3) can be applied to a variety of structures. The results of Section 3.3 can be generalized for other than LEC transistors and gaussian profiles. Thus we may state that $G_e$ increases more or less linearly with the $n^+$ emitter diffusion depth, provided that the surface concentration $N(o)$ is kept constant.

From the other experiments, described in Section 3, we determined the following values for the recombination velocities: at the interface Al contact/silicon, $s_{Al} \sim 3\times10^5$ cm/$s$, at the interface oxide/silicon, $s_{ox} \sim 5000$ cm/$s$ and at the interface poly/monosilicon, $s_{poly} \sim 5000$ cm/$s$.

We must remark that these $s$-values were obtained on the assumption that the non-ideality factor $m$ was equal to 1. Often the different base current components have $m$-values between 1.1 and 1.4. In that case the recombination velocities become voltage-dependent. Further we must stipulate that, e.g. $J_{ox}$ in Fig. 9 may also contain the recombination currents of the bulk and space charge region underneath the oxide. Expressed in terms of surface recombination they increase the value of $s_{ox}$.

Concerning the recombination velocity at the poly/monosilicon interface, its value may strongly depend on the impurity concentration of the monosilicon.

In general we must say that the values of recombination velocities at surfaces and interfaces should be used very carefully.

*Acknowledgements*—The authors wish to express their gratitude to J. W. A. Scholte who supplied them with the LEC transistors, and to J. A. Timmermans, who did some of the measurements.

## REFERENCES


1. R. J. Whittier and D. A. Tremere, *IEEE Trans. Electron Dev.* **ED-16**, 39 (1969).
2. H. C. de Graaff, *Philips Res. Repts.* **26**, 191 (1971).
3. S. M. Sze, *Physics of Semiconductor Devices*, p. 269. Wiley-Interscience, New York (1969).
4. H. K. Gummel, *Proc. I. R. E.* **49**, 834 (1961).
5. H. C. de Graaff and J. W. Slotboom, *Solid-St. Electr.* **19**, 809 (1976).
6. J. Lindmayer and Ch. Y. Wrigley, *Fundamentals of Semiconductor Devices*, Van Nostrand, New York (1965).
7. R. v. Overstraeten, R. Mertens and H. de Man, *Essderc, Munich*, paper B 10.3 (1973).
8. W. W. Sheng, *IEEE Trans. Electron Dev.* **ED-22**, 25 (1975).
9. H. J. J. de Man, *IEEE Trans. Electron Dev.* **ED-18**, 833 (1971).
10. D. P. Kennedy and P. C. Murley, IRE Trans **ED-9**, 136 (1962).
11. S. C. Choo, *Solid-St. Electr.* **15**, 11 (1972).
12. J. W. Slotboom, *Electron. Lett.* **5**, 677 (1969); and M. S. Mock, *Solid-St. Electr.* **16**, 1251 (1973).
13. J. Graul, A. Glasl and H. Murrmann, *Int. Electr. Dev. Meeting*, Washington, paper 20.2 (1975).
14. I. C. Irvin, *Bell Syst. Techn. J.* **41**, 387 (1962).
15. J. W. Slotboom, *Solid-St. Electron* **20**, 167 (1977).
16. E. H. Putley and W. H. Mitchell, *Proc. Phys. Soc. London*, **A72**, 193 (1958).
17. J. W. Slotboom and H. C. de Graaff, *Solid-St. Electron* **19**, 857 (1976).
18. J. D. Beck and R. Conradt, *Solid-State Commun.* **13**, 93 (1973).
19. H. E. J. Wulms, *IEEE Int. Solid-St. Circ. Conf.*, Philadelphia, paper THAM 9.1 (1976).


## APPENDIX

For the hole and electron current densities we write

$$J_p = -qD_p\frac{dp}{dx} + q\mu_p p\left(-\frac{d\Psi}{dx} + \frac{kT}{qn_i}\frac{dn_i}{dx}\right) \tag{16}$$

$$J_n = qD_n\frac{dN}{dx} + q\mu_n N\left(-\frac{d\Psi}{dx} - \frac{kT}{qn_i}\frac{dn_i}{dx}\right). \tag{17}$$

In these equations bandgap narrowing is included by the fact that $n_i$ is a function of $x$. Further we have put $n = N$, $N$ being the net doping concentration. This is allowed under quasi-neutrality and low injection conditions. In equilibrium $J_n = 0$ and

$$-\frac{d\Psi}{dx} = \frac{kT}{q}\left(\frac{1}{n_i}\frac{dn_i}{dx} - \frac{1}{N}\frac{dN}{dx}\right). \tag{18}$$

The right-hand side of eqn (18) is the built-in electric field, modified by bandgap narrowing effects. In non-equilibrium

situations we assume that the field in the heavily doped emitter region, given by eqn (18), will not change significantly and we can write

$$J_p = -qD_p \frac{dp}{dx} + qD_p p \left( \frac{2}{n_i} \frac{dn_i}{dx} - \frac{1}{N} \frac{dN}{dx} \right) \quad (19)$$

or

$$\frac{dp}{dx} + p \frac{d}{dx} \left[ \ln \left\{ \frac{N(x)}{n_i^2(x)} \right\} \right] = -\frac{J_p(x)}{qD_p} \quad (20)$$

The differential eqn (20) is valid in the neutral emitter region and for low injection conditions. These are not severe restrictions for normal emitters. The general solution of eqn (20) is:

$$p(x) = \frac{n_i^2(x)}{N(x)} \left\{ \text{const} + \frac{1}{q} \int_0^x \frac{N(t)}{D_p(t)} \frac{J_p(t)}{n_i^2(t)} dt \right\}. \quad (21)$$

The boundary conditions are:

$$p(0) - p_o \approx p(0) = \frac{J_p(0)}{qs} \quad (22a)$$

and

$$p(W_e) = \frac{n_i^2(W_e)}{N(W_e)} \exp(qV_{be}/kT) \quad (22b)$$

From (21) and (22a) it follows that

$$-J_p(W_e) = \frac{qn_{io}^2 \exp(qV_{be}/kT)}{\frac{N(o)}{s}\left(\frac{n_{io}^2}{n_i^2(o)}\right) \cdot \frac{J_p(o)}{J_p(W_e)} + \int_0^{W_e} \frac{N(x)}{D_p(x)} \frac{J_p(x)}{J_p(W_e)} dx} \quad (23)$$

which proves eqn (3).

