Nat.Lab. Unclassified Report 2001/804

Date of issue: 1990

Physically Based Compact Modelling of Lateral PNP Transistors

F.G. O'Hara B.E.

Unclassified Report
©Koninklijke Philips Electronics N.V. 2001

NAT. LAB. REPORT 6503

F.G. O'Hara B.E.

Physically Based Compact Modelling of Lateral PNP Transistors

Abstract

In the design of bipolar analog integrated circuits, greater flexibility is often achieved when both npn and pnp transistors are incorporated in the circuit design. Most present day bipolar production processes use the conventional *lateral* pnp as the standard pnp transistor structure. Unfortunately, the range of circuit designs in which these lateral pnps are used has been restricted due to the lack of an accurate physically based compact model for computer-aided circuit design. Existing lateral pnp compact models tend to be adaptations of their npn counterparts which neglect the complex two-dimensional nature of lateral pnp physics thereby leading to semi-empirical formulations.

The goal of this research is to provide a new highly accurate compact model for lateral pnp integrated circuit transistors which is based directly on device physics. This new model, named MODELLA[1], uses a new physical modelling approach whereby the main currents and charges are independently related to bias dependent minority carrier concentrations. Current crowding effects, high injection effects and a bias dependent output impedance are also modelled by MODELLA.

This thesis describes how the equivalent circuit topology, analytic equations and parameters of MODELLA are related to the physics and structure of the lateral pnp. Crucial aspects of compact model development such as the parameter determination strategy, predictive capability testing, temperature scaling and suitability to circuit simulation are treated in detail. Results and comparisons with existing lateral pnp compact models show the superiority of this new physically based compact model formulation. MODELLA should therefore prove to be a very attractive option to circuit designers when accurate circuit simulations involving lateral pnp transistors are required.

Note: In order to provide a complete treatment of MODELLA, this thesis also includes a description of the research done by Jan van den Biesen when he was developing the original protopype of MODELLA (Nat. Lab. Report Nr. 6131).

Keywords: lateral pnp, bipolar transistor, compact model, MODELLA, device physics, numerical device simulation, parameter extraction, temperature scaling, circuit simulation, modelling.

© N.V. Philips' Gloeilampenfabrieken

Contents

1 Introduction **1**

1.1 Background 1

1.2 The lateral pnp transistor 2

1.3 Compact modelling 4

1.4 Project aims 6

1.5 Thesis organisation 8

2 Lateral PNP physics **10**

2.1 Introduction 10

2.2 A qualitative description 11

2.3 A quantitative description 17

2.3.1 The collector current and the Early effect 17

2.3.2 The substrate current 22

2.3.3 The base current 25

2.3.4 Charges, capacitances and transit times 30

2.4 Results 36

2.5 Discussion 38

2.6 Summary and conclusions 42

3 Compact model formulation **45**

3.1 Introduction 45

3.2 Derivation of the basic equations 47

3.2.1 The collector current and the Early effect 47

3.2.2 The substrate current 51

3.2.3 The base current 52

3.2.4 Charges, capacitances and transit times 53

3.3 Modella 54

3.4 Behavioural description and parameter dependencies 59

3.5 Conclusions 62

4 Parameter determination **64**

4.1 Introduction 64

4.2 Parameter determination with Bipar 65

4.3 Parameter determination strategy 68

4.3.1 D.C. model parameters 68

4.3.2 Charge model parameters 78

4.3.3 Example 81

© N.V. Philips' Gloeilampenfabrieken

4.4 Discussion 86
4.5 Conclusions 89

5 Temperature Dependence Modelling **90**
5.1 Introduction 90
5.2 Temperature scaling rules 91
5.2.1 Background 91
5.2.2 Derivation of scaling rules 93
5.3 Verification 97
5.3.1 Introduction 97
5.3.2 Procedure 97
5.3.3 Results 98
5.4 Conclusions 105

6 Model comparisons **106**
6.1 Introduction 106
6.2 The extended Gummel-Poon model 107
6.3 Modella improvements 110
6.4 Results 111
6.5 Conclusions 119

7 Circuit simulation environment **120**
7.1 Introduction 120
7.2 Implementation procedure 121
7.3 Implementation verification 122
7.4 Results 125
7.5 Conclusions 126

8 Conclusions **129**

Bibliography **133**

Appendices

A Model equations and parameters **137**

B Parameter optimisation theory **151**
B.1 The optimisation process 151
B.2 The model evaluation routine 152
B.2.1 The solution process 153

C Proceedings from the BCTM '90 **158**

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 1

Introduction

1.1 Background

In the design of bipolar integrated circuits, the use of computer-aided circuit simulation is an invaluable tool. Computer predictions of circuit behaviour mean that circuit designs can be tested and optimised before the processing of silicon has even begun. In this way, development times and costs can be minimised. The accuracy and reliability of circuit simulators like E-SPICE[2] and PANACEA[3] are, however, directly proportional to the quality of the transistor compact models used in these simulators. These compact models predict transistor behaviour for any given operating condition and so when they are combined with a circuit solver, a prediction of circuit behaviour can be obtained. As the accuracy requirements of circuit simulators and therefore of compact models increase, research is required to develop more accurate *physically based* compact models.

To date, this research has concentrated mainly on the *npn* transistor. However, designers also need *pnp* transistors because of their complementary properties. Using pnp transistors means that a greater design flexibility is often achieved when they are incorporated with npn transistors in the same circuit design. Most present day bipolar processes use the conventional *lateral* pnp[1] as the standard pnp transistor structure. The quality of lateral pnp compact models has been sadly lacking, however, because present models tend to be adaptations of npn formulations which neglect the complex two dimensional nature of lateral pnp physics. Because of this inadequacy in the compact models (and thus in the circuit simulations), the range of circuit

[1] So called because the main current, Ic, flows in a mainly lateral direction parallel to the surface.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

designs in which lateral pnps are used is severely restricted. This results in a reduction in the flexibility of circuit design.

The aim of this research is to redress this imbalance in the quality of lateral pnp compact modelling by providing a new highly accurate compact model, based directly on device physics, for lateral pnp integrated circuit transistors.

1.2 The lateral pnp transistor

The conventional lateral pnp structure is one of several pnp structures available in bipolar processes today. Because the lateral pnp avails of the p-type base of the npn transistor to form its emitter and collector regions (see figure 1.1), no special masks or extra processing steps are required in its manufacture. It is therefore a cost effective pnp implementation which is relatively easy to manufacture and does not increase process development time.

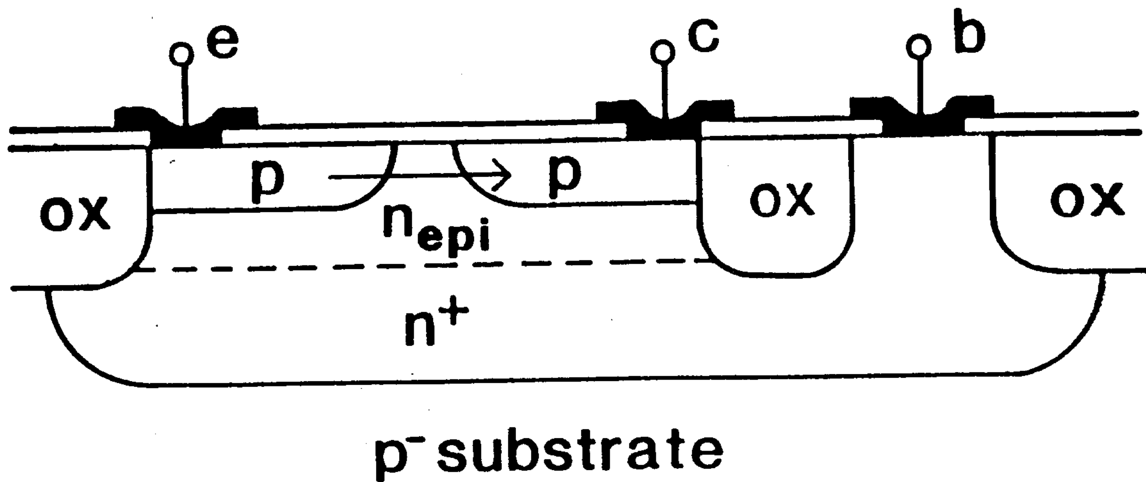
The main disadvantage of the lateral pnp is that, relative to the npn transistor, its performance in terms of gain and speed is quite poor. Optimised lateral pnp structures for analog applications rarely exhibit cut-off frequencies[2] above 100MHz or current gains above 100 [4]. Recently, high performance *vertical* pnp structures with cut-off frequencies in the GHz range have been reported in the literature. This type of high performance structure is obtained, though, at the cost of extra processing steps and usually requires two to three extra masks. These structures may well play a role in future bipolar production processes when applications demand high performance pnp transistors. However, unless this type of performance is a prerequisite for the circuit application, the lateral pnp structure will remain an effective and readily manufacturable pnp implementation suitable for general applications.

Lateral pnps have a variety of uses in present day bipolar integrated circuits. Examples of analog applications include buffers, current mirrors, level shifting circuitry, current sources and signal amplifiers . Furthermore, the lateral pnp constitutes an an integral part of I^2L[5] technology.

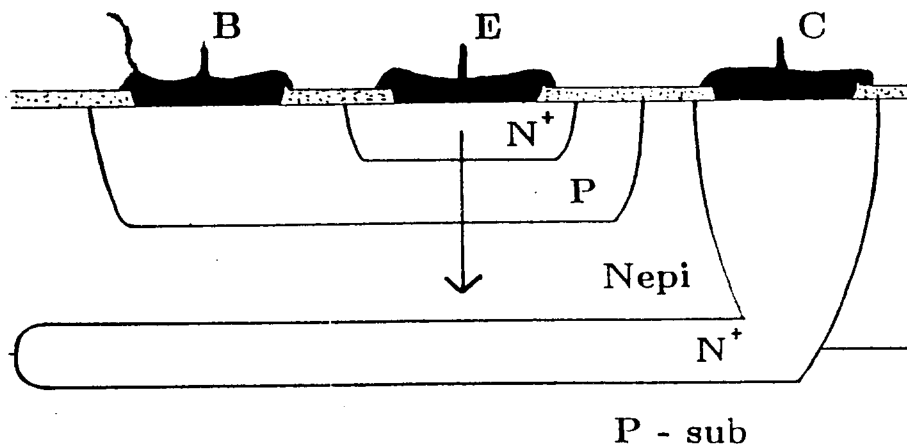
[2]The cut-off frequency, f_t, is a figure of merit which provides an indication of the maximum operating frequency of a transistor.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only



(a)



(b)

Figure 1.1: *Cross-sections of (a) the oxide-isolated lateral pnp transistor and (b) the standard vertical npn transistor.*

A simplified cross-section of an *oxide-isolated* lateral pnp structure, along with a standard vertical npn for comparison purposes, are shown in figure 1.1. The direction of main current flow in these devices is indicated by the arrows, showing why these devices are termed lateral and vertical. In the lateral pnp structure, silicon dioxide (SiO_2) is used to isolate the device

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

from its surroundings. The emitter and collector are formed by two p regions which are also used for the base of the npn. The active base of the lateral pnp is composed of a homogeneous epitaxial n-type doping with a highly doped buried leading to the base contact. A parasitic vertical pnp transistor (i.e. regions $pn^-n^+p^-$) also constitutes an inherent part of this structure.

It can be seen, therefore, that the lateral pnp operates very much as a two dimensional device, unlike the mainly one dimensional operation of the vertical npn. It follows that the behaviour and physics of such a device will be different in many respects to the npn. The device physics of a lateral pnp must thus be studied in its own right if we are to be able to accurately predict and model its behaviour. An accurate compact model *must* therefore reflect this two-dimensional nature if effects resulting from this specific structure are to be modelled correctly.

1.3 Compact modelling

In principle there are two ways of modelling the behaviour of semiconductor devices.

- **Numerical device simulation:** This involves computer simulation by solving numerically the Poisson and carrier continuity equations with appropriate boundary conditions and physical models. The aim of this type of modelling is to give predictions of device behaviour and also to give an insight into the internal physical operation of devices.
- **Compact modelling:** This involves finding an equivalent circuit and a compact set of analytic equations describing the electrical device characteristics. The aim of this type of modelling is to predict the terminal behaviour of devices.

We are concerned here with the latter of these two approaches. Although it is sometimes less accurate than the first approach, it has a great advantage in terms of its speed of evaluation and is therefore used in circuit simulators like E-SPICE and PANACEA. The first approach could in principle be used in circuit simulators but the large C.P.U. time would be prohibitive.

A compact model, therefore, consists of an equivalent circuit and a set of analytic equations preferably in explicit form. The numerical constants in these equations are called parameters and their values depend on the geometry of the transistor and the process in which the transistor is made. An example of a model parameter is the saturation value, $\mathbf{I_s}$, of the collector

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

current. Values for these model parameters can be obtained either from process and layout data in a process block[3] or directly from curve fitting the device characteristics obtained from measurements or numerical device simulation results.

Compact models can be further subdivided into *physically based* and *empirically based* formulations. Physical compact models are those in which the model equations and equivalent circuit are derived directly from device physics. Empirical models usually consist of rather simplified equivalent circuits and possess analytical expressions that are of a curve fitting nature and not primarily based on device physics. In practice, compact models usually fall somewhere between these two categories. Present day state of the art compact models are almost exclusively physically based formulations. This is because of the many advantages associated with these formulations:

- the model parameters have a physical significance and this can be used to check the plausibility of values obtained by parameter extraction procedures. In addition, parameters which have a physical significance, rather than an empirical fitting purpose, provide for a better intuitive understanding of compact modelling.
- a very useful advantage is the forecasting ability of these models. The correlation between model parameters and process quantities like doping concentrations is based on device physics. Therefore, if the doping level for example is altered, it is possible to forecast device behaviour by utilising this correlation.
- geometrical scaling rules can be applied with confidence. This contrasts with empirical models in which the correlation between the model parameters and geometry is lacking.
- the correlation between the model parameters is governed by device physics which can be used for realistic statistical modelling[4].

[3] A process block is a computer program which enables the circuit designer to calculate the compact model parameters for a given device geometry and a given process. It contains the relations between model parameters on one hand and process and layout information on the other. These relations are derived for a given process using a combination of theoretical and experimentally observed dependencies. A set of compact model parameters can therefore be generated for any given combination of doping levels, geometries, etc. This procedure to generate compact model parameters is highly suited to physical compact models due to the physical significance of the parameters.

[4] It should be noted, however, that parameter extraction procedures on measured device characteristics can induce an unphysical interdependence of some parameters due to the similar influence some parameters have on the characteristics. An example of this is the effect that resistance parameter values can have on the determination of the high injection parameter.

© N.V. Philips' Gloeilampenfabrieken

An important consequence of a physically based formulation is that the formation of process blocks, for the generation of model parameter sets, now has a theoretical as well as an experimental basis. This means that instead of simply experimenting to try to find parameter dependencies on process and geometric quantities, a theoretical prediction of these dependencies can be made beforehand. The development of process and geometric scaling rules for these parameters should therefore have a more scientific basis compared to ad hoc empirical scaling.

Well known examples of compact models for npn transistors are the Ebers-Moll[6] and Gummel-Poon[7] models. More recently, another npn compact model called MEXTRAM[8,9] has been developed. This is a physically based formulation which provides greater modelling accuracy especially in the high current ranges of the f_t (cut-off frequency) and h_{fe} (D.C. current gain) fall offs. No equivalent lateral pnp compact model formulation is presently available. Instead, lateral pnp compact models are based on either the Ebers-Moll or Gummel-Poon approaches. The equivalent circuits and equations of these npn formulations are adapted in an attempt to model the lateral pnp electrical behaviour. These models neglect important physical effects in the lateral pnp and, consequently, flaws in the physical description have to be compensated for by choosing *non-physical* parameter values. Indeed, many of the model parameters are of a non-physical nature to begin with, having been derived in a (semi-)empirical manner.

1.4 Project aims

Having defined the problems associated with present day lateral pnp compact modelling approaches, this project aims to develop a compact model which will provide the circuit designer with the same level of accuracy for lateral pnp transistors as Mextram does for vertical npn transistors. This involves adopting new modelling approaches for the lateral pnp based on a detailed knowledge of the device physics.

The best presently available lateral pnp models within Philips are the Gummel-Poon based model TPL-301[10] and various variations based on this e.g. TPLHEA[11] and TPLHLEAK[12]. These variations or hybrid models are attempts to reformulate certain aspects of TPL-301 in order to compensate for model inadequacy. TPLHEA for example, is identical to TPL-301 except that an equation has been reformulated to try to introduce some (unphysical) form of

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

collector-base voltage (Vcb) dependency of the Early voltage[5]. When this project is completed, it is hoped that Modella will prove to be a more accurate and more physical alternative to all of these models, thereby providing *one* model for *all* applications.

Since a theoretical prototype of Modella has already been proposed[14], this project is not starting from first principles. Rather, it is concerned with **1)** investigating this virtually untested proposal for a lateral pnp compact model and **2)** developing this prototype to the stage where is provides a realistic and attractive option to designers using lateral pnps in circuit simulations.

In order to explain the implications of these general aims, a more detailed description of the project aims follows.