# 5.8. "Some aspects of LEC transistor behaviour"

Solid-State Electr., Vol. 19, p. 809, 1976.

## SOME ASPECTS OF LEC TRANSISTOR BEHAVIOUR

H. C. DE GRAAFF and J. W. SLOTBOOM
Philips Research Laboratories, Eindhoven, The Netherlands

(*Received* 12 *November* 1975; *in revised form* 19 *February* 1976)

**Abstract**—Some properties of bipolar transistors with low emitter concentrations are investigated both theoretically and experimentally. It turns out that the base current of LEC transistors at medium and high injection levels is the same as in double diffused transistors and can be explained by Auger recombination in the emitter $n^+$ region. The cut-off frequency $f_T$ is rather low, due to extra charge storage in the lightly doped emitter region. Small $n^+$ emitter areas, surrounded by a $p$-ring may introduce anomalies such as kinks in the ($I_{ce}$, $V_{ce}$) characteristics and negative resistances.

### 1. INTRODUCTION

A new bipolar transistor structure with Low Emitter Concentrations (LEC transistor) was first reported in the literature by Yagi *et al.*[1]. The emitter region of these structures consisted of an $n^+n$ barrier whose purpose was to reflect the minority carriers (holes), injected from the base, thus keeping the base current low and making the current gain high.

We investigated similar structures, whose cross-section and doping profile are given in Fig. 1. The lightly doped emitter region was an epitaxial layer in all cases; the base was either a buried layer made by ion implantation, or another epitaxial layer. Several layer thicknesses and impurity concentrations were used for both emitter and base regions: $N_e$, $N_b \sim 10^{15}$–$10^{17}\,\mathrm{cm}^{-3}$ and $W_e$, $W_b \sim 1$–$5\,\mu\mathrm{m}$. The emitter $n^+$ region was made by a normal phosphorus diffusion (surface concentration $\sim 10^{21}\,\mathrm{cm}^{-3}$). In this paper we report the results of our investigations, with special emphasis on base current and current gain and $f_T$ behaviour. A short treatment will also be given of some anomalies which arise when the emitter $n^+$ area is smaller than the total emitter area and/or a $p$-type ring around the $n^+$ area is present. Throughout this paper saturation effects in the collector are excluded.

### 2. CHARGE STORAGE AND COLLECTOR CURRENT

Apart from the normal charge storage in the base, which can be written as

$$Q_b = Q_{b0} + \Delta Q_b$$

and

$$\Delta Q_b = \frac{1}{4} q N_b W_b \left\{ -1 + \sqrt{\left[ 1 + \left( \frac{2n_i}{N_b} \right)^2 \exp(\gamma V_{be}) \right]} \right\} \tag{1}$$

there is also a charge storage in the lightly doped emitter region. At medium injection levels ($400\,\mathrm{mV} < V_{be} < 700\,\mathrm{mV}$) it is given by:

$$\Delta Q_e = \frac{1}{2} q N_e W_e \left\{ -1 + \sqrt{\left[ 1 + \left( \frac{2n_i}{N_e} \right)^2 \exp(\gamma V_{be}) \right]} \right\}. \tag{2a}$$

In these equations $\gamma = (q/kT)$. They are derived in the Appendix. The charges are given for $1\,\mathrm{cm}^2$ of emitter area. The charge distributions of $\Delta Q_e$ and $\Delta Q_b$ are sketched in Fig. 2. At high injection levels in the $n^+$ emitter layer the minority carrier distribution will show a negative gradient $(\mathrm{d}p/\mathrm{d}x) = (J_c/2qD_n)$. The eqn (2a) is no longer valid and has to be replaced by

$$\Delta Q_e = q n_i W_e \left\{ 1 - \frac{(2W_e + W_b) - \sqrt{(W_b^2 + 4W_e W_b)}}{4W_e} \right\} \exp\left( \frac{1}{2} \gamma V_{be} \right), \tag{2b}$$

see Appendix. From eqns (1) and (2) it will be clear that the stored charges increase with $\exp(\gamma V_{be})$ at low and medium injection levels, whereas at high injection levels they increase with $\exp(\frac{1}{2}\gamma V_{be})$. One should bear in mind that high injection conditions can exist in the emitter layer together with low injection conditions in the base and vice versa, depending on the impurity concentrations $N_e$ and $N_b$.

The collector current is written as

$$I_c = \frac{q^2 D_n n_i^2 A_e}{\Delta Q_e + Q_b} \exp(\gamma V_{be}). \tag{3}$$

Under high injection conditions $(\Delta Q_e + Q_b) \gg Q_{b0}$ and $I_c$ increases with $\exp(\frac{1}{2}\gamma V_{be})$.

### 2. THE BASE CURRENT

Figure 3 shows a one-dimensional computer simulation of the total recombination within the device. As recombination mechanisms we have taken a Shockley–Read–Hall process with a constant life time and an Auger process:

$$R = \frac{pn - n_i^2}{\tau_0(p + n + 2n_i)} + C_A(p + n)pn \tag{4}$$

with $\tau_0 = 10^{-6}\,\mathrm{s}$ and $C_A = 1.5 \cdot 10^{-31}\,\mathrm{cm}^6\,\mathrm{s}^{-1}$[2]. We can distinguish four components:

1. Recombination in the emitter-base depletion layer.

$$I_1 = I_{10} \exp(\tfrac{1}{2}\gamma V_{be}). \tag{5}$$

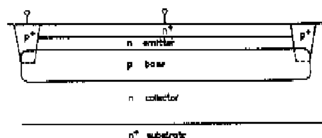



Fig. 1a. Cross-section of the LEC transistor.