- To consolidate and extend a knowledge of lateral pnp device physics.
- To provide an extensive evaluation of the prototype and to make any appropriate alterations. More specifically this will involve:
  - Testing the D.C. and A.C. behaviour of Modella on a wide variety of lateral pnp structures from different processes. This should involve determining the accuracy of the model's predictive capability, defining the range of the model's validity and liaising with designers to ensure that a correct balance is maintained between the accuracy and complexity of the model.
  - Assuring continuity of functions and derivatives since discontinuities can lead to problems with convergence in circuit simulations and/or incorrect model predictions.
- The development of a compact model does not only deal with deriving the model equations and equivalent circuit but it must also develop methods to extract the various parameter values. Indeed, the model evaluation described above cannot be carried out effectively without the availability of an efficient and reliable parameter determination strategy. Developing this strategy will involve finding the most effective optimisation methodology to determine parameters from measured device characteristics by, for example, investigating the sensitivity of these characteristics to the different parameters. Other measurement

[5]The Early voltage is defined as $V_{eaf} = Ic|\frac{dVcb}{dIc}| + Vcb$ where Ic and Ib are the collector and base currents respectively. This figure of merit refers to the Early effect[13] which is a measure of transistor output impedance or, expressed another way, the sensitivity of Ic to Vcb variation.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

techniques to determine some of the parameters in a more direct manner, possibly without the use of optimisation methods, will be investigated. The physical correctness of parameters, measurement induced parameter correlation and other aspects of parameter determination must be examined.

- In order to increase the range of validity of Modella, the temperature scaling of model parameters will be investigated. Without temperature scaling, a parameter set is only valid for the temperature at which it was determined. When the temperature is altered the device behaviour changes due to the temperature dependence of basic physical quantities like the intrinsic carrier concentration[15]. Temperature scaling attempts to scale parameters values (in a physical way) by relating them to the basic physical quantities on which they depend and for which the temperature dependence is known.

- Suitability to circuit simulation is another crucial aspect of model development. The equivalent circuit and equation formulations must lend themselves to solution by a circuit solver. Any convergence difficulties must be identified and solved in the circuit simulation implementation. The speed of convergence must be evaluated to provide the designer with an estimate of the model's C.P.U. intensity.

- Comparisons with existing compact models in terms of accuracy, speed, range of validity, convergence properties, physical basis, etc. form an integral part of model development. Advantages and disadvantages of the model must be analysed and communicated to designers so that its suitability for use with particular problems can be established.

1.5 Thesis organisation

The following is a list of chapters and their main subjects.

- Chapter 2 provides a summary both in qualitative and quantitative terms of lateral pnp device physics. Numerical device simulations are used where appropriate to complement this discussion.

- In chapter 3 the equivalent circuit and analytic equations of the compact model prototype are derived from device physics. A short analysis of model behaviour helps to provide an insight into its complex functionality.

© N.V. Philips' Gloeilampenfabrieken

- Chapter 4 treats the difficult yet important area of parameter determination with a view to developing an efficient parameter determination strategy.
- Temperature scaling rules for the model parameters are derived in chapter 5. An example of the determination of temperature parameters required for this scaling is given and some results are presented.
- An extensive comparison with the best presently lateral pnp compact model is presented in chapter 6.
- The suitability of Modella to circuit simulation and the methods used to verify correct implementation are discussed in chapter 7.
- Finally, chapter 8 presents conclusions and suggests future work. It discusses the results and achievements of this research and outlines its contribution to the field of bipolar compact modelling.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 2

Lateral PNP physics

2.1 Introduction

Since the first publication on the lateral pnp in 1964 by Lin et.al.[16], many theoretical and experimental investigations [17] - [25] have been carried out in order to improve our understanding of this device. Numerical techniques and analytical formulations were developed in order to understand, predict and model device behaviour. When reviewing this literature, one encounters many varying and often conflicting theories, especially with respect to the relative significance of the various geometrical effects and physical mechanisms that determine device behaviour.

This problem is compounded by the fact that improved processing techniques in present day bipolar processes have rendered some previous analyses obsolete. These improvements have affected both the underlying physical mechanisms and geometrical considerations. For example, increased bulk carrier lifetimes have led to a reduction in the contribution of bulk recombination to the base current. In addition, vertical downscaling of npn transistors has led to reduced junction depths in the lateral pnp thereby increasing the significance of two dimensional effects in this device.

The purpose of this chapter is to summarise lateral pnp physics relevant to present day bipolar processes. This summary has its origins in the literature and to a large extent in the work of Van den Biesen[14]. Numerical device simulations using the two-dimensional package, CURRY[26], and the one-dimensional package, TRAP[27], are used to provide an insight into the device physics of typical lateral pnp structures. Results from these simulations are also used

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

to check predictions from analytical formulations.

After a brief qualitative introduction to the internal operation of the lateral pnp in the next section, section **2.3** describes how analytical equations can be formulated to describe device behaviour. The effects of ohmic voltage drops (i.e. voltages dropped across internal resistances) are not included here since they cannot be accounted for analytically. Section 2.4 compares results from these analytical formulations to numerical device simulation results. A discussion on the consequences that certain structural variations, three-dimensional effects, and series resistances have for device behaviour follows in section **2.5**. A summary of lateral pnp device physics and some conclusions are provided in section **2.6**.

2.2 A qualitative description

A cross section of a typical lateral pnp transistor was shown schematically in figure 1.1. In this example, oxide (SiO_2) is used to isolate the lateral pnp from other devices. The emitter and collector regions are formed by using the p-type base implantation of the npn transistor. These are implanted into the homogeneously doped n-type epitaxial layer which forms the active base region. An n^+ buried layer completely isolates the epitaxial region from the p^- substrate region and an n^+ implantation is generally used under the base contact.

To understand how this device operates internally, it is useful to consider the results of numerical device simulations on the simplified structure shown in figure 2.1. The simplification involves assuming abrupt, rectangularly shaped junctions (with homogeneously doped regions therefore assumed in the emitter, collector and buried layer regions). The doping profiles in the lateral and vertical directions are shown in figure 2.2 and figure 2.3 respectively. The third dimension is of unit length and is assumed to have no influence on device behaviour thereby facilitating two-dimensional representation. Simulation results are obtained using the program CURRY which solves the Poisson equation and carrier continuity equations in two dimensions. It includes standard models for physical effects such as band gap narrowing, surface and bulk recombination, doping and field dependent mobility, etc.

Consider firstly figure 2.4 which shows the hole current flow lines at *moderate* forward bias. In this figure only the *direction* of hole current flow is indicated; the magnitude of the current

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

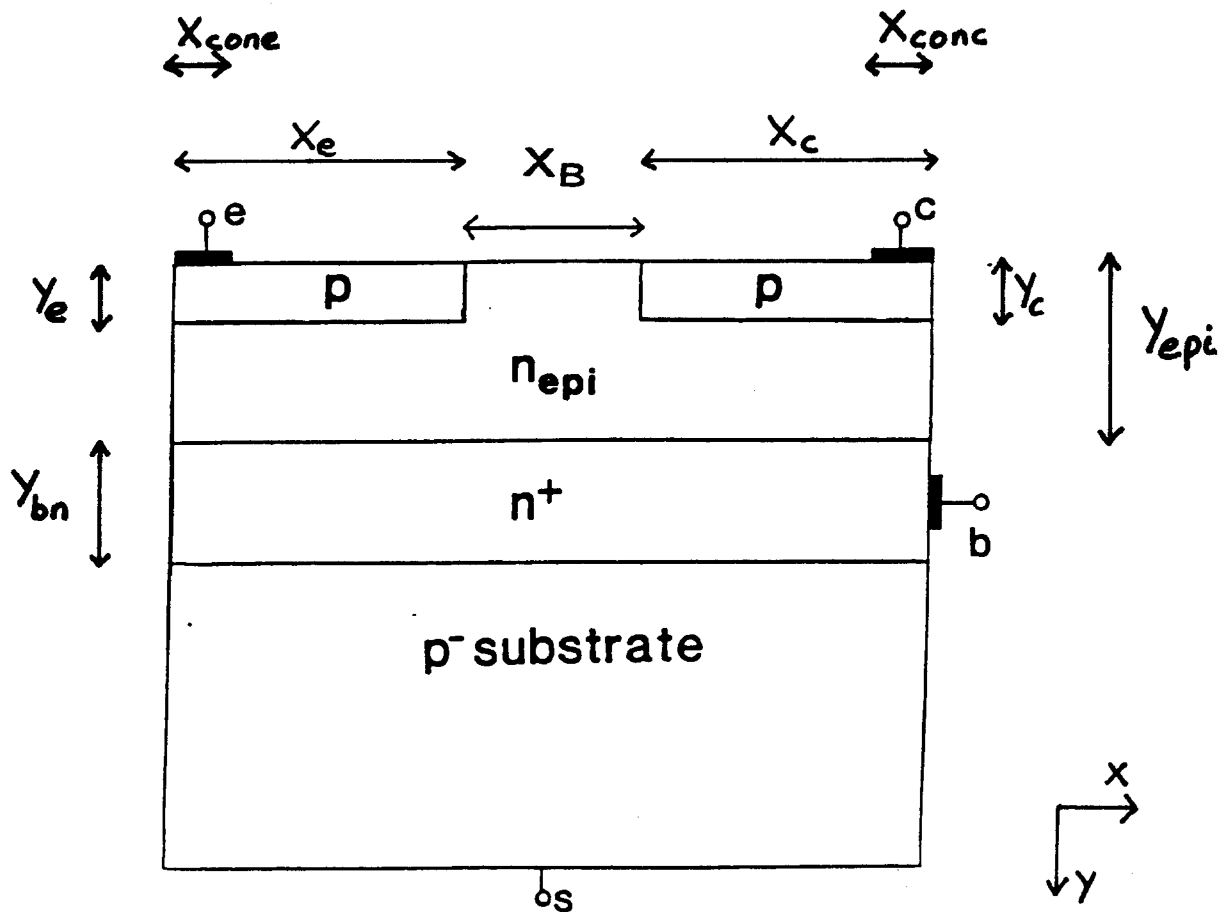


Figure 2.1: *Simplified cross section of a lateral pnp transistor used in numerical device simulations. The relevant geometrical variables are indicated.*

density being treated in section **2.3**. A portion of the holes injected by the emitter into the epitaxial region laterally crosses the base and the reverse biased collector-base junction to reach the collector. This gives rise to part of the collector current Ic. Another portion of the injected holes arrives at the collector, but along curved trajectories in the lower part of the epitaxial layer. The remainder of the injected holes either recombines in the buried layer or crosses the reverse biased substrate-base junction to reach the substrate thereby giving rise to the substrate current Is.

In figure 2.5 the electron current flow lines at moderate forward bias are indicated (the direction is that of conventional current flow). Most of the electrons are injected from the base into the emitter thereby giving rise to the base current Ib.

© N.V. Philips' Gloeilampenfabrieken

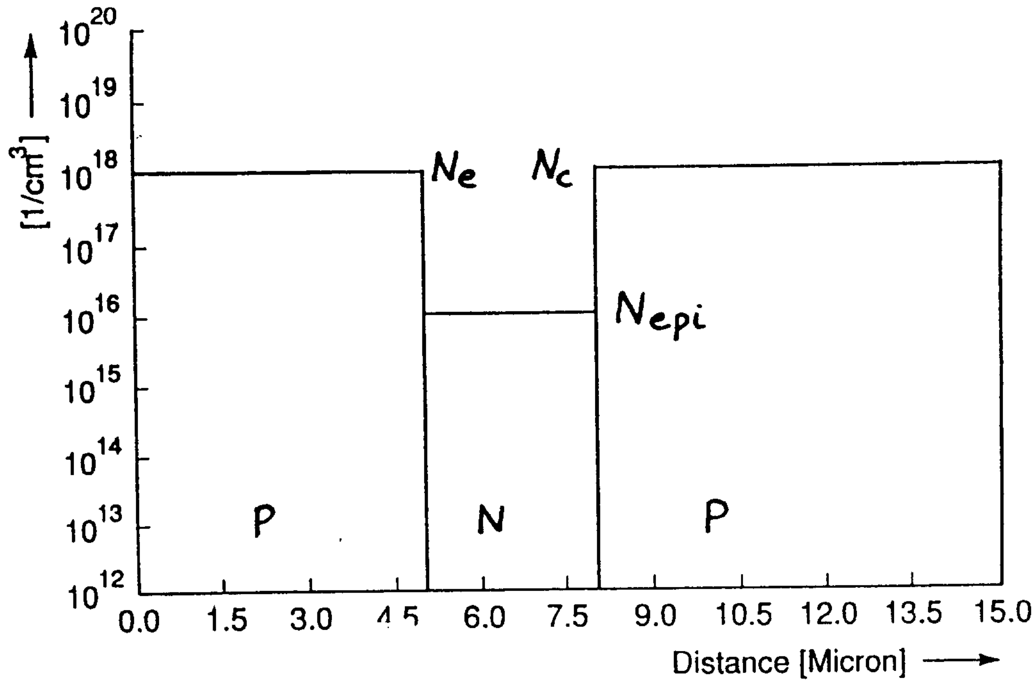


Figure 2.2: *Lateral cross section of the simplified lateral pnp transistor structure used in numerical device simulations.*

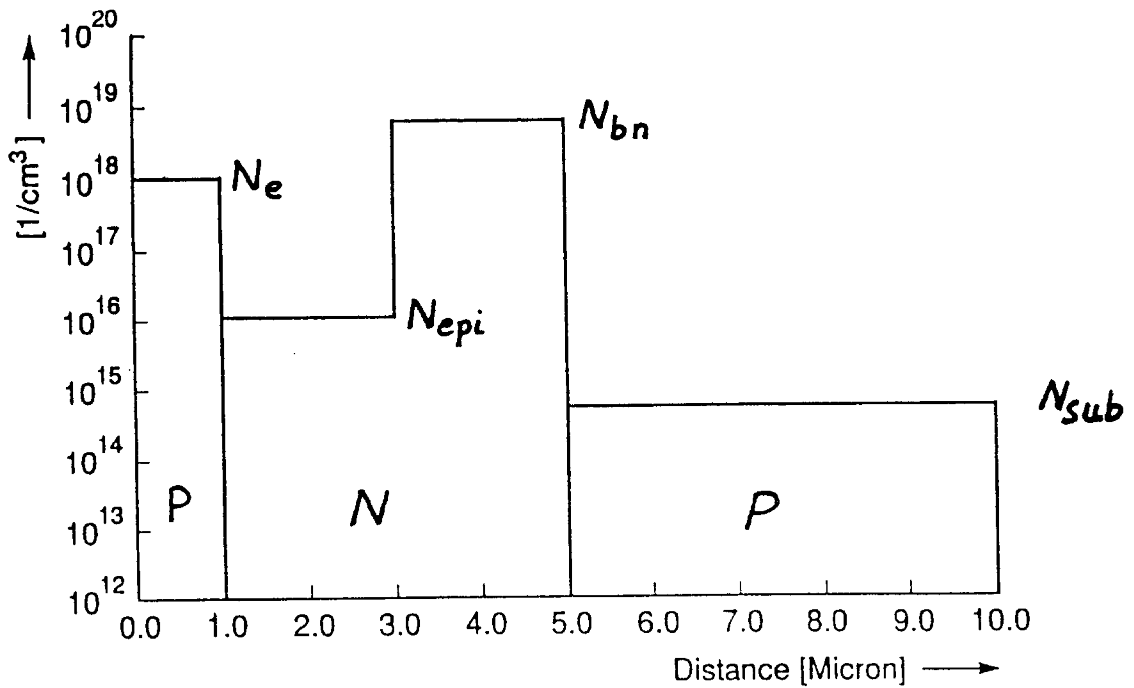


Figure 2.3: *Vertical cross section of the simplified lateral pnp transistor structure used in numerical device simulations.*

© N.V. Philips' Gloeilampenfabrieken

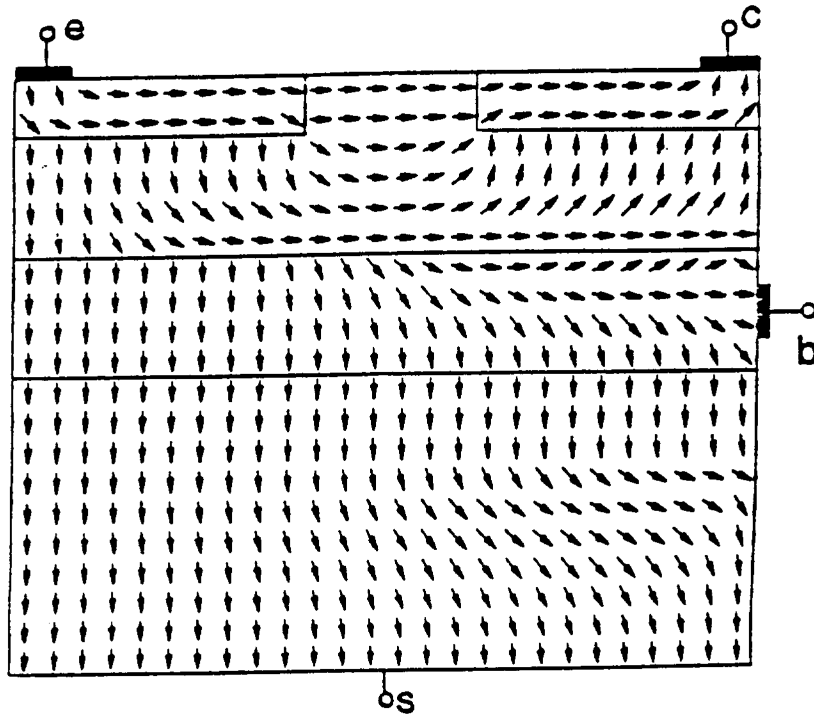


Figure 2.4: *Hole current flow lines in the simplified structure at moderate forward bias ($Veb = 0.5V, Vcb = -1.0V$ and $Vsb = -1.0V$).*