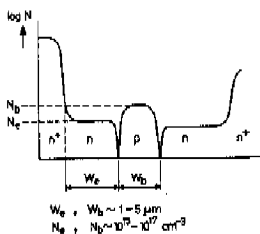



Fig. 1b. Doping profile.

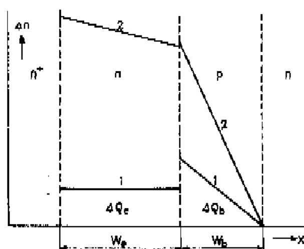



Fig. 2. Sketch of the distribution of the excess electron concentration $\Delta n = n - n_0$. The integral of $\Delta n$ is the total stored charge. Curve 1: low injection. Curve 2: high injection.

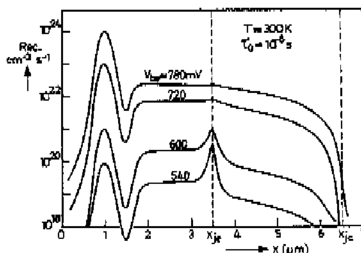



Fig. 3. Computer simulation of the recombination. Recombination mechanisms are SRH and Auger processes.

This component is of importance only at low injection levels.

2. Recombination in the lightly doped emitter region of the stored minority carrier charge,

$$I_2 = A_s \cdot \frac{\Delta Q_e}{\tau_e}. \tag{6}$$

3. Same as sub. 2., but now in the base region,

$$I_3 = A_s \cdot \frac{\Delta Q_b}{\tau_b}. \tag{7}$$

4. Auger recombination in the $n^+$ emitter region,

$$I_4 = \frac{q^2 D_p n_{ie}^2 \cdot A_e}{Q_{e0}} \exp(\gamma V_{be}/m) \tag{8}$$

where $n_{ie^+}$ = effective intrinsic concentration in the $n^+$ emitter. $Q_{e0}$ = effective impurity charge in the $n^+$ emitter.

$Q_{e0}$ can be written as $qN^+L_{p,eff}$, where the value of the effective recombination length $L_{p,eff}$ is determined by the doping profile of the $n^+$ emitter region and the Auger constant $C_A$. The non-ideality factor $m$ is approximately equal to 1, as can be seen from the voltage dependence of the maxima in Fig. 3. Physically speaking the Auger component is brought about by a recombination process with an impurity dependent lifetime[2]. This effect is enhanced by an increased minority carrier concentration due to bandgap narrowing[3, 4].

Surface recombination and extra recombination at the $n^+n$ interface[5, 6] are omitted in the computer simulation; the latter has the same voltage dependence as the $I_2$-component. This means that this component would also increase with $\exp(\gamma V_{be})$ at low and medium injection levels, but with $\exp(\frac{1}{2}\gamma V_{be})$ at high injection. This is not found experimentally, see Fig. 4. In this figure measured

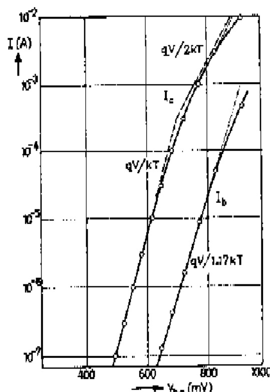



Fig. 4. Measurements of $I_c$ and $I_b$ versus $V_{be}$. The dotted lines are corrected for internal ohmic voltage drops, the dashed lines give the asymptotes.

values of $I_c$ and $I_b$ versus $V_{be}$ are plotted for one of our LEC structures. If $V_{be}$ is corrected for ohmic voltage drops across the series resistances we can clearly see that even under high injection conditions ($V_{be} \geq 700$ mV, $I_c \sim \exp(\frac{1}{2}qV_{be})$) $I_b = I_s \sim \exp(qV_{be}/1.17\,kT)$.

We conclude from these experiments that recombination at the $n^+n$ interface can be neglected and also that the lifetimes $\tau_n$ and $\tau_p$ are sufficiently high to keep the currents $I_2$ and $I_3$ small. This leaves the Auger component $I_4$ as dominant for the base current at medium and high injection levels, although the value $m = 1.17$ indicates that the other components are still present.

In Fig. 5 the base current density of several LEC transistors is compared with that of a typical double diffused transistor. The $n^+$ emitters of the LEC and double diffused transistors were made by the same process, except the one indicated by open squares. It is obvious that at medium and high level injection the base current density of an LEC transistor is not smaller than that of a double diffused one. Other evidence indicating that the base currents have normal values in LEC transistors is that the "Gummel number in the emitter", defined as $G_e = h_{FE\,max} \times G_b$ ($G_b$ = Gummel number in the base) has values of $1-3\times10^{13}$ scm$^{-4}$. The same values are found for double diffused transistors. It should be noted that the $I_1$ component (eqn (5)) in some LEC transistors is very small, which can be ascribed to an improved lifetime in the emitter-base junction. In that case the Auger component $I_4$ is still dominant, even at low injection levels (see Fig. 5, black squares).

## 4. THE CURRENT GAIN $h_{FE}$

Figure 6 shows a typical example of the current gain $h_{FE}(I_c)$ for two different temperatures. The $h_{FE}$ fall-off at high current is due to high injection (see Fig. 4). From

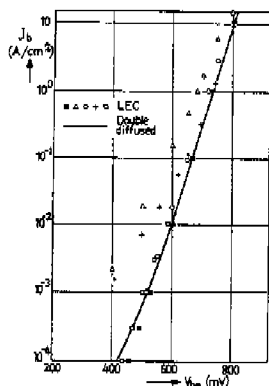



Fig. 5. Measured values of base current density of several LEC transistors in comparison with a double diffused transistor.

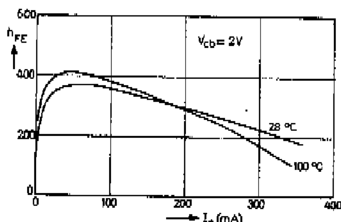



Fig. 6. Current gain $h_{FE}$ versus collector current $I_c$, at two different temperatures, for a typical LEC transistor.

eqns (3) and (8) it can be deduced that

$$h_{FE\,max} \simeq \frac{Q_{e0}}{Q_{b0}} \cdot \frac{D_n}{D_p} \left(\frac{n_i}{n_{ie^+}}\right)^2. \qquad (9)$$

In eqn (9) $D_n/D_p$ and $(n_i/n_{ie^+})^2$ are temperature-dependent;

$$\frac{D_n}{D_p} \sim T^{-\alpha} \quad \text{and} \quad \left(\frac{n_i}{n_{ie^+}}\right)^2 \sim \exp\left(\frac{-q\Delta V_{gap}}{kT}\right),$$

where $\Delta V_{gap}$ = bandgap difference between base and emitter. From eqns (1), (2b), (3) and (8) we can also deduce that at high injection levels

$$h_{FE}\,(\text{high inj.}) \simeq \frac{4Q_{e0}}{(W_b+2W_e)^2}\frac{D_n^2}{D_p}\left(\frac{n_i}{n_{ie^+}}\right)^2\frac{1}{I_c}. \qquad (10)$$

If we put also $(D_n^2/D_p) \sim T^{-\alpha}$, then in both cases the temperature dependence of the $h_{FE}$ can be written in the form

$$\frac{T}{h_{FE}}\frac{dh_{FE}}{dT} = -\alpha + q\Delta V_{gap}/kT. \qquad (11)$$

With a base dope $N_b = 10^{16}$ cm$^{-3}$ the values of $\alpha$ are 2.6 for the factor $D_n/D_p$ and 3.8 for the factor $D_n^2/D_p$ [7], whereas $(T/h_{FE})(dh_{FE}/dT)$ is slightly positive for $h_{FE\,max}$, and slightly negative for $h_{FE}$ at high injection. From this it follows that $\Delta V_{gap} \simeq 80$ mV, which is about the same as found elsewhere[8, 9].

## 5. $f_T$ BEHAVIOUR

The maximum $f_T$ is mainly determined by the charge storage in the lightly doped emitter and the base. We can distinguish between two extreme cases:

1. Low injection conditions both in emitter and base. With the help of eqns (1), (2a) and (3) we can write down the delay $\tau_1$ as

$$\tau_1 = A_e\frac{d(\Delta Q_e + \Delta Q_b)}{dI_c} = G_b\left(\frac{W_e}{N_e} + \frac{\eta - 1 + e^{-\eta}}{\eta^2}\cdot\frac{W_b}{N_b}\right) \qquad (12)$$

where $G_b$ is the Gummel number in the base [$G_b$ = $(W_bN_b/D_n)$] and $\eta$ is a measure of the built-in drift field in the base[10].

2. High injection conditions in emitter and base. The delay is given by

$$\tau_h = \frac{1}{16 D_n} (2W_e + W_b + \sqrt{[W_b^2 + 4W_e W_b]})^2 \qquad (13)$$

which follows from eqns (1), (2b) and (3) by eliminating $V_{be}$ and differentiating the charge with respect to $I_c$.

By disregarding the influence of the depletion capacitances (expressed by the term $(kT/qI_c)(C_e + C_c)$) for a moment, we can say that, if $\tau_h > \tau_f$, the $f_T$ falls off with increasing current due to high injection and the $f_{T\max}$ is determined by the value of $\tau_f$. If $\tau_h < \tau_f$ the $f_T$ increases with current, even under high injection conditions. These two different kinds of $f_T$ behaviour are illustrated in Fig. 7, which gives a computer simulation of two selected examples. Table 1 gives a comparison with measurements. With $W_e$ and $W_b$ a few microns, $N_e$ and $N_b \sim 10^{15}$–$10^{16}$ cm$^{-3}$, the calculated transit times in Table 1 correspond to maximum $f_T$ values of 30–100 MHz.

### 6. SOME ANOMALIES

Some anomalies may occur if the area of the $n^+$ region in the emitter is made small with respect to the total emitter area and a ring of $p$-type material around the $n^+$ region is present (see Fig. 8a). In the first place the $p$-ring may act as the collector of a parasitic $pnp$ transistor, made up of the $p$-ring, the lightly doped emitter under this ring as base and the base region as emitter. In the second place the lateral resistances in the emitter and base regions, which in some of our samples were of the order of magnitude of 1 $k\Omega$, can no longer be neglected. Moreover, the lateral base resistance may be modulated if the collector-base junction becomes forward biased. All

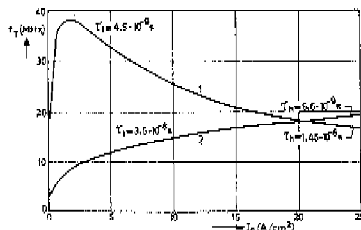



Fig. 7. Computer simulation of two quite different examples of $f_T$ behaviour. Curve 1: $W_e = 10\ \mu$m, $W_b = 5\ \mu$m, $N_e = 2 \times 10^{15}$ cm$^{-3}$, $N_b = 1 \times 10^{15}$ cm$^{-3}$. Curve 2: $W_e = 2\ \mu$m, $W_b = 5\ \mu$m, $N_e = 1 \times 10^{15}$ cm$^{-3}$, $N_b = 1 \times 10^{16}$ cm$^{-3}$.