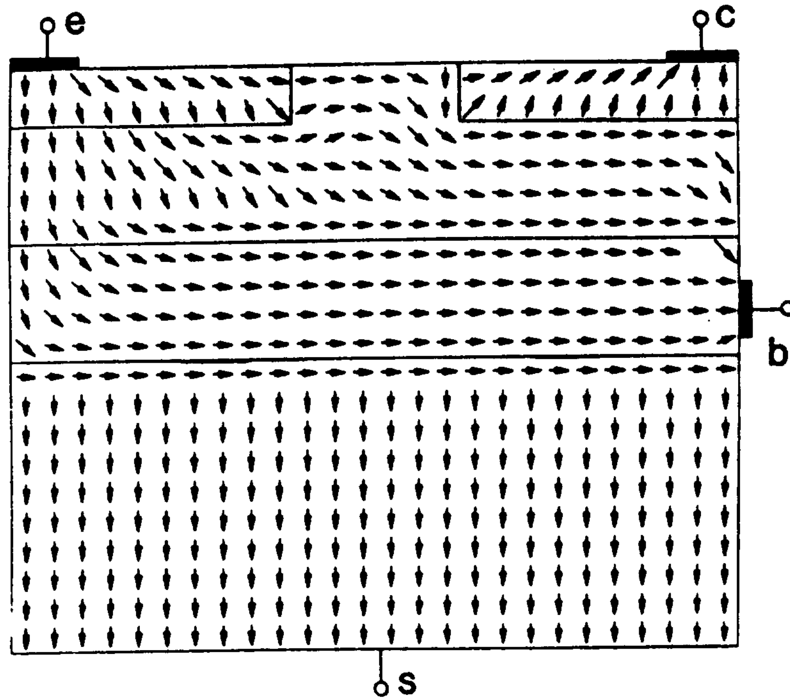


Figure 2.5: *Electron current flow lines in the simplified structure at moderate forward bias ($Veb = 0.5V, Vcb = -1.0V$ and $Vsb = -1.0V$).*

© N.V. Philips' Gloeilampenfabrieken

At moderate forward bias, the four terminal currents (the emitter current, $Ie = Ic + Ib + Is$) exhibit an ideal exponential dependence on the emitter-base voltage i.e. $I \propto \exp \frac{Veb}{Vt}$ where $Vt = \frac{kT}{q}$, the thermal voltage. This ideal behaviour at moderate forward bias leads to constant current gains in both the lateral pnp (e-b-c) transistor (i.e. $h_{fe} = \frac{Ic}{Ib} = constant$) and in the parasitic vertical pnp (e-b-s) transistor (i.e. $h_{fs} = \frac{Is}{Ib} = constant$).

At *high* forward bias two phenomena give rise to non-ideal behaviour: *high injection* and *current crowding*.

- High injection: This effect occurs when the injected minority carrier concentration exceeds the impurity concentration. Because the epitaxial layer concentration is relatively low, the injected hole concentration soon exceeds this value as the forward bias is increased. The collector current then loses its ideal exponential dependence on Veb and starts to behave according to $Ic \propto \exp \frac{Veb}{2Vt}$. As a consequence of this non-ideal behaviour, the current gain h_{fe} falls off at high current levels. Figure 2.6[1] shows that the hole current flow pattern in the epilayer is unchanged at high forward bias. This is an important result as it shows that the direction of hole current flow which constitutes Ic is unaffected by the injection level.
- Current crowding: At high current levels, voltages dropped across internal resistances can lead to reduced values of junction voltages. In the emitter region this ohmic voltage drop in the region between the emitter contact and the emitter sidewall leads to a debiasing or reduction of the sidewall emitter-base voltage. The emitter-base voltage directly underneath the emitter contact is however unaffected by this voltage drop across the lateral emitter resistance. As a result, the current tends to 'crowd' into this region. Figure 2.7 shows the hole current at high forward bias, this time using default mobility values so that ohmic voltage drops are included. When this is compared to the case without ohmic voltage drops (figure 2.6), it can be seen that the pattern has changed dramatically. Due to the ohmic voltage drop across the lateral part of the emitter region, the collector current now originates from hole injection directly underneath the emitter contact. This effect leads to a reduction in Ic and therefore in h_{fe} at high current levels.

[1]The effect of ohmic voltage drops in this simulation has been eliminated by increasing the values of *majority* carrier mobilities in the various regions. Since the value of resistance is inversely proportional to these mobilities, ohmic voltage drops are therefore excluded.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

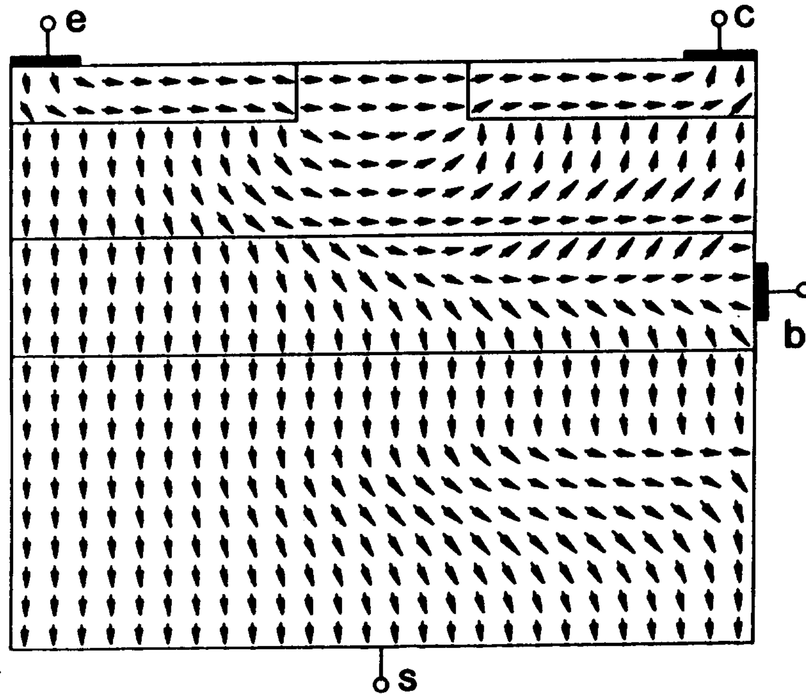


Figure 2.6: *Hole current flow lines in the simplified structure at high forward bias ($Veb = 0.9V, Vcb = -1.0V$ and $Vsb = -1.0V$). Ohmic voltage drops are excluded in this simulation.*

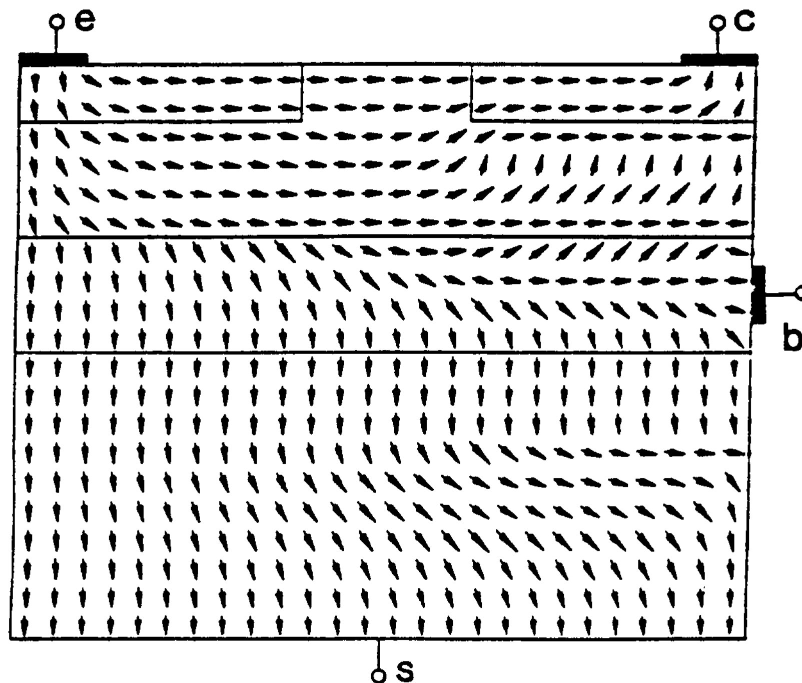


Figure 2.7: *Hole current flow lines in the simplified structure at high forward bias ($Veb = 0.9V, Vcb = -1.0V$ and $Vsb = -1.0V$). Ohmic voltage drops are included in this simulation.*

This brief introduction to some of the main physical effects at moderate to high forward bias will be extended and described analytically in the next section.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

2.3 A quantitative description

In this section ohmic voltage drops are not accounted for in any analytical derivations. Effects arising from these ohmic voltage drops are instead discussed separately in a less quantitative way by observing predictions from numerical device simulations. With regard to biasing, this section assumes forward active bias with the reverse active bias case following from symmetry.

2.3.1 The collector current and the Early effect

The collector current

In the last section it was shown that two collector current components can be distinguished: an almost purely lateral flow and a flow along curved trajectories. This section outlines a method to describe these two components analytically:

- *The lateral component*: This component is quite straightforward to describe as the problem is basically one-dimensional. The contribution of this lateral hole flow to the collector current is given approximately by $Jpx(0) \times Ye$. Here $Jpx(0)$ is a one dimensional hole current density which travels in the x direction directly under the oxide (at $y = 0$) between emitter and collector (see figure 2.1). Ye is the emitter-base junction depth. Simulations with the TRAP program show that $Jpx(0)$ is equal to the hole current density of a one-dimensional pnp transistor[2] with the doping profile of figure 2.2. An exact solution to this one dimensional problem can be derived [28] which also holds under high injection conditions[3].

$$Jpx(0) = q\, D_p\, N_{epi} \times (2p_0 - ln(1 + p_0)) \,/\, Xb \qquad (2.1)$$

where

D_p	=	the hole diffusion coefficient
N_{epi}	=	the epitaxial doping concentration
Xb	=	the base width

[2] The surface recombination is in general negligible which means that the hole current density has no component normal to the surface in this part of the base region.

[3] The main assumptions made in this derivation are 1) quasi-neutrality and 2) negligible electron current density in the base.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

p_0 is the reduced (i.e. normalised to the background doping concentration) injected minority carrier concentration in the epitaxial region and is given by

$$p_0 = \left(\sqrt{1 + 4 (\frac{n_i}{N_{epi}}^2) \exp \frac{Veb}{Vt}} - 1 \right) / 2 \tag{2.2}$$

where $n_i =$ the intrinsic carrier concentration. The limits of low and high injection are given by $p_0 \ll 1$ and $p_0 \gg 1$ respectively. Equation (2.1) therefore predicts that

– at low injection

$$\begin{aligned} Jpx(0) &= q\, D_p\, N_{epi}\, p_0 \,/\, Xb \\ and \;\; p_0 &\propto \exp \frac{Veb}{Vt} \end{aligned} \tag{2.3}$$

– at high injection

$$\begin{aligned} Jpx(0) &= 2\, q\, D_p\, N_{epi}\, p_0 \,/\, Xb \\ and \;\; p_0 &\propto \exp \frac{Veb}{2\,Vt} \end{aligned} \tag{2.4}$$

The factor of two increase at high injection can be described as the electric field enhancement of the minority carrier diffusion coefficient, D_p, since it translates into a reduction, by a factor of two, in the time taken for the holes to travel from emitter to collector (the Webster effect[29]).

- *The contribution from curved trajectories*: As described above, these curved hole current trajectories originate from the bottom of the emitter and terminate at the bottom of the collector. In order to account for the contribution from these trajectories, the collector current can be expressed as

$$Ic = Jpx(0) \times Ye'' \tag{2.5}$$

where the effective emitter-base junction depth, Ye'', is somewhat larger than Ye. The dependence of Ye'' on geometry (thickness of the epitaxial layer, emitter and collector junction depths and the base width) has been calculated numerically by Chou[17]. Berger et.al.[20] suggested using $Ye'' = Yepi$ but this yields quite inaccurate results compared to the numerical solution. Stoutjesdijk[30] found a simple analytical approximation for

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Ye'' by assuming that the curved trajectories consist of quarter circles and straight line segments. In this approximation, the contribution of each trajectory was taken as being inversely proportional to its length giving:

$$\begin{aligned} Ye'' &= Ye + Xb \int_0^{Yepi-Ye} \frac{1}{Xb + \pi y} dy \\ &= Ye + \frac{Xb}{\pi} ln \left(1 + \pi \frac{Yepi - Ye}{Xb}\right) \end{aligned} \quad (2.6)$$

The accuracy of this analytical approximation in describing the contribution from the two dimensional current flow was investigated by comparing it to the numerical solution obtained by Chou for various geometries[4]. Agreement between analytic and numerical results was good over the entire range of geometries found in practical devices.

In these calculations the widths of the emitter-base and collector-base depletion widths (i.e. Web and Wcb) were neglected. These depletion regions can be assumed to extend entirely into the epilayer because of its low doping concentration (i.e. one-sided abrupt junctions are assumed). Accordingly, the electrical dimensions that really determine quantities like the collector current are (see figure 2.1):

$$\begin{aligned} Xe' &= Xe + Web \\ Ye' &= Ye + Web \\ Xc' &= Xc + Wcb \\ Yc' &= Yc + Wcb \\ Xb' &= Xb - Web - Wcb \end{aligned} \quad (2.7)$$

The analytical formulation of Stoutjesdijk can be extended to incorporate these electrical dimensions and the consequent asymmetry between emitter and collector yielding:

$$Ye'' = Ye' + \frac{Xb'}{\pi} ln \left[\left(1 + \pi \frac{Yc' - Ye'}{2\,Xb'}\right) \left(1 + \pi \frac{2\,Yepi - Ye' - Yc'}{2\,Xb'}\right) \right] \quad (2.8)$$

[4]Because Chou's numerical investigations used curved junctions, a realistic comparison was only possible by expanding the analytical formulation to include junction curvature. This was done according to the formulation of Seo and Kim[21].

© N.V. Philips' Gloeilampenfabrieken

In order to illustrate the physical plausibility of this approach (i.e. $Ic = Jpx(0) \times Ye''$) in describing the collector current, the reader is referred to figure 2.8. Here the x component of the hole current distribution in the epilayer between emitter and collector (normalised to the value at $y = 0$) is shown. The value of Ye'' obtained using equation **(2.8)** is seen to be quite realistic. Figure 2.8 also shows the hole current distribution at high injection. In agreement with figures 2.4 and 2.6, this distribution retains its shape relative to the low injection case. It can therefore be deduced that equation **(2.1)**, which was derived to describe the hole current density of a one-dimensional pnp transistor *at all injection levels*, can also be applied to the collector current of a lateral pnp transistor *at all injection levels.* Section 2.5 shows how well the collector current predicted by equation **(2.5)** agrees with two-dimensional numerical device simulations.

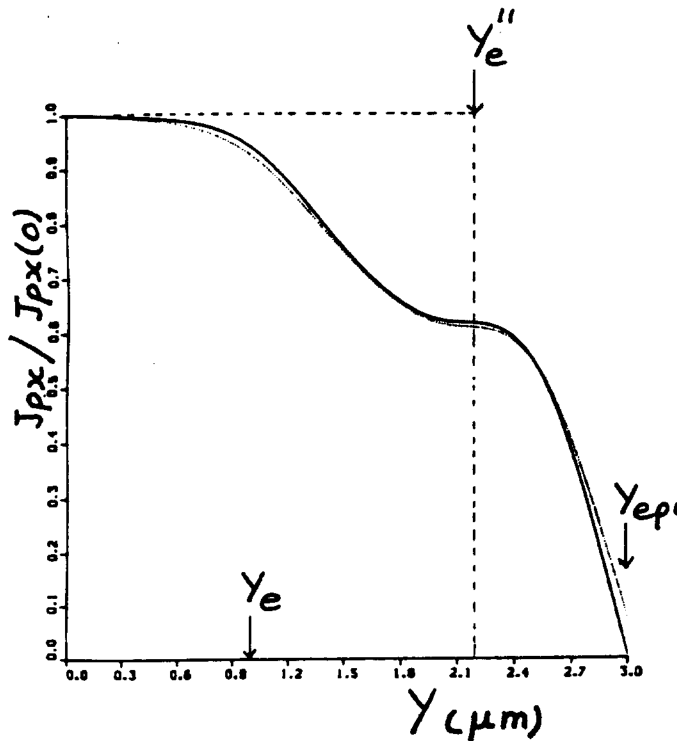


Figure 2.8: *The x component of the hole current distribution in the epilayer between emitter and collector at moderate (full line) and at high (dotted line) forward bias. Ohmic voltage drops excluded.*

The Early effect

The Early effect[13] relates to the variation of the collector current Ic with the collector-base voltage Vcb. An increase in magnitude of Vcb leads to an increase in the collector-base depletion width Wcb which in turn increases Ic. This is an important effect in the lateral pnp because the

© N.V. Philips' Gloeilampenfabrieken

low epilayer doping concentration results in a low value of the Early voltage V_{eaf}. This figure of merit is commonly defined as

$$V_{eaf} = Ic \left| \frac{dVcb}{dIc} \right| + Vcb \tag{2.9}$$

where the emitter-base voltage Veb or equivalently Ib is taken at a constant value. Therefore, the lower the Early voltage the greater the Early effect. In the above analytical formulation for Ic (equation **(2.5)**), the Early effect is modelled in one dimension by the dependence of $Jpx(0)$ on Xb' and in two dimensions by the dependence of Ye'' on Xb' and Yc'. Using equation **(2.9)** to determine the Early voltage of a lateral pnp we find:

$$V_{eaf} = \frac{Jpx(0)}{\frac{dJpx(0)}{dVcb} + \frac{Jpx(0)}{Ye''}\frac{dYe''}{dVcb}} + Vcb \tag{2.10}$$

In the literature, the Early voltage of a lateral pnp has been described in terms of a one-dimensional device (i.e. $V_{eaf} = Jpx(0)/\frac{dJpx(0)}{dVcb}$) [23]. Indeed, the contribution of the second term in the denominator of equation **(2.10)** turns out to be less that about 20% for geometries occurring in practical devices. It can therefore be deduced that the Early effect in a lateral pnp transistor (as in figure 2.1) can be roughly approximated by the Early effect in a one-dimensional transistor (as in figure 2.2).