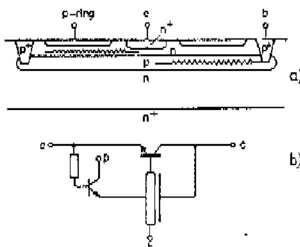



Fig. 8. (a) Cross-section of an LEC transistor with a small $n^+$ emitter region and a $p$-type ring surrounding the $n^+$ region. The large lateral resistances are also indicated. (b) Equivalent circuit for the structure given in (a).

these effects are represented by the equivalent circuit in Fig. 8b. Figure 9 shows the $I_c$ and $I_b$ versus $V_{be}$ for an LEC transistor with the $n^+$ region area 100 times smaller than the total emitter area. The peculiar kink in the characteristics can be modelled by considering two

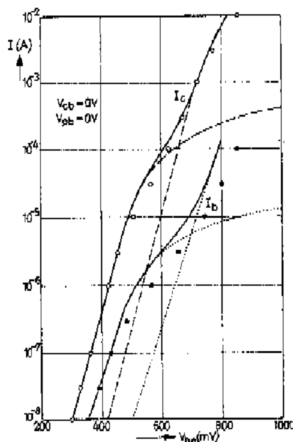



Fig. 9. Measurements of $I_c$ and $I_b$ vs $V_{be}$ for an LEC transistor with a small $n^+$ area. The fully drawn lines are the results of the two-transistor model.

Table I.

| sample | quantity | calculated with eqs(12) and (13) | measured from $f_{T\max}$ |
|---|---|---|---|
| 8 | $\tau_f$ | $1.8 \times 10^{-9}$ s | $1.2$–$1.6 \times 10^{-9}$ |
| E16 | $\tau_h$ | 5.2 | 4.0 |
| K35 | $\tau_h$ | 1.8 | 3.2 |
| RI | $\tau_h$ | 3.2 | 3.3 |
| LECP | $\tau_f$ | 3.0 | 3.0 |

transistors in parallel, one of which has a large (lateral) emitter series resistance of 1 kΩ. This means that for $V_{be} \geq 750$ mV only the region under the $n^+$ emitter is active in this LEC structure. The base currents of the two transistors in parallel are assumed to be Auger recombination currents with a non-ideality factor $m$ of 1.22. When a negative voltage is applied to the $p$-ring, the $pnp$ transistor becomes active; the $p$-ring (its collector) draws a large current, which must be supplied via the base. This gives an extra voltage drop across the large lateral base resistance, thus increasing the $V_{be}$. The results on the $(I_c, V_{be})$ characteristics are shown in Fig. 10. It is obvious that also the $h_{FE}$ is influenced by the voltage of the $p$-ring.

If the $p$-ring is floating ($I_p = 0$) its potential is influenced somewhat by the collector-base junction in forward bias: when this forward bias is reduced, the $p$-ring potential becomes slightly more negative. This influence is sufficient for the existence of a region with negative resistance in the $(I_c, V_{cb})$-characteristics (see Fig. 11), provided $V_{be}$ is kept constant.

## 7. CONCLUSIONS

From the foregoing the following conclusions can be drawn. Auger recombination in the $n^+$ region of the emitter dominates the base current at medium and high injection levels. The base current density in LEC structures is the same as in double diffused ones and the

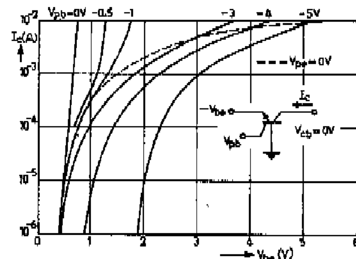



Fig. 10. Measurements of $I_c$ vs $V_{be}$ for an LEC transistor with $p$-ring, showing the influence of the $p$-ring potential on the characteristics.

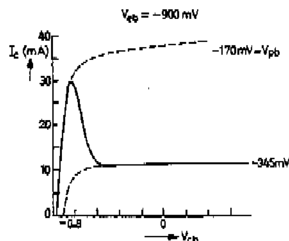



Fig. 11. Negative resistance phenomena with $V_{be}$ = constant and floating $p$-ring. The $p$-ring potential varies between −170 and −345 mV.

$n^+n$ barrier in the emitter has the same reflecting properties as the $n^+$ emitter of a conventional double diffused transistor. The lightly doped region offers no advantage in this respect. However, LEC structures may help to keep the recombination in the emitter-base depletion layer low. The cut-off frequency $f_T$ is relatively low due to the extra charge storage in the lightly doped emitter layer. The behaviour of $f_T$ versus $I_c$ depends on the ratio of the delay times at low and high injection levels. Finally, LEC transistors with $p$-rings and small $n^+$ emitters show some anomalies in their characteristics (kinks and negative resistance ranges).

*Acknowledgements*—The authors wish to express their gratitude to Mr. A. Schmitz and Mr. J. Scholte, who supplied the samples discussed in this paper. They would also like to thank Dr. P. Hart, Mr. W. Smulders and Mr. J. Timmermans for helpful discussions.

## REFERENCES

1. H. Yagi, T. Tsuyuki, K. Koma and Y. Miyazawa, *International Electron Devices Meeting*, Washington D.C. (1974).
2. J. D. Beck and R. Conradt, *Solid-St. Commun.* 13, 93 (1973).
3. R. J. van Overstraeten, H. J. de Man and R. P. Mertens, *IEEE Trans. Electron Devices* ED-20, 290 (1973).
4. J. W. Slotboom and H. C. de Graaff, *5th Eur. Solid St. Device Res. Conf.*, Grenoble (1975).
5. R. W. Dutton and R. J. Whittier, *IEEE Trans. Electron Devices* ED-16, 458 (1969).
6. F. M. Klaassen, *IEEE Trans. Electron Devices* ED-22, 145 (1975).
7. W. W. Gärtner, *Semicond. Prod.* 29 (1960).
8. W. L. Kauffman and A. A. Bergh, *IEEE Trans. Electron Devices* ED-15, 732 (1968).
9. D. Buhanan, *IEEE Trans. Electron Devices* ED-16, 117 (1969).
10. J. Lindmayer and Ch. Y. Wrigley, *Fundamentals of Semiconductor Devices*, van Nostrand (1965).

## APPENDIX

The charge, stored in the neutral base region, is

$$\Delta Q_b = \frac{1}{2} q n_b W_b. \tag{14}$$

$n_b$ is the minority carrier concentration at the edge of the depletion layer:

$$n_b(n_b + N_b) = n_i^2 \exp(\gamma V_{be}). \tag{15}$$

Solving this quadratic equation for $n_b$ and substituting into the expression for $\Delta Q_b$ gives eqn (1).

From the current equations for $J_n$ and $J_p$ and the charge neutrality condition we can derive for the minority carrier gradient in the lightly doped emitter region:

$$\frac{dp}{dx} = \frac{J_p - \frac{\mu_p}{\mu_n}\left(1 + \frac{N_e}{p}\right) J_n}{q D_p \left(2 + \frac{N_e}{p}\right)}. \tag{16}$$

For $p \ll N_e$ ($V_{be} < 400$ mV) is $(dp/dx) > 0$, because both $J_n$ and $J_p$ are negative. For $p \gg N_e$ ($V_{be} > 700$ mV) is $(dp/dx) < 0$ and in between is $(dp/dx) = 0$. In that case we can write $\Delta Q_e = q p_e W_e$ and $p_e(p_e + N_e) = n_i^2 \exp(\gamma V_{be})$. Elimination of $p_e$ leads to eqn (2a). For $p \gg N_e$ we can write $(dp/dx) = (J_n/2qD_n)$ or for the decrease in carrier concentration:

$$\Delta p = \frac{J_n W_e}{2 q D_n}. \tag{17}$$

The stored charge $\Delta Q_e$ then is

$$\Delta Q_e = qW_e\left\{n_i \exp\left(\frac{1}{2}\gamma V_{be}\right) - \frac{1}{2}\Delta p\right\}. \tag{18}$$

From eqn (3) it follows that

$$J_c = \frac{q^2 D_n n_i^2 \exp(\gamma V_{be})}{\Delta Q_e + \Delta Q_n}$$

and from this

$$\frac{\Delta p}{n_i} = \frac{(2W_e + W_n) - \sqrt{(W_n^2 + 4W_e W_n)}}{2W_e}\exp\left(\frac{1}{2}\gamma V_{be}\right). \tag{19}$$

Substitution of $\Delta p$ into eqn (18) gives eqn (2b).

## SUMMARY

The investigations and publications in this thesis concern the electrical behaviour of bipolar transistors in relationship to such process and material dependent quantities as the profile of the impurities and the mobility and lifetime of the carriers. These quantities appear in Poisson's equation and the continuity equations for the holes and electrons, which govern the electrical behaviour. This is discussed in chapter 1. Because these differential equations are nonlinear, rigorous solutions require numerical methods. This offers the opportunity to take into account realistic data about the doping profile, mobilities and lifetimes and results in an accurate description of the electrical behaviour in quantitative terms.

In 1964 Gummel published a numerical method for solving the one-dimensional equations (see chapter 2). In order to be able to deal also with two-dimensional transistor structures, we have developed a numerical method, which is presented in the first two publications reprinted in chapter 5. Non-uniform current density distributions, such as current crowding due to lateral voltage drop (see chapter 2), or spreading effects in very small devices, can be studied in this way. When the results of these calculations are compared with the results of experiments, the behaviour of the collector current as a function of the applied voltages agrees rather well; the description of the base current, on the other hand, is very unsatisfactory.

In modern bipolar transistors the base current is usually determined by the minority carrier injection into the emitter. Experiments and calculations, discussed in chapter 3 and 4, show that, as a result of the high impurity concentration in the emitter, this injection is strongly affected by two phenomena. First, the minority carrier lifetime decreases strongly at higher impurity concentrations. In recent publications of lifetime measurements in silicon this is ascribed to Auger recombination (see chapter 1, fig. 1.8). Secondly, the pn-product increases strongly at higher impurity concentrations (e.g. 70 times at an impurity concentration of $2.3 \cdot 10^{19}$ $cm^{-3}$).

Chapter 3 describes in detail how the pn-product in the base region was determined by using the transistor itself as a measuring vehicle. Measurements and calculations indicate that the pn-product at higher impurity concentrations is exponentially related (as it is in the case of pure silicon) with a quantity that can be interpreted as the bandgap extrapolated towards zero degree Kelvin. An increase in pn-product corresponds to a decrease in bandgap. Although bandgap narrowing has been discussed in the literature for many years, quite considerable discrepancies exist between different theoretical calculations and disagreement also exists with optical absorption measurements. From our measurements of the pn-product it follows that bandgap narrowing is larger and starts at lower impurity concentrations than is derived from the optical absorption measurements. The reason for this is not yet clearly known (see the discussion at the end of chapter 3). The measurements of both effects can be characterized by simple analytical approximations.

We developed a one-dimensional numerical transistor program including these effects, and that has a number of improvements in the algorithm (chapter 2). From the calculated recombination it appears rather unexpectedly that, apart from the well-known recombination maximum in the emitter-base junction, a second and often more important peak in the recombination appears in the heavily doped region of the emitter. *This is caused by the interaction of the above-mentioned phenomena:* on the one hand the strong reduction in lifetime and on the other hand the relatively large and only gradually decreasing minority carrier concentration. It is not possible to discriminate between each of these effects from current gain measurements alone. Our computer simulations do not only describe well these current gain measurements, but agree also with all kinds of other experiments, such as the measurement of the base current at low emitter-base forward voltages, the decrease of current gain when small layers of emitter material are etched away, the current gain characteristics as a function of temperature, etc.. If alternative models for the lifetime and bandgap narrowing are substituted in these simulations, the results do not agree with all of these measurements at the same time.

Using the detailed output data of these calculations, analytical approximations for the base current are discussed in chapter 4 and applied for different emitter structures, e.g. the LEC transistor, which has an extra, low-doped region in the emitter.

The results described in this thesis are not only important for the analysis of bipolar transistors but are also directly applicable in the analysis of many other silicon devices.