It is often assumed that the Early voltage of a one-dimensional transistor (as in figure 2.2) is a constant. However, by using the analytical formulation for $Jpx(0)$ and noting that the depletion layer width of the one-sided abrupt collector-base junction is given by

$$Wcb = \sqrt{\frac{2\epsilon}{qN_{epi}}(Vd - Vcb)} \tag{2.11}$$

where ϵ = the permittivity
Vd = the diffusion voltage

it can easily be shown that

$$V_{eaf} \propto Wcb(Xb - Wcb) \tag{2.12}$$

The Early voltage is therefore quite a strong function of Vcb, increasing in value as the reverse bias is increased.

© N.V. Philips' Gloeilampenfabrieken

2.3.2 The substrate current

As was seen in figure 2.4, the substrate current is composed of holes which are injected downwards by the emitter. The holes travel through the epilayer and buried layer regions and are collected by the substrate. As will be shown in the analysis below, the magnitude of these current density flow lines is proportional to the value of the injected hole concentration in the epilayer along the epilayer-buried layer interface. Numerical simulations and the work of Chou[17] have shown that three regions can be distinguished in which good approximations for the hole concentrations can be made. The epilayer directly under the the collector region can be assumed to have zero hole concentration and therefore the contribution to the substrate current, Is, from this hole current flow will be neglected. This leaves the region directly under the emitter and the region between emitter and collector. A method to analytically describe these two components of Is is outlined below.

- *The component under the emitter region:* The contribution of this hole flow originating from the bottom of the emitter region is given approximately by $Jpy(0) \times Xe'$. Here $Jpy(0)$ is the hole current density which travels downwards along the left hand wall of the device shown in figure 2.1 (at $x = 0$). Xe' is the electrical emitter width defined in equation **(2.7)**. From simulations, $Jpy(0)$ is seen to be equal to the hole current density of a one-dimensional pnp transistor with the doping profile of figure 2.3. The blocking action of the highly doped buried layer is seen to be mainly due to its high Gummel number[5] and not so much due to recombination. As a consequence of the high Gummel number, $Jpy(0)$ has an ideal exponential dependence on Veb up to very high forward bias. For normal values of epitaxial thickness, the injected hole concentration, p_{epi}, in the epilayer can be assumed to have a value which does not vary in the y direction between the emitter-base junction and the epilayer-buried layer interface. This value is equal to the value under the emitter-base junction i.e. $p_{epi} = N_{epi} \times p_0$, where p_0 is the reduced injected hole concentration given by equation **(2.2)**. $Jpy(0)$ is therefore a diffusion current caused by

[5] The Gummel number of the buried layer in figure 2.3 is defined as

$$\begin{aligned} G_{bn} &= \int_{Y_{epi}}^{Y_{epi}+Y_{bn}} \frac{n}{D_{pbn}}\, dy \\ &= \frac{Ybn \times N_{bn}}{D_{Pbn}} \end{aligned}$$

where D_{pbn} is the hole diffusion coefficient in the buried layer.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

the fall off of the hole concentration in the buried layer:

$$\begin{aligned} Jpy(0) &= q\, D_{pbn} \frac{dp}{dy} \\ &= q\, D_{pbn} \frac{p_{bn0}}{Ybn} \end{aligned} \tag{2.13}$$

where D_{pbn} = the hole diffusion coefficient in the buried layer

p_{bn0} = the injected hole concentration in the buried layer at the epilayer-buried layer interface

The pn product, which is given by $pn = n_i \exp \frac{Veb}{Vt}$ is constant[6] across the epilayer-buried layer interface and this allows p_{bn0} to be written in terms of p_{epi}, the hole concentration in the epilayer. Neglecting high injection in the buried layer (because of its high doping concentration), we have:

$$\begin{aligned} p_{bn0}\, n_{bn0} &= p_{epi}\, n_{epi} \\ => \quad p_{bn0} &= \frac{(N_{epi} + p_{epi})\, p_{epi}}{N_{bn}} \end{aligned}$$

The contribution to Is from this component is therefore given by

$$Is = Jpy(0) \times Xe' \tag{2.14}$$

where

$$\begin{aligned} Jpy(0) &= \frac{q\, D_{pbn}\, (N_{epi} + N_{epi} p_0)\, N_{epi}\, p_0}{Ybn\, N_{bn}} \\ &= \frac{q\, D_{pbn}\, N_{epi}^2 (1 + p_0)\, p_0}{Ybn\, N_{bn}} \end{aligned} \tag{2.15}$$

In the case of high voltage devices with thick epitaxial layers, $Jpy(0)$ may be reformulated to account for the influence of the charge in the epilayer. The Gummel approach can be used in this case.

- *The component between the emitter and collector*: The contribution of the hole flow through the epilayer and buried layer located between emitter and collector is given approximately

[6]Bandgap narrowing has been omitted here for the sake of clarity. It can be included by simply using the appropriate *effective* intrinsic carrier concentrations in regions with different doping concentrations[15].

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

by $Jpy(x) \times Xb'$. Xb' is the electrical base width defined by equation (2.7) and $Jpy(x)$ is the as yet unknown hole current density which travels downwards between emitter and collector. Using assumptions similar to above, we can say:

$$Jpy(x) \;=\; q\, D_{pbn} \frac{p_{bn0}(x)}{Ybn} \tag{2.16}$$

Similarly

$$p_{bn0}(x) \;=\; \frac{(N_{epi} + p_{epi}(x))\, p_{epi}(x)}{N_{bn}}$$

According to Chou[17] and Lindmayer and Wrigley[31], the injected hole concentration in this epilayer region falls off linearly from $N_{epi}\, p_0$ at the emitter side to zero at the collector. Therefore assuming an average value for p_{epi} of $\frac{N_{epi}\, p_0}{2}$ we have:

$$\begin{aligned} Jpy(x) &= \frac{q\, D_{pbn}\, N_{epi}^2 (1 + \frac{p_0}{2})\, \frac{p_0}{2}}{Ybn\, N_{bn}} \\ &= Jpy(0) \frac{1}{2} \frac{(1 + \frac{p_0}{2})}{(1 + p_0)} \end{aligned} \tag{2.17}$$

This component can therefore be also expressed in terms of the one-dimensional hole current of the pnp transistor in figure 2.3. Because of symmetry in the hole concentration in the epilayer between emitter and collector, this contribution has a weight factor of $\frac{1}{2}$ which reduces to $\frac{1}{4}$ at high injection[7] .

The substrate current can now be expressed as

$$Is \;=\; Jpy(0) \times Xe'' \tag{2.18}$$

[7] It should be noted that it can be shown in a more exact analysis that this factor is actually $\frac{1}{3}$. This can be derived by taking the linear fall off in p_{epi} between emitter and collector and integrating from Xe' to $Xe' + Xb'$. Because the approximation of a constant value of p_{epi} is not made, we then obtain

$$Jpy(x) \;=\; Jpy(0) \frac{1}{2} \frac{(1 + \frac{2p_0}{3})}{(1 + p_0)}$$

© N.V. Philips' Gloeilampenfabrieken

where Xe'', the *effective* emitter width, is larger that Xe' and is given by:

$$Xe'' \;=\; Xe' + \frac{1}{2}\frac{(1 + \frac{p_0}{2})}{(1 + p_0)}\,Xb' \tag{2.19}$$

In figure 2.9 it can be seen that the width of the distribution of $Jpy(x)$ calculated with CURRY indeed changes at high injection, in fair agreement with the change in Xe'' given by equation (**2.19**). Therefore, in spite of the ideal exponential dependence of $Jpy(0)$ on Veb, the substrate current turns out to deviate from ideality at high injection. The accuracy of equations 2.18 and 2.19 in predicting the substrate current is shown in 2.5.

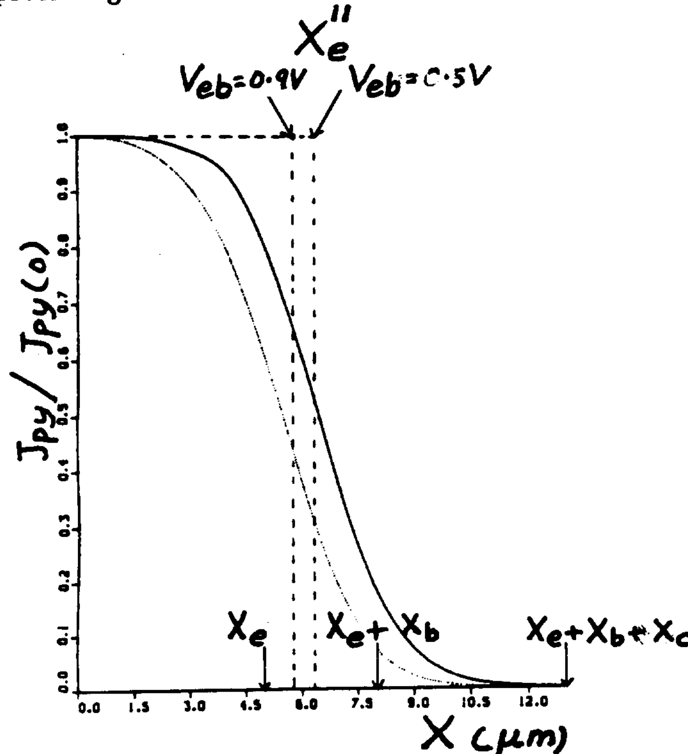


Figure 2.9: *The y component of the hole current distribution in the epilayer as a function of x at moderate (full line) and at high (dotted line) forward bias. Ohmic voltage drops excluded.*

2.3.3 The base current

The origins of the base current in the lateral pnp have been analysed by many authors in the literature[20,17] with many varying theories as to the relative significance of the various contributions. Numerical simulations on devices made in present day bipolar processes have shown that bulk recombination in the emitter and base regions no longer plays a significant role. Instead, the following contributions have been deemed to be important.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

At low forward bias the base current largely stems from recombination in the emitter-base depletion layer. This results in the well known non-ideal base current component which behaves according to $Ib \propto \exp\frac{Veb}{m\,Vt}$, where $1 < m < 2$ (this behaviour can be described well by the standard SRH[32] formulation).

At moderate forward bias, recombination at the *emitter contact* of electrons injected from the base into the emitter predominates, yielding an ideal exponential dependence on Veb. Only at extremely high forward bias can high injection effects give rise to a deviation from ideality, but this is beyond the region of interest in normal device operation (this is so because the current gain h_{fe} has already dropped to a fraction of its peak value at these bias conditions).

In order to analytically describe the base current at moderate to high forward bias, the base current can be expressed as:

$$Ib \;=\; Jny(0) \;\times\; X''_{cone} \tag{2.20}$$

where $Jny(0)$ = the electron current density along the left hand wall of the device of figure 2.1 (at $x = 0$)

X''_{cone} = an *effective* emitter contact width somewhat larger than X_{cone} shown in figure 2.1

Expressions for these two quantities will now be derived.

- $Jny(0)$: This is seen to be equal to the electron current density of the one-dimensional pnp transistor of figure 2.3. Neglecting high injection effects in the emitter, this current can be expressed as

$$Jny(0) \;=\; q\,D_{ne}\,\frac{dn}{dy} \tag{2.21}$$

where D_{ne} is the electron diffusion coefficient in the emitter. The value of the injected electron concentration in the emitter at the emitter-base junction (at $y = Ye$) is known and is given by $n_e = N_e \times n_0$. Here N_e is the emitter doping concentration and n_0 is the

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

reduced injected electron concentration in the emitter given by;

$$n_0 = \left(\sqrt{1 + 4(\frac{n_i}{N_e}^2)\exp\frac{Veb}{Vt}} - 1\right) / 2 \tag{2.22}$$

Because the injected electron concentration at the emitter contact (at $y = 0$), n_c, is not known, the minority carrier distribution in the emitter must be solved. Assuming zero bulk recombination in the emitter, this distribution is given in one-dimension by $\frac{d^2n}{dy^2} = 0$, which has as solution

$$n(y) = \alpha y + \beta \tag{2.23}$$

where α and β must be solved from boundary conditions. At $y = 0$, $n(0) = n_c \;\; => \beta = n_c$. At the emitter contact a mixed or 'Robbins' boundary condition, which is also used in CURRY and TRAP, can be applied. In one-dimension this yields the electron current density expressed as a function of the electron concentration at the contact, n_c, and the effective electron surface recombination velocity, s_n:

$$Jn = q\, s_n\, n_c \tag{2.24}$$

Now we can write

$$\begin{aligned} \alpha = \frac{dn}{dy} &= \frac{Jn}{qD_{ne}} \\ &= \frac{s_n\, n_c}{D_{ne}} \end{aligned} \tag{2.25}$$

Therefore

$$n(y) = \frac{s_n\, n_c}{D_{ne}}\, y + n_c$$

at $y = Ye$:

$$n(Ye) = n_e = \frac{s_n\, n_c\, Ye}{D_{ne}} + n_c \tag{2.26}$$

$$=> \; n_c = \frac{n_e}{1 + \frac{s_n Ye}{D_{ne}}} \tag{2.27}$$

From this expression for the injected electron concentration in the emitter, it can be seen that in the limiting cases of $s_n = 0$ and $s_n = \infty$ we have $n_c = n_e$ and $n_c = 0$ respectively. After substitution of equation (**2.27**) and equation (**2.25**) into equation (**2.21**) the result

© N.V. Philips' Gloeilampenfabrieken

becomes:

$$Jny(0) = \frac{q\, s_n\, n_e}{1 + \frac{s_n Ye}{D_{ne}}} \tag{2.28}$$

where n_e is be found using equation **(2.22)** and s_n is equal to the value used in numerical simulations ($s_n \approx 3 \times 10^5 cm/s$). Note that the limits of $s_n = 0$ and $s_n = \infty$ yield $Jny(0) = 0$ and $Jny(0) = \frac{q D_{ne} n_e}{Ye}$ respectively.

- X''_{cone}: The contribution to the base current from the electron current density in the region directly below the emitter contact is given by $Jny(0) \times X_{cone}$, where X_{cone} is the contact width shown in figure 2.1. The electron current density flow line pattern in the emitter is, however, not purely restricted to this region. Electrons injected vertically into the emitter from that part of the base region not directly below the emitter contact also contribute somewhat to the base current. The flow pattern in the emitter, to some extent visible in figure 2.5, resembles the fringing field in a parallel plate capacitor. Maxwell[33] found that the fringing field correction to the capacitance can be accounted for by adding a certain fixed fraction, f, of the plate separation to the plate length in the familiar expression for the capacitance. This suggests using an effective emitter contact width with the present problem:

$$X''_{cone} = X_{cone} + f \times Ye \tag{2.29}$$

Using the assumption that $Ye \ll (Xe - X_{cone})$, an assumption usually valid for practical devices, Hurkx[34] analytically derived that f was equal to $\frac{2ln2}{\pi} \approx 0.4413$. Figure 2.10 shows how well equation **(2.29)** approximates the y component of the electron current density distribution determined by numerical simulations. The ability of equation **(2.20)** to predict Ib is shown in section **2.5**.

Oxide interface recombination

In the above analysis the contribution to the base current from recombination at the $Si - SiO_2$ interface has been assumed negligible. The validity of this assumption is dependent on the quality of the oxide interface. Because this oxide quality is process dependent, the possibility of a significant oxide interface recombination component cannot be completely ruled out. Numerical simulations were therefore done by the author to examine the behaviour of this recombination current and the results are now summarised briefly.

© N.V. Philips' Gloeilampenfabrieken

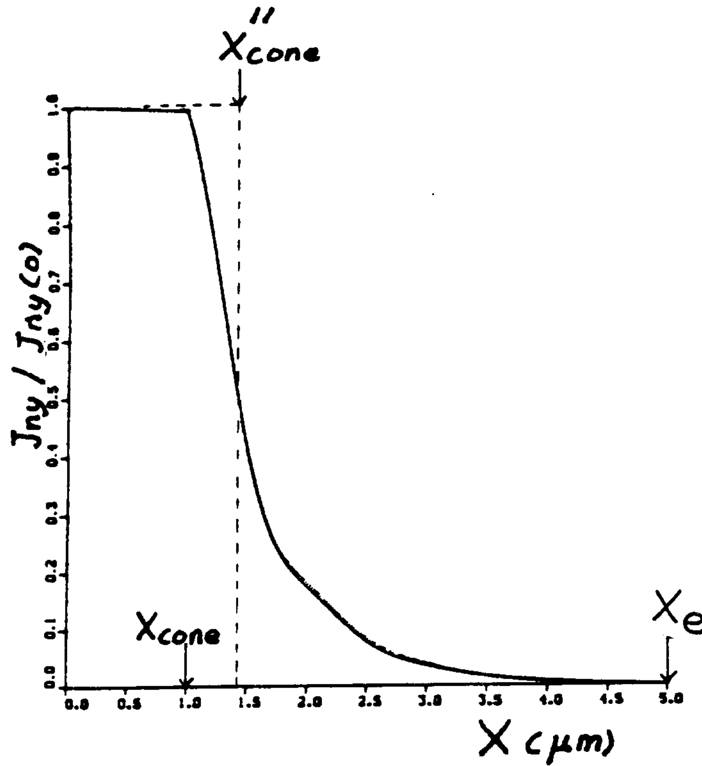


Figure 2.10: *The y component of the electron current distribution in the epilayer as a function of x at moderate (full line) and at high (dotted line) forward bias. Ohmic voltage drops excluded.*

In these simulations, the oxide interface recombination current, J_{ox}, was described by a SRH[32] type formulation[8] using an effective surface recombination velocity, s_{ox}. At low to moderate forward bias the oxide interface recombination in the neutral base region (and less importantly in the neutral emitter region) exhibited an ideal exponential dependence on Veb. In the emitter-base depletion region, however, the dependency was non-ideal and varied according to $\exp \frac{Veb}{mVt}$ where the non-ideality factor m was between one and two. The influence of a fixed charge density in the oxide region on these recombination currents was also investigated. Because both positive and negative values of this charge can exist depending on the processing and stressing conditions, both cases were investigated:

[8]More specifically:

$$J_{ox} = q \int_{X_{cone}}^{X_e + X_b} U_s$$

where

$$U_s = s_{ox} \frac{pn - n_i^2}{p + n + 2n_i} \quad (2.30)$$

This formulation is only strictly applicable to one value of the interface potential as the dependence of s_{ox} on this potential has been neglected.

© N.V. Philips' Gloeilampenfabrieken

- *Positive oxide charge density*: This had the effect of reducing the width of the emitter-base depletion layer under the oxide, thus causing a reduction in the ratio of non-ideal to ideal components. *Accumulation* in the neutral base region under the oxide resulted in a lower ideal component due to a reduced injected minority carrier concentration at the interface.
- *Negative oxide charge density*: As the magnitude of the negative oxide charge was increased[9], the ratio of non-ideal to ideal components increased. This was because the width of the emitter-base depletion layer increased and also because at sufficiently large values of oxide charge a depletion layer formed in the 'neutral' base region directly under the oxide. Further increases in the charge magnitude resulted eventually in inversion layer formation and it was interesting to note that a non-ideality factor, m, greater than two was observed in agreement with a recent publication by Ghannan and Mertens[35].

These simulations illustrated some interesting properties of the oxide interface and the effect of oxide charge. Even though the contribution to Ib of oxide interface recombination components is usually negligible (assuming good oxide quality), it may be worthwhile having some knowledge of the physics associated with this process when the oxide quality is suspect.

2.3.4 Charges, capacitances and transit times

The electron and hole distributions for the lateral pnp structure of figure 2.1 are shown, at moderate forward bias, in figure 2.11 and figure 2.12 respectively. Minority carrier concentrations in different regions of this lateral pnp, having been introduced somewhat in the last few sections, will be now be treated in some detail.

This minority charge storage in the neutral regions plays an important role in the cut-off frequency,[10] f_t. In the epilayer base regions both under the emitter and between the emitter and collector regions, a relatively large amount of minority hole charge is stored. At moderate forward bias, this charge storage predominates and is one of the main causes of low f_t values in

[9] In practice this increased charge could correspond to the injection of hot carriers into the oxide when the collector-base junction is subject to stressing bias conditions

[10] The common emitter cut-off frequency, f_t, is a figure of merit commonly used to characterise the high frequency behaviour of bipolar transistors. It is defined as the frequency at which the small signal current gain, $\hat{h}_{fe} = \frac{i_c}{i_b}$, has dropped to a value of one and therefore defines the maximum frequency at which the transistor can still be used as a gain element.

© N.V. Philips' Gloeilampenfabrieken

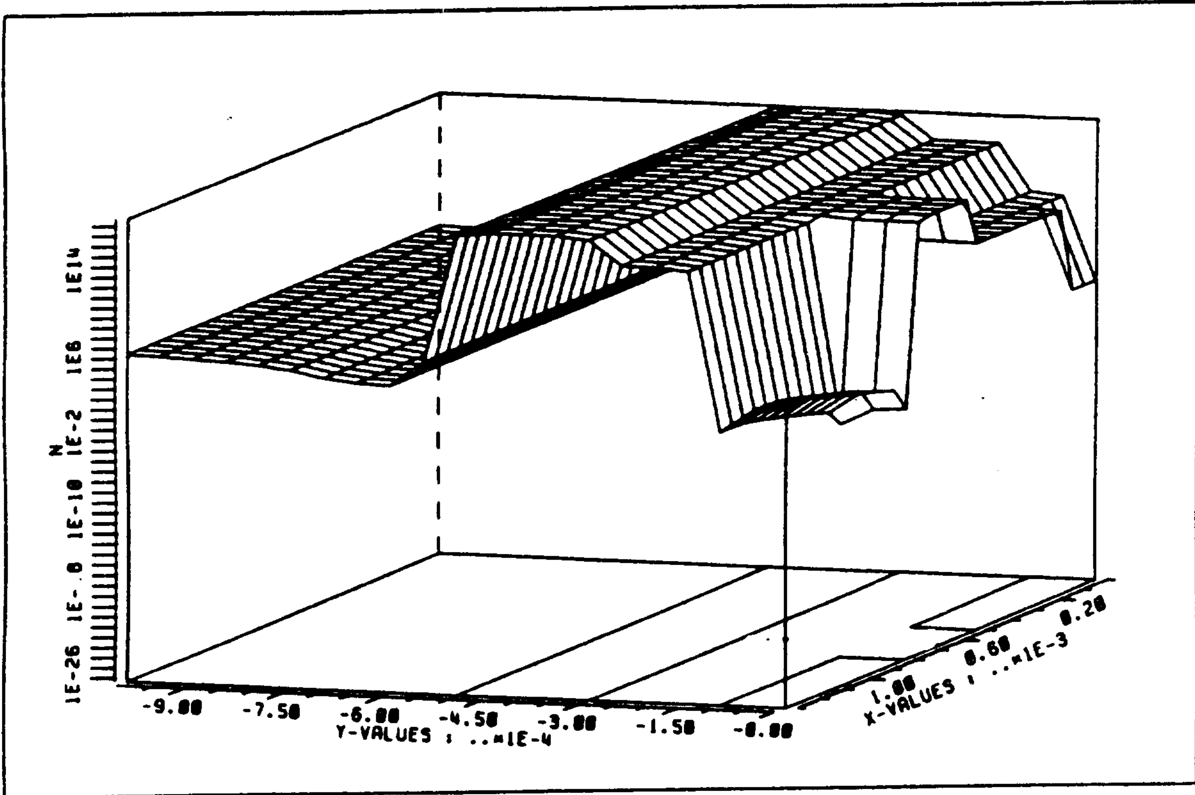


Figure 2.11: *Electron distribution at moderate forward bias.*

the lateral pnp compared to the npn transistor.

Figure 2.13 shows schematically the approximate average minority carrier concentrations in each region of a lateral pnp transistor. These regional divisions and their average values have been deduced from the results of numerical device simulations and also from the literature. In Last[36] and Eltoukhy[22], for example, the approximations for regions IV, V and VI can be found. The reduced electron concentration injected into the emitter, n_0, and the reduced hole concentration injected into the epilayer, p_0, have already been defined by equations 2.22 and 2.2 respectively. In the buried layer, the reduced injected hole concentration, p_{bn0}, is given by

$$p_{bn0} = \left(\sqrt{1 + 4(\frac{n_i^{\ 2}}{N_{bn}}) \exp \frac{Veb}{Vt}} - 1 \right) / 2 \tag{2.31}$$

The reason why the average minority charge in region VIII decreases at high injection is due to the behaviour of the hole concentration in region V which has been described in section **2.3.2**.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

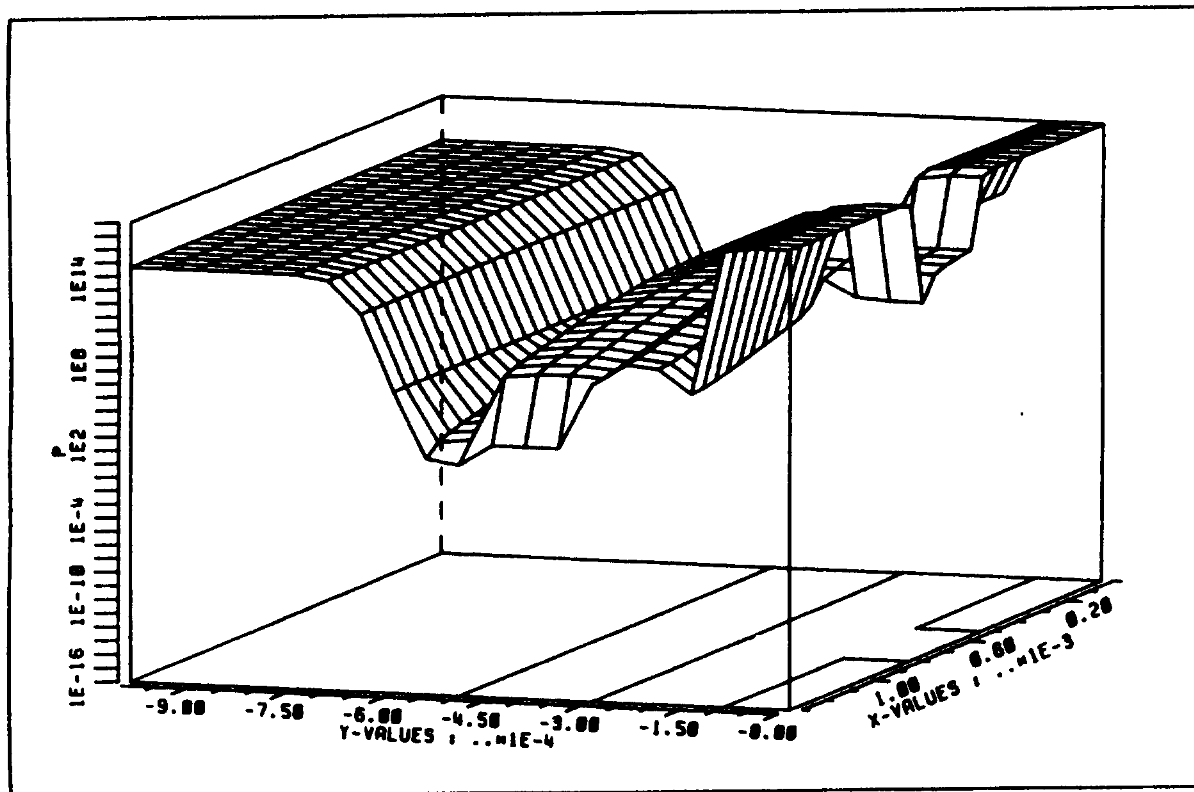


Figure 2.12: *Hole distribution at moderate forward bias.*

In order to relate this storage of minority carriers in the neutral regions to the cut-off frequency, f_t, transit times are introduced. The transit time, τ, represents the finite amount of time required by a charge carrier to travel through a particular region and it is defined as:

$$\tau = \frac{dQ}{dIc}|_{dVec=0} \tag{2.32}$$

The total transit time can be divided into the constituents originating from the regional division of minority carrier charge storage. Because the total transit time is related to f_t by

$$f_t = \frac{1}{2\pi\tau_{tot}} \tag{2.33}$$

we therefore have a method of relating stored charges to f_t. In addition to the contributions of the transit times in the neutral regions to the total transit time, we also have the contributions from the charging times of depletion regions. The total transit time, τ_{tot}, can therefore be

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

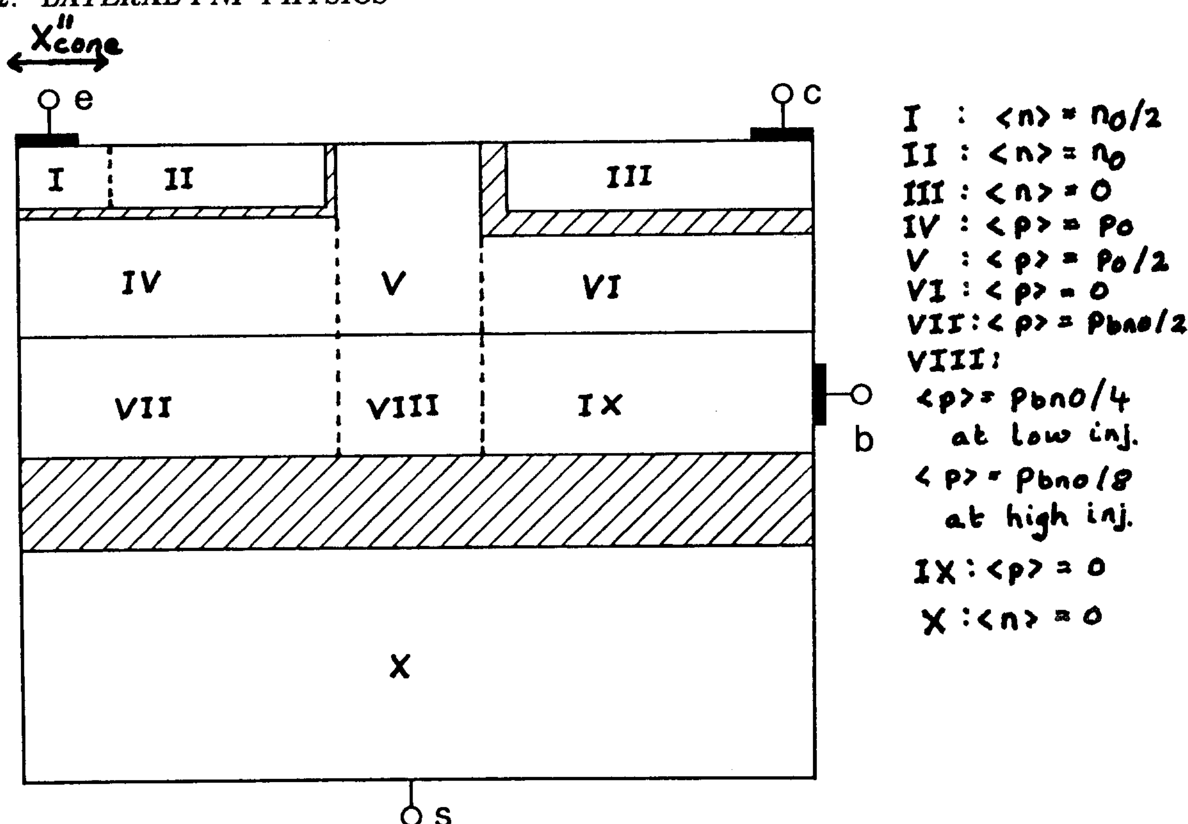


Figure 2.13: *Average minority concentrations. Ohmic voltage drops excluded.*

written as

$$\tau_{tot} = \tau_{eb} + \tau_{cb} + \tau_{sb} + \tau_{epi} + \tau_e + \tau_{bn} \tag{2.34}$$

where

τ_{eb}	=	the charging time of the emitter-base depletion capacitance
τ_{cb}	=	the charging time of the collector-base depletion capacitance
τ_{sb}	=	the charging time of the substrate-base depletion capacitance
τ_{epi}	=	the transit time due to charge storage in the epilayer
τ_e	=	the transit time due to charge storage in the emitter
τ_{bn}	=	the transit time due to charge storage in the buried layer

Neutral charge storage in the emitter-base depletion layer can be neglected[37].

Figure 2.14 shows the bias dependency of the total transit time and its individual components calculated by CURRY. The depletion charging times τ_{eb}, τ_{cb} and τ_{sb} (and the associated depletion capacitances) show the usual behaviour[37], dominating at low forward bias. As expected, the epitaxial base transit time, $\tau_{epi}(= \frac{dQ_{epi}}{dI_c})$, dominates at moderate forward bias but

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

also shows a *reduction* at high injection clearly visible in figure 2.14. This reduction by a factor of two of the base transit time is called the Webster effect[29] and is related to the enhanced minority carrier diffusion coefficient observed at high injection levels (see section **2.3.1**) [11].

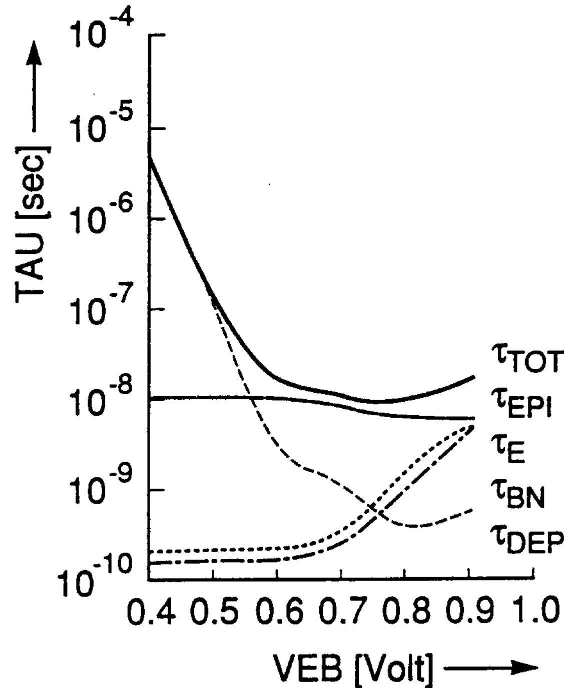


Figure 2.14: *Transit times as a function of forward bias. Ohmic voltage drops excluded.*

The transit times due to charge storage in the emitter $\tau_e(=\frac{dQ_e}{dIc})$ and the buried layer $\tau_{bn}(=\frac{dQ_{bn}}{dIc})$ are seen to *increase* at high injection levels. This increase is due to the fact that, in contrast to Ic, the minority carrier charge storage in these highly doped regions has an ideal exponential dependence on Veb up to very high forward bias.