## SAMENVATTING


De onderzoekingen en publicaties die in dit proefschrift worden besproken hebben tot doel een beter inzicht te geven in het electrisch gedrag van bipolaire transistoren en hoe dit samenhangt met proces- en materiaalafhankelijke grootheden, zoals het concentratieprofiel van de verontreinigingsatomen en de beweeglijkheid en levensduur van de ladingsdragers. Deze grootheden komen voor in de vergelijking van Poisson en de beide continuïteitsvergelijkingen voor de gaten en electronen, die het electrisch gedrag beheersen. Dit wordt besproken in hoofdstuk 1. Aangezien deze differentiaalvergelijkingen niet-lineair zijn, zijn voor het verkrijgen van nauwkeurige oplossingen numerieke methoden vereist. Dit biedt tevens de mogelijkheid experimentele gegevens over het concentratieprofiel van de verontreinigingsatomen en de beweeglijkheid en levensduur van de ladingsdragers, zonder vereenvoudigingen, in rekening te brengen. Hierdoor wordt een zo goed mogelijk quantitatief beeld van het electrisch gedrag verkregen.

In 1964 publiceerde Gummel een methode waarmee de vergelijkingen, zij het eendimensionaal, numeriek kunnen worden opgelost (zie hoofdstuk 2). Teneinde ook tweedimensionale berekeningen aan transistoren uit te kunnen voeren hebben we een numerieke methode ontwikkeld, die wordt beschreven in de eerste twee publicaties van hoofdstuk 5. Hiermee kunnen niet-uniforme stroomdichtheidsverdelingen, zoals bijv. stroomverdringing tengevolge van laterale spanningsverschillen (zie hoofdstuk 2), of spreidingseffecten in zeer kleine devices worden onderzocht. Wanneer we de resultaten van deze numerieke berekeningen vergelijken met experimenten, dan blijkt in het algemeen, dat het gedrag van de collectorstroom als functie van de aangelegde spanningen goed overeenstemt, terwijl daarentegen het gedrag van de basisstroom alleen met behulp van allerlei, experimenteel niet verifieerbare, aannamen is te verklaren en dan nog slechts gebrekkig.

In moderne bipolaire transistoren wordt de basisstroom meestal bepaald door injectie van minderheidsladingsdragers in de emitter. Experimenten en berekeningen, besproken in de hoofdstukken 3 en 4, tonen aan dat hierbij, als gevolg van de hoge concentratie van verontreinigingsatomen in de emitter, twee effecten een belangrijke rol spelen. In de eerste plaats neemt de levensduur van de minderheidsladingsdragers bij hogere concentraties van verontreinigingsatomen sterk af. In recente publicaties van levensduurmetingen in silicium wordt dit toegeschreven aan Auger-recombinatie (zie hoofdstuk 1, fig. 1.8). Ten tweede blijkt bij hoge concentraties het pn-product aanzienlijk toe te nemen (bijv. 70 maal bij een concentratie van $2.3 \cdot 10^{19}$ cm$^{-3}$).

In hoofdstuk 3 wordt uitvoerig uiteengezet hoe wij er in geslaagd zijn het pn-product in de transistor zelf, met name in het basisgebied, te bepalen voor een aantal transistoren met verschillende concentraties van verontreinigingsatomen in de basis.

Metingen en berekeningen tonen aan dat het pn-product ook bij hoge concentraties van verontreinigingsatomen (geheel analoog aan het geval van zuiver silicium) exponetieel gerelateerd is aan een grootheid die geïnterpreteerd kan worden als de naar nul graden Kelvin geëxtrapoleerde bandafstand. Een toename van het pn-product komt dan overeen met een afname van de bandafstand. Hoewel bandvernauwing reeds jarenlang onderwerp van discussie in de literatuur is, bestaan er grote verschillen tussen berekeningen onderling, terwijl optische absorptiemetingen er niet mee overeenstemmen. Uit onze metingen van het pn-product volgt dat de bandvernauwing veel groter is en bij lagere concentraties begint dan uit optische absorptiemetingen kan worden afgeleid. Een duidelijke verklaring hiervoor is nog niet gevonden (zie de discussie aan het eind van hoofdstuk 3). Beide gemeten effecten kunnen worden gekarakteriseerd door eenvoudige analytische benaderingen.

We hebben een eendimensionaal numeriek transistorprogramma ontwikkeld (hoofdstuk 2), waarin deze effecten zijn meegenomen en dat een aantal belangrijke verbeteringen in de oplosmethode bevat. Uit de berekeningen blijkt tamelijk onverwacht dat de recombinatie, behalve het bekende maximum in de emitter-basis-junctie, een tweede en meestal veel belangrijker recombinatiepiek kan vertonen in het ladingsneutrale gedeelte van de emitter. *Dit is het gevolg van beide effecten tezamen:* de sterke afname van de levensduur enerzijds en de relatief hoge concentratie en geleidelijke afval van de minderheidsladingsdragers anderzijds. Het is niet mogelijk om, alleen op grond van het verloop van de stroomversterking als functie van de collectorstroom, beide effecten te onderscheiden. Onze computersimulaties blijken niet alleen dit verloop goed te beschrijven maar ook overeen te stemmen met allerlei andere experimenten, zoals de meting van de basisstroom bij lage emitterbasisspanningen, de afname van de stroomversterking bij wegetsen van dunne laagjes emitter-materiaal, de stroomversterking als functie van de temperatuur, etc.. Indien echter in deze berekeningen andere modellen voor de levensduur en de bandvernauwing als functie van de verontreinigingsconcentratie worden gekozen, zijn de resultaten niet goed in overeenstemming met al deze metingen tezamen.

Uitgaande van de gedetailleerde gegevens van deze berekeningen worden in hoofdstuk 4 analytische benaderingen voor het gedrag van de basisstroom beschreven en toegepast op verschillende emitterstructuren, o.a. de LEC transistor, waarbij het emittergebied, behalve het normale gebied met een hoge concentratie van verontreinigingsatomen, tevens een gebied bevat met lage concentratie.

De resultaten die in dit proefschrift worden beschreven zijn niet alleen van belang voor bipolaire transistoren maar ook direct toepasbaar bij de analyse van vele andere silicium devices.