When τ_{epi}, τ_e and τ_{bn} were calculated using the approximate average concentrations shown in figure 2.13 (the derivatives were calculated numerically), good agreement with the curves shown in figure 2.14 was obtained. Because of a lack of accuracy in the mesh used in numerical simulations, depletion charging times could not be compared to analytical formulations.

In the above discussion, ohmic voltage drops have not been taken into account. Figure 2.15 shows the case when these ohmic voltage drops are included in the numerical simulation. The

[11]In the case of the one-dimensional pnp transistor of figure 2.2, $\tau_{epi}=\frac{Xb'^2}{2D_p}$ which reduces at high injection to $\tau_{epi}=\frac{Xb'^2}{4D_p}$. See for example Chou[19]

© N.V. Philips' Gloeilampenfabrieken

only important difference with figure 2.14 is the increase in τ_{epi} at high current levels after the initial decrease due to the Webster effect. This is a consequence of the ohmic voltage drop across the emitter. A large portion of Q_{epi} is stored in the epilayer under the emitter and is determined by the emitter-base junction voltage along the bottom of the emitter. As a result, Q_{epi} is largely unaffected by this ohmic voltage drop. In contrast, the lateral component of Ic is reduced because it depends on the sidewall junction voltage which decreases due to this ohmic voltage drop.

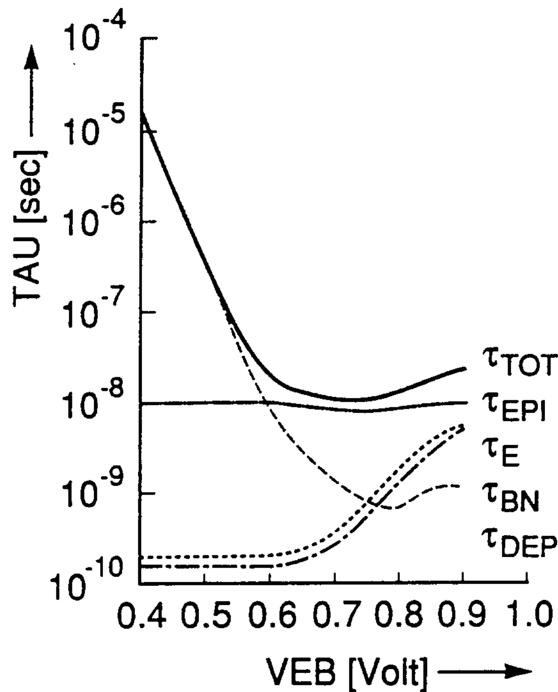


Figure 2.15: *Transit times as a function of forward bias. Ohmic voltage drops included.*

To summarise, it may be said that the relatively low f_t of the lateral pnp is due to the large charge storage in the neutral epitaxial base region. The increase of τ_{tot} at high collector currents and the corresponding fall off of f_t is determined by high injection in the epilayer (which results in an increase in τ_e and τ_{bn}) and by ohmic voltage drop across the emitter (increasing τ_{epi}). The Kirk effect[38] and quasi-saturation[39], which are important effects in the f_t fall off of npn transistors, could only become important at much higher collector current levels and are therefore not relevant for lateral pnp physics.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

2.4 Results

The analytical formulations developed in the previous few sections were combined to form the core of a computer program[14] to predict the device characteristics (current, Early voltages, capacitances, transit times, f_t, surface and bulk recombination) for forward active, reverse active and saturation conditions. Input to the program consists only of doping concentrations, temperature and geometry. The dependence on impurity concentrations and temperature of the physical parameters used in the descriptions of bandgap narrowing, recombination, etc. are equivalent to those used in CURRY. The application range of the program is restricted, however, to structures with abrupt, rectangularly shaped junctions and ohmic voltage drops are not included because they cannot be accounted for analytically.

Figures 2.16 to 2.19 show comparisons between the analytical predictions made by this program and the numerical predictions made by CURRY for the structure shown in figure 2.1. It is emphasised that no fitting of parameters was involved in obtaining any of these results.

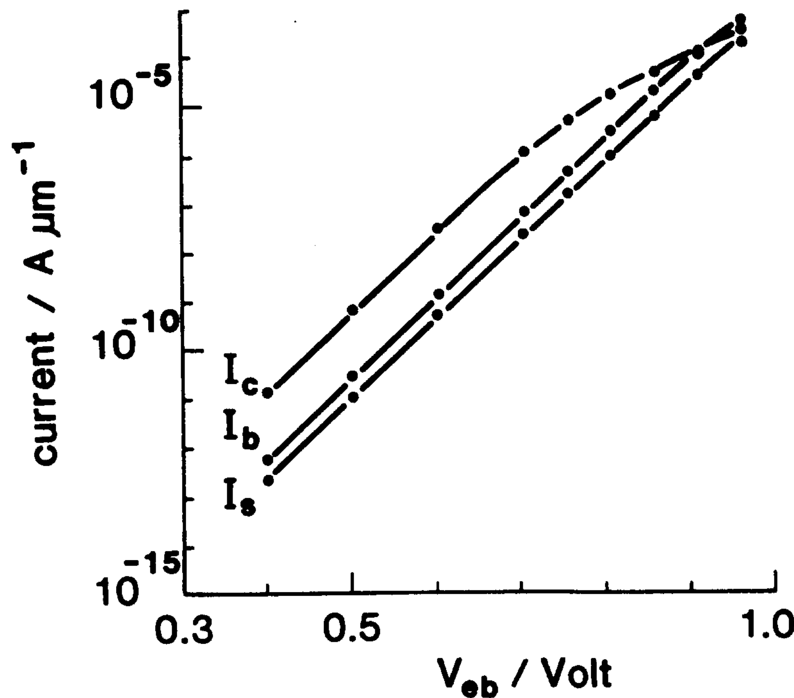


Figure 2.16: *Gummel plot comparison. Ohmic voltage drops excluded.*

Agreement between analytical and numerical predictions for the currents, depicted in the forward active Gummel plot of figure 2.16, was within about 10%. The current gains of the lat-

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

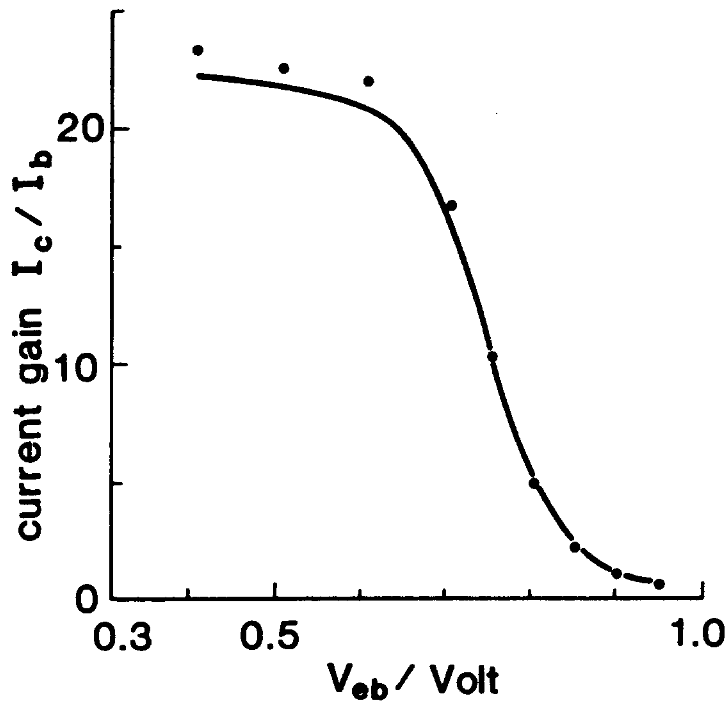


Figure 2.17: *Current gain h_{fe} comparison. Ohmic voltage drops excluded.*

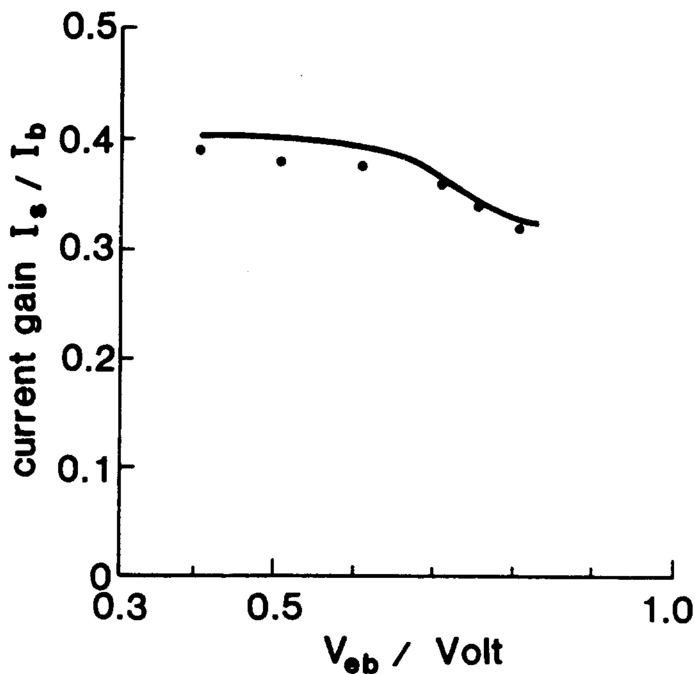


Figure 2.18: *Current gain h_{fs} comparison. Ohmic voltage drops excluded.*

eral (e-b-c) pnp transistor and the vertical parasitic pnp (e-b-s) transistor, shown in figure 2.17 and figure 2.18, also showed good agreement. In the latter plot, the maximum value of Veb shown is $0.8V$ because the influence of high injection in the base current leads to a discernable

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

increase in the gain characteristic at higher forward bias. Because ohmic voltage drops have been excluded in these results, $Veb = 0.8V$ corresponds to a much higher value of the actual applied emitter-base voltage in the case with ohmic voltage drops included. Therefore, values of Veb greater than about $0.8V$ in these plots are outside the region of interest.

Figure 2.19 shows the $Ic - Vec$ characteristic where it can be seen from the slope of the curves that the Early effect is modelled well.

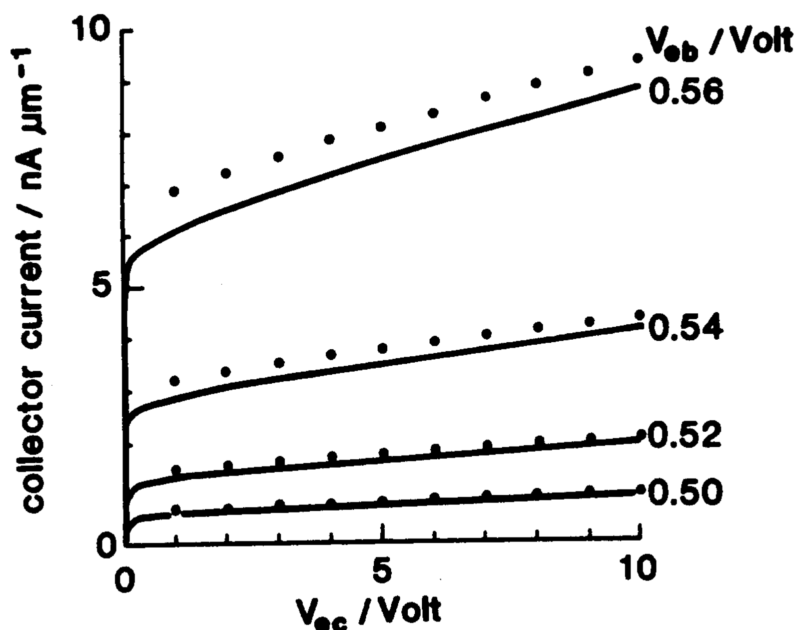


Figure 2.19: *$Ic - Vec$ comparison. Ohmic voltage drops excluded.*

2.5 Discussion

The investigation presented so far in this chapter has outlined the physical effects responsible for the behaviour of lateral pnp devices as predicted by numerical device simulations. In practice, when dealing with the actual lateral pnp integrated circuit device, certain effects may lead to deviation from this simulated behaviour. This section completes a description of lateral pnp physics by discussing the most important of these effects.

The first of these effects concerns the influence of current crowding on the Early voltage at high forward bias. Under current crowding conditions the collector current originates from

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

holes injected into the base region under the emitter contact. Because these holes now have a longer distance to travel to reach the collector, *the effective base width is increased.* From the analysis of the Early effect in section **2.3.1**, this effective increase in base width leads to an increase in the Early voltage, V_{eaf}. It can be concluded that the Early voltage in the lateral pnp is a function not only of Vcb but also of Veb.

Another effect concerns lateral pnp devices which employ junction isolation instead of oxide isolation. A schematic cross-section of such a device in which the reverse biased substrate-base junction acts to isolate the device from its surroundings is shown in figure 2.20. In this structure, highly doped p-type implantations make contact to the underlying substrate, thereby surrounding the device by a p-n junction which is always reversed biased to eliminate external interaction. When the lateral pnp device operates under forward active bias, some of the holes injected downwards by the emitter may leak along the epilayer-buried layer interface to reach the substrate without having passed through the highly doped buried layer. As a result, this extra component of the substrate current will be subject to similar high injection effects observed in the collector current. It should be noted that oxide isolation devices in which the buried layer does not completely separate the epilayer from the substrate (termed 'open' buried layer devices), will experience the same effect. If we denote the fraction of the substrate current which avoids passing through the buried layer as X_{hi} and if we, for the sake of clarity, neglect the second order high injection effect discussed in section **2.3.2**, current gain plots similar to those shown in figure 2.21 can be observed for the parasitic vertical pnp transistor. The fraction of the substrate current subject to high injection is dependent on the particular device geometry and processing. Note that with $X_{hi} = 0.5$, the gain characteristic tends to saturate after an initial fall off in that component of Is subject to high injection effects.

In the reverse active case a large substrate current flows due to the proximity of p-type isolation to the forward biased collector-base junction. The magnitude of this current may often be greater than both the emitter and base currents.

A temperature effect which tends to occur more often in junction isolated devices than oxide isolated, is the *zero base current crossover* phenomena at low forward bias. When the magnitude of the saturation current of the reverse biased substrate-base junction is significant, it compen-

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

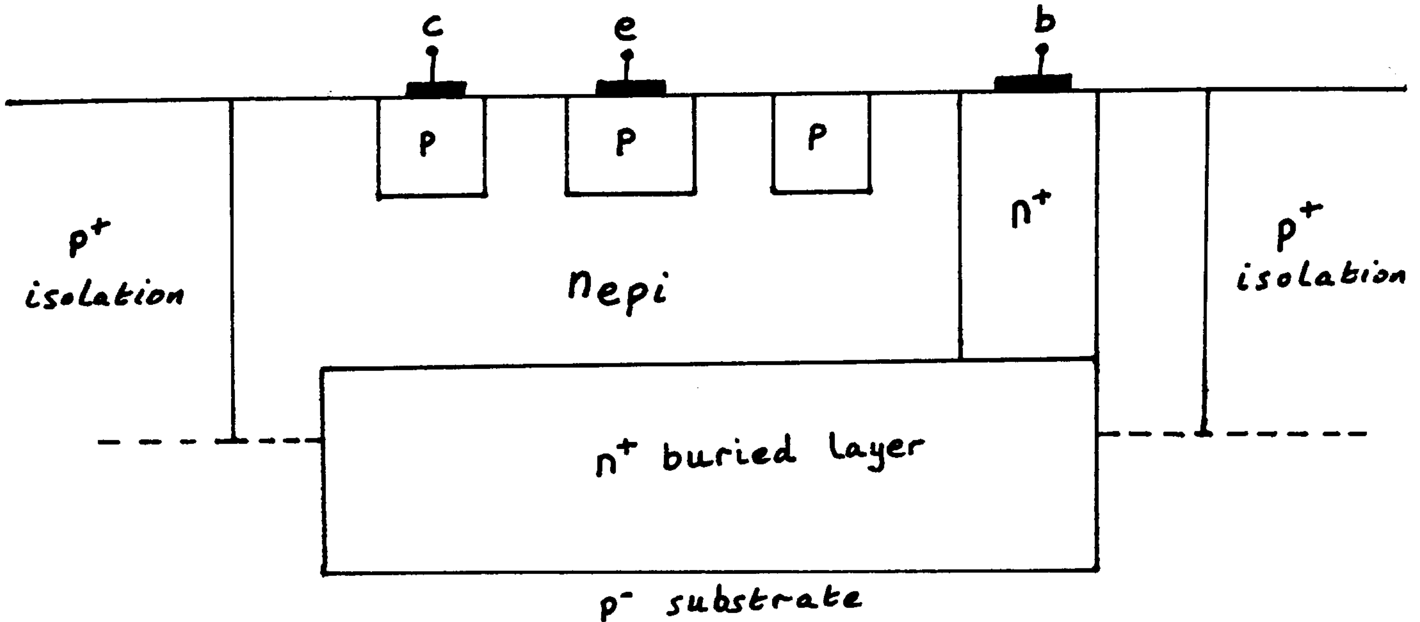


Figure 2.20: *Schematic cross-section of a junction isolated lateral pnp transistor.*

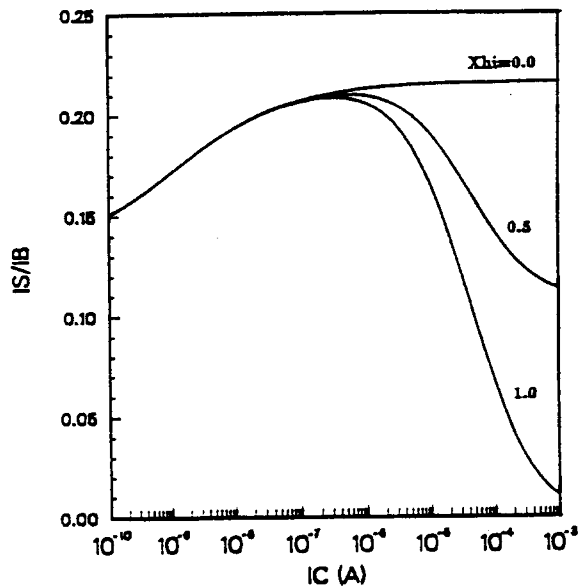


Figure 2.21: *Current gain of vertical parasitic pnp transistor with the fraction subject to high injection effects as parameter. The second order high injection effect is excluded here.*

sates the base current and may even cause it to change sign. In this case, the current gain, h_{fe}, is seen to increase dramatically in the low forward bias region. Due to the larger area of the substrate-base junction with junction isolation, the magnitude of its saturation current is larger

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

than with oxide isolation. However, this phenomena is basically a temperature effect and is usually only observed at temperatures much greater than room temperature. Due to its greater sensitivity to temperature, the saturation current of the substrate-base junction becomes comparable to Ib at higher temperatures, whereas at room temperature its effect is normally negligible.