STELLINGEN

J.W. Slotboom

25 oktober 1977

I

Bij injectie van minderheidsladingsdragers in zwaargedoteerd silicium, zoals bijv. gateninjectie in een $n^{+}p$ diode, bevinden zich de meeste geïnjecteerde ladingsdragers in een dun gebied grenzend aan de junctie. Gewoonlijk wordt aangenomen dat buiten deze "actieve" laag het aantal minderheidsladingsdragers zo klein is, dat het vrijwel geen bijdrage geeft tot de totale recombinatiestroom. In veel gevallen is deze aanname onjuist.

o.a. S.C. Choo, Solid-St. Electr., Vol. 15, p. 11, 1972.
F.A. Lindholm, S.S. Li, C.T. Sah, Record 11th Photovoltaic Specialists Conf., p. 3, 1975.
F.A. Lindholm, A. Neugroschel, C.T. Sah, M.P. Godlewski, N.W. Brandhorst, IEEE Transact. El. Dev., Vol. ED-24, p. 402, 1977.
A. Neugroschel, F.A. Lindholm, C.T. Sah, IEEE Transact. El. Dev., Vol. ED-24, p. 662, 1977.

II

Kannam heeft waargenomen dat de stroomversterking van bipolaire silicium transistoren toeneemt na een warmtebehandeling gedurende 16 uur op 800 °C in een stikstof omgeving. Hij schrijft dit toe aan een vermindering van de roosterspanning in het met fosforatomen gedoteerde emittergebied. Dit effect kan echter eenvoudiger worden verklaard met de door Aiken en Schwettmann onder soortgelijke omstandigheden aangetoonde diffusie van fosfor in het gebied van de emitter-basisjunctie, waardoor een deel van de basisdotering wordt gecompenseerd.

P.J. Kannam, IEEE Transact. El. Dev., Vol. ED-20, p. 845, 1973.
J.G. Aiken, F.N. Schwettmann, Semicond. Silicon 1973, Electrochem. Soc., p. 717.

III

De temperatuurafhankelijkheid van het electrisch gedrag van MOS-stuurcircuits met een "resistive gate" van polykristallijn silicium als centrale stuurelectrode kan worden geregeld door de keuze van de soortelijke weerstand van het polykristallijn silicium.

M.V. Whelan, L.A. Daverveld, J.G. de Groot, Philips Res. Repts., 30, p. 436, 1975.
J.Y.W. Seto, J. Appl. Phys., 46, p. 5247, 1975.

IV

Bij veldberekeningen van discontinuïteiten in golfpijpen wordt veelal gebruik ge-

maakt van de momenten-methode. Hierbij wordt het veld ter weerszijden van de discontinuïteit benaderd door een eindig aantal discrete modi die terplaatse van de discontinuïteit aan elkaar worden gepast. Hoewel deze methode voor gesloten golfpijpen zeer bruikbaar kan zijn, is zij in principe niet geschikt voor berekeningen aan discontinuïteiten in open diëlectrische golfgeleiders.

P.J. Clarricoats, K.R. Slinn, Proc. Inst. Electr. Eng., Vol. 114, p. 878, 1967.
S.F. Mahmoud, J.C. Beal, IEEE Transact. Microwave Theory and Technique, Vol. MIT-23, p. 193, 1975.

V

Omdat Grimbergen, bij de verklaring van zijn IV-metingen aan silicium dioden, veronderstelt dat de diffusieconstante van de electronen in p-type silicium een constante waarde van 10 $cm^2/sec$ heeft, onderschat hij de grootte van het pn-product bij hoge verontreinigingsconcentraties in het p-type gebied.

C.A. Grimbergen, Proefschrift aan de Rijksuniversiteit te Groningen, hoofdstuk 5, 1977.

VI

De verschuiving van de spectrale verdeling van de recombinatiestraling die optreedt bij hoge concentraties gat-electronparen in silicium en germanium, wordt doorgaans toegeschreven aan een verlaging van de bandafstand. In hoeverre dit verschijnsel invloed heeft op de werking van "power devices" dient nader onderzocht te worden.

V.S. Vavilov, E.L. Nolle, Sov. Phys.-Semicond., 2, p. 616, 1968.
V.M. Asnin, A.A. Rogachev, Sov. Phys.-Solid State, 5, p. 1257, 1963.

VII

Bij veel beschouwingen over "pn junction-devices" kan het begrip depletielaag beter worden vervangen door ruimteladingslaag.

VIII

De methode die Lewandowski voorstelt om de toekomstige dichtheidsverdeling van een bepaald consumenten-artikel te voorspellen op grond van de huidige verdeling ervan voor verschillende typen van huishoudens, is aanvechtbaar.

R. Lewandowski, Prognose- und Informationssysteme, Bd. 1, p. 355, Walter de Gruyter, 1974.

IX

Roosterfouten in ionenkristallen zijn er tendele voor verantwoordelijk, dat de experimenteel bepaalde capaciteit aan het grensvlak kristal/electroliet-oplossing veel lager is dan volgt uit berekeningen gebaseerd op de electrische dubbellaagtheorie van Gouy en Chapman.

H.R. Kruyt, Colloid Science, Vol. I, Hfdst. 4, Elsevier Publ. Comp., 1952.
E.P. Honig, Trans. Faraday Soc., 64, p. 2248, 1969.

X

In tekstboeken over geïntegreerde schakelingen wordt vaak de vertragingstijd van TTL poorten onjuist berekend.

L. Strauss, Wave Generation and Shaping, p. 177, McGraw-Hill, 1970.
A. van der Ziel, Introductory Electronics, p. 245, Prentice Hall, 1974.

XI

De ICRU beveelt een bol met een diameter van 30 cm aan als model voor het menselijk lichaam en definieert de index van de geabsorbeerde stralingsdosis als volgt: "The absorbed dose index at a point is the maximum absorbed dose within a 30 cm diameter sphere centered at this point and consisting of material equivalent to soft tissue with a density of 1 g/cm$^3$."
De hier gedefinieerde grootheid maakt het mogelijk dat de dosisbelasting in een punt van het lichaam wordt bepaald door de dosis in denkbeeldige anatomische structuren buiten het lichaam.

Intern. Comm. on Radiation Units, Report No. 25, 1976.