One final point concerning differences between two-dimensional simulations and measurements on actual devices relates to the geometry of the third dimension which up to now has been assumed to have unit length having no effect on two-dimensional device behaviour. Basically three different layout geometries sketched in figure 2.22 are used in practice: stripe, U-shaped and cylindrical (octagonal) geometries.

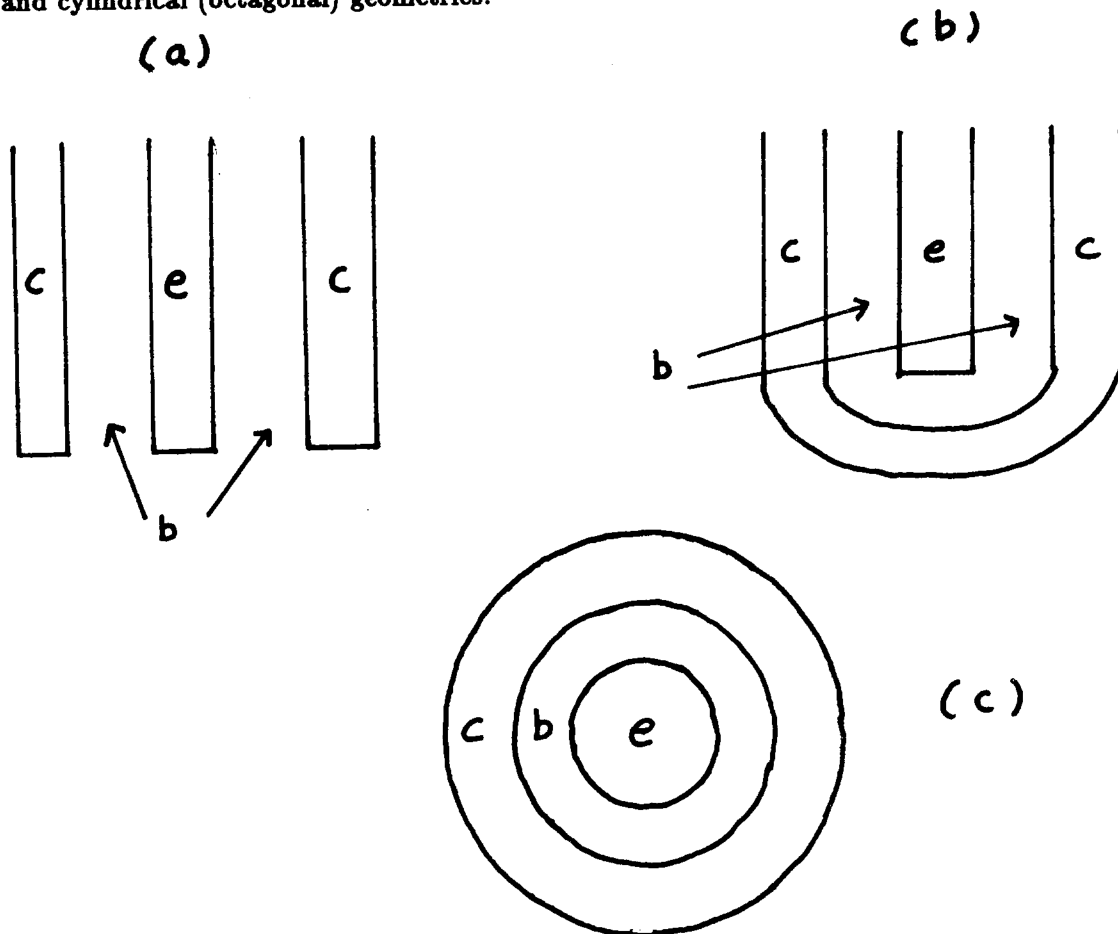


Figure 2.22: *Layout geometries: a) stripe b) U-shaped and c) cylindrical.*

With stripe geometry, the current gain and f_t values obtained from two-dimensional simulations are usually quite representative if the length of the third dimension is large enough. With cylindrical geometry the collector completely surrounds the emitter and measurements

© N.V. Philips' Gloeilampenfabrieken

on devices show better current gain and f_t values than those predicted by two dimensional simulations. This is because of the asymmetric scaling to three dimensions of two dimensional results for the collector current on one hand and the base current and charges on the other. In addition, the use of cartesian co-ordinates in two dimensional simulations underestimates the collector current compared to a more exact solution obtained from cylindrical co-ordinates. On the basis of these observations one can derive simplified scaling rules for cylindrical structures. The performance improvements obtained from two to three dimensional scaling of results for U-shaped devices lie somewhere between those of stripe and cylindrical devices.

2.6 Summary and conclusions

The device physics of the lateral pnp transistor has been investigated, primarily by means of numerical device simulations on a basic structure. By combining information from the literature with the results of these simulations it was possible to distinguish the various geometrical effects and physical mechanisms that determine the behaviour of devices made in present day bipolar processes.

Some aspects of the *geometry* and *doping levels* of the lateral pnp were found to have profound effects on device performance:

With reference to *geometry*, the mainly lateral flow of the collector current restricted the current handling capability of the lateral pnp as the injecting emitter area, at moderate forward bias, was only effective at the sidewall and corner regions. This, combined with the fact that the parasitic vertical pnp transistor was 'on' in forward active mode (with the resultant loss of injected holes to the substrate current), had a detrimental effect on the performance of the lateral pnp with respect to current gain. Lateral pnp A.C. behaviour, characterised by f_t, suffered from the large lateral dimensions which led to large depletion capacitances and charge storage in the neutral regions.

With reference to the *doping levels* of the lateral pnp, a *moderately doped* emitter had two main consequences for device performance:

- the emitter efficiency was rather low as electrons injected into the emitter from the base

© N.V. Philips' Gloeilampenfabrieken

constituted a base current which contributed to restricting the current gain, h_{fe}, to within a range of about 10 to 100.

- the lateral emitter resistance was significant and led to current crowding under the emitter at high current levels.

In addition, a *low homogeneous doping* level in the epitaxial base region caused the following effects:

- high injection effects dominated device behaviour when the injected hole concentration exceeded the epilayer doping concentration.
- charge storage in the epilayer regions between the emitter and collector and under the emitter led to a large epitaxial base transit time dominating the f_t characteristic. Typical values for f_t were restricted to within a range of about 10 to 100MHz.
- a significant Early effect resulted from the sensitivity of the collector-base depletion region, which extended more or less completely into the lowly doped epilayer, to the variation of the collector-base voltage. The Early voltage was typically between 5 and 50V.

It was found that, at high forward bias, high injection and current crowding dominated device behaviour.

- high injection led to a non-ideal exponential dependence of the collector current on the emitter-base forward bias. A similar behaviour in *part* of the substrate current was observed for certain structures in addition to a second order high injection effect. The transit time due to charge storage in the epitaxial base region, τ_{epi}, decreased by a factor of two at high injection (the Webster effect) whereas transit times due to charge storage in the neutral emitter and buried layer regions increased at high injection.
- current crowding under the emitter led to a reduction in the collector current, an increase in τ_{epi} and, because of an increase in the effective base width, an increase in the Early voltage at high forward bias.

In terms of the device characteristics, therefore, it can be said that the fall off in the current gain, h_{fe}, and f_t characteristics at high current levels was due to high injection and to current crowding under the emitter. The fall off in the current gain of the parasitic vertical pnp, h_{fs},

© N.V. Philips' Gloeilampenfabrieken

at high current levels was due to a second order high injection effect and with junction isolation or 'open' buried layers was also due to the normal high injection effect in a certain fraction of the substrate current. The Early voltage was found to be a function not only of Vcb but also, due to current crowding, of Veb.

When ohmic voltage drops were excluded it was shown that the currents and charges in the basic simulation structure could be modelled by analytical formulations. The collector, base and substrate currents could be expressed in terms of the current densities of one-dimensional pnp devices. Furthermore, the Early effect in a one-dimensional pnp transistor turned out to be a reasonably good approximation to the Early effect in the lateral pnp. Transit times in neutral regions could be modelled by employing regional approximations for minority carrier charge storage.

Due firstly, to the insight into lateral pnp physics afforded by numerical device simulations and secondly, to the good agreement between analytical formulations and numerical simulations on the basic simulation structure, it can be concluded that the device physics of the lateral pnp transistor is now well understood. The next stage of the project is use this knowledge to develop the set of equations and the equivalent circuit topology which constitutes a *physically* based compact model. This is the subject of discussion in the next chapter, where the formulation of a prototype compact model called Modella is described.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 3

Compact model formulation

3.1 Introduction

Having investigated lateral pnp physics in great detail, the next stage in the development of a physically based compact model is to use this knowledge to develop a prototype compact model. The development of this prototype is the subject of this chapter.

Advanced compact models for npn transistors (see for example [8]) are usually based on the Gummel concept. Here the collector current is expressed in terms of the majority charge in the neutral base, Q_b :

$$Ic \propto \frac{1}{Q_b} \tag{3.1}$$

This base charge, or equivalently the Gummel number, G_b, plays a key role in the behaviour of the compact model:

- the increase of Q_b at high injection gives rise to the familiar non-ideal exponential dependence of the collector current on Veb (i.e. $Ic \propto \exp\frac{Veb}{2Vt}$), above a certain knee current Ik.
- the base transit time, τ_b, is given by $\frac{dQ_b}{dIc}$.
- the Early voltage, V_{eaf}, is determined by the ratio of the collector-base depletion charge, Q_{tc}, to Q_b.

The Gummel concept, however, is based on a strictly one-dimensional derivation. This is usually suitable for the vertical npn transistor but simple arguments show that it fails for the

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

two-dimensional lateral pnp device. Consider a case where the collector is enlarged in the lateral direction, keeping the base width, Xb, fixed (see figure 2.1). According to the Gummel concept, the resulting increase in the base charge should lead to a decrease in Ic, an increase in τ_b and a change in V_{eaf}. In section **2.3** it was shown that the collector current flows mainly in the centre part of figure 2.1 and that virtually no minority charge is stored under the collector. Accordingly, a lateral enlargement of the collector would only marginally affect Ic, τ_b and V_{eaf}. In addition to these observations, the use of Q_{tc} to model the Early effect in the lateral pnp in inherently unphysical. This is because Q_{tc} mainly originates from the collector-base depletion region under the collector whereas the Early effect has been shown to be dependent on the depletion width at the collector-base sidewall. It may be concluded that, for a lateral pnp transistor, no unique value of Q_b exists which can simultaneously model the high injection effect in Ic, the base transit time and the Early effect.

Because the Gummel concept is unsuitable for the lateral pnp compact model, a new physically based approach is required. The analysis in chapter 2 will therefore be made suitable for compact modelling purposes. This compact MODEL for the LAteral pnp, called MODELLA (this is Italian for mannequin), will be presented in two stages:

1. the most important compact modelling equations will be derived for the forward active case in section **3.2**. Most of the equations apply to the one-dimensional transistors discussed in section **2.3**.

2. a combination with series resistances, to model the ohmic voltage drops, results in the equivalent circuit of Modella presented in section **3.3**. The dependence of the model equations on the various internal junction voltages which arise from the introduction of series resistances will also be discussed.

This presentation of Modella will be followed by a discussion of model behaviour in section **3.4** so that an insight into its rather complex functionality and sensitivity to parameter values can be gained. Some conclusions on this new physically based approach to lateral pnp compact modelling are then presented in section **3.5**.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

3.2 Derivation of the basic equations

3.2.1 The collector current and the Early effect

It was shown in chapter 2 that the high injection and Early effects in the collector current of a lateral pnp transistor can be effectively described in terms of a one-dimensional transistor with a homogeneously doped base (figure 2.2). This one-dimensional analysis will now be adapted to derive expressions more suitable for compact modelling. The manner in which these equations are formulated will take account of the appropriate choices of *physical* model parameters.

The collector current

Unfortunately, the exact expression given by equation (2.1) for the hole current is not suitable for compact modelling because of the computing time required for logarithms. By using two *additional* approximations in the derivation of this equation it is, however, possible to arrive at a simpler yet accurate expression. The electron and hole current densities are given by:

$$Jnx(0) = qD_n \frac{dn}{dx} + q\mu_n E(x) n(x) \tag{3.2}$$

$$Jpx(0) = -qD_p \frac{dp}{dx} + q\mu_p E(x) p(x) \tag{3.3}$$

The following assumptions, used also in the derivation of the exact solution, are made

- quasi-neutrality => $n(x) = p(x) + N_{epi}$ and $\frac{dn}{dx} = \frac{dp}{dx}$
- the electron current density is negligible at all injection levels:
  $Jnx(0) = 0$ => $E(x) = -\frac{D_n}{\mu_p} \frac{1}{p(x)+N_{epi}} \frac{dp}{dx}$.

  Using the Einstein relation and substituting into equation (3.3) we obtain:

$$Jpx(0) = -qD_p \left(1 + \frac{p(x)}{N_{epi} + p(x)}\right) \frac{dp}{dx} \tag{3.4}$$

Note that the diffusion coefficient effectively doubles at high injection when $p(x) >> N_{epi}$.

The two additional approximations mentioned above are as follows.

- the distribution of the injected hole concentration in the base is assumed to be linear. We can therefore replace $\frac{dp}{dx}$ by $\frac{N_{epi} p_0}{Xb}$, where p_0 is the reduced injected hole concentration in

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

the base at the emitter-base junction and is given by equation (2.2). The validity of this assumption is verified by fig. 7.3 of [31] in which the hole distribution closely resembles a linear distribution at *all* injection levels.

- $p(x)$ can be replaced by a suitable average value. The obvious choice with a linear fall off from $N_{epi}p_0$ to zero is $\frac{N_{epi}p_0}{2}$.

These assumptions result in:

$$Jpx(0) = -\frac{q\, D_p\, N_{epi}\, p_0}{Xb} \left(\frac{2+2p_0}{2+p_0}\right) \tag{3.5}$$

In order to manipulate this expression into a form suitable for modelling the collector current, we note that p_0 can be expressed as:

$$p_0 = \left(\sqrt{1 + 16\,\frac{I_f}{Ik}} - 1\right) / 2 \tag{3.6}$$

which leads to the following expression for the collector current:

$$Ic = \frac{4I_f}{3 + \sqrt{1 + 16\,\frac{I_f}{Ik}}} \tag{3.7}$$

where	I_f	=	the ideal reference current
		=	$I_{f0} \exp \frac{Veb}{Vt}$
where	I_{f0}	=	the saturation current parameter
		=	$\frac{qD_p n_i^2}{Xb\, N_{epi}}$
and where	Ik	=	the high injection knee current parameter
		=	$\frac{4qD_p N_{epi}}{Xb}$

The model parameters in this equation for Ic are I_{f0} and Ik. Because circuit designers are familiar with Gummel-Poon type formulations, the equation for Ic has been expressed in this way for ease of comprehension, rather than in terms of p_0. The deviation of the internal emitter-base junction voltage, $Ve'b'$, from Veb depends on series resistances which will be dealt with later in section 3.3.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

The maximum deviation of equation (**3.7**) from the exact solution (equation (**2.1**) and equation (**2.2**)) is only 4%, occurring as expected at the knee current value. This compares to a maximum deviation of 15% with the expression of Chou[17] who took $N_{epi}\, p_0$ as the average hole concentration instead of $\frac{N_{epi}\, p_0}{2}$. The Gummel-Poon expression is seen to deviate by a maximum of 13% from the exact solution. It should be noted that the Gummel-Poon expression, in contrast to the new formulation and that of Chou, does not account for the reduction of the base transit time by a factor of two at high injection (the Webster effect).

The Early effect

As described in section **2.3.1**, the Early effect arises from the dependence of the collector current on Vcb through variation of the collector-base depletion layer width. From equation (**2.1**), with the collector-base depletion width, Wcb, taken into account we find:

$$Ic \quad \propto \quad \frac{1}{Xb \; - \; Wcb} \tag{3.8}$$

Wcb is given by equation (**2.11**) for a one-sided abrupt junction i.e. $Wcb \propto \sqrt{Vd - Vcb}$ where the diffusion voltage Vd is given by[40] $Vd = 2\,Vt(ln\frac{N_{epi}}{n_i} - 1)$. If we denote the collector current in which the Early effect is not modelled as Ic', the collector current *with* Early effect can be written as

$$\begin{aligned} Ic &= \frac{Ic'}{1 - \frac{Wcb}{Xb}} \\ &= \frac{Ic'}{1 - \frac{\sqrt{1-\frac{Vcb}{Vd}}}{\gamma}} \end{aligned} \tag{3.9}$$

where $\gamma \quad = \quad Xb/\sqrt{\frac{2\epsilon\, Vd}{qN_{epi}}}$

Using the definition of the Early voltage $V_{eaf} = Ic|\frac{dVcb}{dIc}|$, we can write

$$V_{eaf} \quad = \quad 2\gamma\, Vd\sqrt{1 - \frac{Vcb}{Vd}}\, \left(1 - \frac{\sqrt{1 - \frac{Vcb}{Vd}}}{\gamma}\right) \tag{3.10}$$

At $Vcb = 0$ this expression yields

$$V_{eaf0} \quad = \quad 2\gamma\, Vd - 2\, Vd$$

© N.V. Philips' Gloeilampenfabrieken

$$=> \gamma = 1 + \frac{V_{eaf0}}{2\,Vd}$$

where V_{eaf0} is the Early voltage at zero collector-base bias. Substitution of this expression for γ into equation (3.9) yields

$$\text{Ic} = \frac{Ic'}{1 - \frac{\sqrt{1-\frac{V_{cb}}{Vd}}}{1+\frac{V_{eaf0}}{2Vd}}} \qquad (3.11)$$

In this equation the Early voltage at zero collector-base bias, V_{eaf0}, is the model parameter. This physical approach to describing the Early effect is, in principle, equivalent to Schneider[23] who also made direct use of the dependence of Ic on Wcb. In their formulation they chose an 'apparent' punch through voltage as the free modelling parameter. However, the choice of an Early voltage at $Vcb = 0V$ as the physical model parameter is intuitively more obvious.

The diffusion voltage, Vd, can easily be shown to be determined by the modelling parameters in Ic' according to

$$Vd = Vt\left(ln\frac{Ik}{4\,I_{f0}} - 2\right) \qquad (3.12)$$

In order to have one expression to describe both the forward and reverse Early effect, the emitter-base depletion width is taken into account in a manner similar to above. Because the continuity of functions and derivatives must be assured in a compact model, the square roots are modified to deal with zero or negative arguments. Replacing Veb and Vcb by their internal junction values $Ve'b'$ and $Vc'b'$, we obtain an expression suitable for modelling the high injection effect and the forward and reverse Early effect of a transistor with a homogeneously doped base:

$$\text{Ic} = \frac{4\text{I}_f \Big/ \left(3 + \sqrt{1 + 16\,\frac{\text{I}_f}{\text{Ik}}}\right)}{1 - \frac{\sqrt[4]{(1-\frac{V_{eb}}{Vd})^2+\delta}}{1+\frac{V_{ear0}}{2Vd}} - \frac{\sqrt[4]{(1-\frac{V_{cb}}{Vd})^2+\delta}}{1+\frac{V_{eaf0}}{2Vd}}} \qquad (3.13)$$

where			
	V_{eaf0}	=	the forward Early voltage at zero collector-base bias
	V_{ear0}	=	the reverse Early voltage at zero emitter-base bias
	δ	=	0.01

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

This equation leads to a forward Early voltage which is dependent on Vcb and a reverse Early voltage which is dependent on Veb in accordance with the observations in section **2.3.1**. V_{eaf0} and V_{ear0} are equal for a one-dimensional transistor and also for symmetric two-dimensional transistors as in figure 2.1. With some of the layout geometries in actual three-dimensional devices, like octagonal and U-shaped, no symmetry exists between emitter and collector. Because of the difference in curvature of emitter-base and collector-base junctions in these cases, the electric fields at the same applied junction voltages are different, leading to different values for V_{eaf0} and V_{ear0}.

An additional advantage of equation (**3.13**) is that punch through occurs at more physical values of Vcb. The approximate value of Vcb at which $Wcb = Xb$ is given by $Vcb \approx -\frac{V_{eaf0}^2}{4Vd}$ which is usually well above the avalanche breakdown voltage, BV_{ce0}. In the conventional modelling of the Early effect $\left(Ic = \frac{Ic'}{1+\frac{Vcb}{V_{eaf}}}\right)$, punch through occurs at very low values of Vcb, i.e. $Vcb = -V_{eaf}$, which leads to numerical problems.

3.2.2 The substrate current

Any formulation to model the substrate current, Is, should reflect the different types of behaviour associated with each of the hole current components reaching the substrate. Of the two components passing through the buried layer to the substrate, both exhibit an ideal exponential dependence on $Ve'b'$ up to very high forward bias but one exhibits a second order high injection effect which was described in section **2.3.2** by means of a decrease in the effective emitter width, Xe''. When the buried layer does not completely separate the epilayer from the substrate a third component exists (see section **2.5**) which stems from trajectories through a homogeneously doped base region with impurity concentration N_{epi}. The high injection effect occurring in this component can therefore be modelled by the formulation derived for Ic. If we denote the fraction of Is subject to this high injection effect as X_{hi} and the ratio between the saturation currents of Is and Ic as X_s, then the following formulation for Is is obtained.

$$Is = (1 - X_{hi})\, X_s\, I_f + \frac{4\, X_{hi}\, X_s\, I_f}{3 + \sqrt{1 + 16\, \frac{I_f}{Ik}}} \tag{3.14}$$

The model parameters are X_{hi} and X_s. Note that this equation will model current gain (i.e. h_{fe}) behaviour similar to that depicted in figure 2.21 because the fraction X_{hi} of Is will disappear

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

at high injection. The use of the *same* parameter Ik to model high injection in both Ic and Is can be understood by considering the fact that it is the ratio $\frac{I_f}{Ik}$ that is common to both equations. From section **3.2.1** this ratio is given by $\frac{I_f}{Ik} = \frac{n_i^2}{4N_{epi}} \exp\frac{Veb}{Vt}$ which is independent of any geometrical effects. High injection modelling is therefore directly related to Veb and is characterised by a knee voltage i.e. when $I_f = Ik \Rightarrow Veb = V_k$, the knee voltage[1]. This is physically correct because it is the injected hole concentration that determines when high injection effects occur and this concentration depends on Veb (equation (**2.2**)). The internal emitter-base voltage at which $I_f = Ik$ corresponds to a value of p_0 of $(\sqrt{1+16}-1)/2 \approx 1.56$. One knee current parameter can therefore be used to describe when the injected hole concentration becomes comparable to N_{epi} and, therefore can characterise the onset of high injection effects on both Ic and Is.

Since the decrease of Xe'' at high injection also leads to a reduction of Is by a certain factor at roughly the same value of p_0 (equation (**2.19**)) and since there is no way to distinguish this second order effect from the normal high injection effect using measured device characteristics only, both effects are modelled by equation (**3.14**). The fraction of Is subject to high injection effects, X_{hi}, will therefore increase somewhat to account for the second order effect.

3.2.3 The base current

The base current at moderate to high forward bias was shown in section **2.3.3** to have an ideal exponential dependence on $Ve'b'$. This component is modelled simply by

$$Ib = \frac{I_f}{\beta_f} \tag{3.15}$$

where the forward current gain, β_f, is the free modelling parameter. At low forward bias the base current exhibits a non-ideal dependence on Veb due to recombination in the emitter-base depletion region. Starting from the standard SRH[32] formulation the following equation, used

[1]Contrary to the analysis presented here, Berger[20] has suggested that V_k is a function of Xb, increasing for larger base widths. However, no references to such a dependence can be found in the rest of the literature and no such dependence is predicted by numerical device simulations. On closer examination of the paper presented by Berger, the author has found certain assumptions in this procedure to be rather dubious.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

also in Mextram, can be derived[15].

$$Ile = \frac{I_{bf}(\exp\frac{Veb}{Vt} - 1)}{\exp\frac{Veb}{2Vt} + \exp\frac{Vlf}{2Vt}} \tag{3.16}$$

The free parameters I_{bf} and Vlf are the saturation current and cross-over (from ideal to non-ideal behaviour) voltage respectively of the non-ideal base current. For $Veb << Vlf$ this equation predicts an ideal dependence on Veb whereas above Vlf, Ile varies according to $\exp\frac{Veb}{2Vt}$.

3.2.4 Charges, capacitances and transit times

A formulation to model the depletion charges which provides a physical description in the region of the diffusion voltage and is also suitable for circuit simulation purposes is the Poon-Gummel[41] expression.

$$Q_{tx} = \frac{-Cjx}{1-Px} \frac{Vdx - Vjx}{\{(1-\frac{Vjx}{Vdx})^2 + \delta\}^{\frac{Px}{2}}} \tag{3.17}$$

The model parameters are the diffusion voltage, Vdx, the zero bias depletion capacitance, Cjx, and the grading coefficient, Px. Here, x refers to e, c or s; the emitter-base, collector-base and substrate-base junctions. Vjx is the junction voltage and $\delta = 0.01$. Note that the equation and its derivatives are continuous functions of Vjx.

The storage of charge in the neutral emitter and buried layer regions was shown to have an ideal dependence on $Ve'b'$ up to very high forward bias. This charge storage can therefore be combined into one equation. Using the charge control principle we obtain:

$$Q_{fn} = \tau_{fn} I_f \tag{3.18}$$

The model parameter, τ_{fn}, is the forward transit time at low injection due to charge storage in the emitter and buried layer. It should be noted that the actual transit time which is given by $\frac{dQ_{fn}}{dIc}$ is not a constant. At low injection levels $\frac{dQ_{fn}}{dIc} = \tau_{fn}$ but at high injection levels $\frac{dQ_{fn}}{dIc} >> \tau_{fn}$ as was shown in section **2.3.4**.

To model the storage of injected hole charge in the epilayer by means of the charge control principle would be unphysical. At high injection levels, the dependence of the injected hole

© N.V. Philips' Gloeilampenfabrieken

concentration in the epilayer on $Ve'b'$ deviates from its ideal exponential dependence at low injection levels. To use the charge control principle would be to neglect this fact. Instead, the physical approach outlined in section **2.3.4**, relating the charge storage directly to the injected hole concentration, is used. Assuming a linear hole distribution in the base of the one-dimensional transistor of figure 2.2 and using equation (**2.2**) one can write:

$$\begin{aligned} Q_f &= \frac{1}{2}\, q\, Xb\, N_{epi}\, p_0 \\ &= \frac{Xb^2}{2D_p}\, \frac{q\, D_p\, N_{epi}}{Xb} \left(\frac{1}{2}\sqrt{1 + 16\,\frac{I_f}{I_k}} - \frac{1}{2}\right) \\ &= \tau_f\, \frac{I_k}{8} \left(\sqrt{1 + 16\,\frac{I_f}{I_k}} - 1\right) \end{aligned} \tag{3.19}$$

The free modelling parameter, τ_f, is the forward transit time at low injection due to charge storage in the epilayer. Due to the phy²ⁱᶜal approach taken in deriving this equation, the actual transit time $\frac{dQ_f}{dI_c}$ is seen to decrease by a factor of two at high injection, thereby modelling the Webster effect.

In a one-dimensional transistor the effect of collector-base and emitter-base depletion widths on Q_f can be accounted for by means of the same Early factor as used for Ic.

$$Q_f = \tau_f\, \frac{I_k}{8} \left(\sqrt{1 + 16\,\frac{I_f}{I_k}} - 1\right) \left(1 - \frac{\sqrt[4]{\left(1 - \frac{V_{eb}}{Vd}\right)^2 + \delta}}{1 + \frac{V_{ear0}}{2Vd}} - \frac{\sqrt[4]{\left(1 - \frac{V_{cb}}{Vd}\right)^2 + \delta}}{1 + \frac{V_{eaf0}}{2Vd}}\right) \tag{3.20}$$

Because of symmetry in the hole distribution, the reverse transit time, τ_r is nearly always equal to τ_f.

3.3 Modella

From one-dimensional analyses, the basic equations for use in Modella were derived in the last section. The two-dimensional aspects of the lateral pnp are incorporated by modifying and extending the basic equations to include the effects of series resistances, most notably current crowding. Modelling current crowding in a lateral pnp compact model is crucial because it affects the current gain h_{fe}, the cut-off frequency f_t and most dramatically the Early voltage V_{eaf}.

The choice of series resistances to use in Modella's equivalent circuit was based primarily

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

on the results of numerical device simulations. It was a compromise between minimising the number of internal nodes in the equivalent circuit and yet maintaining a high level of accuracy. Figure 3.1 shows the physical origins of these series resistances in the lateral pnp device. This figure also shows how the hole current flow lines which make up the collector current can be roughly divided into two distinct components: a purely lateral flow which originates at the emitter sidewall and a flow along curved trajectories which originates from the bottom of the emitter. At low current levels the sidewall component dominates. At high current levels the voltage drop across the lateral emitter resistance, $\mathbf{R_{ein}}$[2], leads to a reduced sidewall junction voltage thereby reducing the contribution from this component. The second component then dominates and current crowding in the region under the emitter contact is observed.

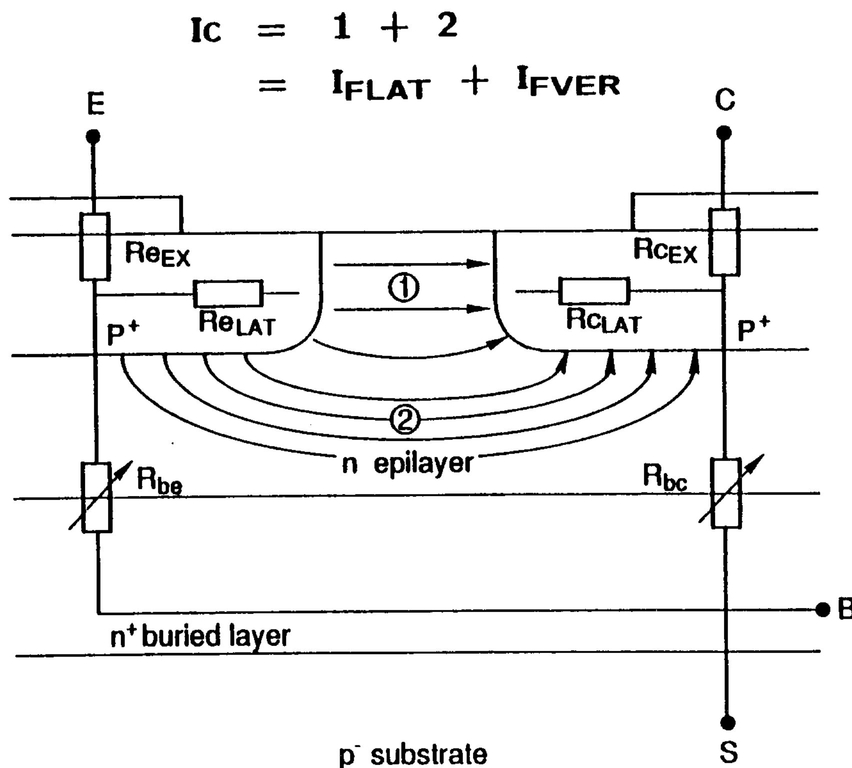


Figure 3.1: *Schematic cross section of a lateral pnp transistor showing the physical origin of the resistances chosen for the equivalent circuit. The two-dimensional hole current flow in the epitaxial base region, which constitutes Ic, is shown to consist of two components: 1. I_{flat} - a purely lateral flow originating from the emitter-base sidewall and 2. I_{fver} - a flow along curved trajectories originating from under the emitter.*

[2]Model parameters used in Modella will be appear in bold face from this point onwards.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

These observations lead us to Modella's equivalent circuit which is shown in figure 3.2. Apart from its symmetry, the most striking features of this equivalent circuit are the split emitter and collector series resistances along with double emitter and collector diodes and current sources (I_{flat}, I_{fver}, I_{rlat} and I_{rver}). Considering the forward active case, Modella models current crowding by using double emitter diodes and current sources to represent the sidewall (I_{flat}) and bottom (I_{fver}) components of I_c. I_{flat} and I_{fver} are therefore functions of the internal junction voltages $Ve1b$ and $Ve2b1$ respectively. These currents are all modelled by using the basic modelling equation for I_c (equation (**3.13**)) but with different internal junction voltages, saturation currents[3] and Early voltage parameters. At low current levels I_{flat} dominates due to its larger saturation current, whereas at high current levels I_{fver} takes over due to the voltage drop across $\mathbf{R_{ein}}$. Because of the larger effective base width that has to be crossed by I_{fver}, its Early voltage parameter (i.e $\mathbf{E_{afv}}$ - see Appendix A) will have a larger value than that of I_{flat} (i.e. $\mathbf{E_{afl}}$). This models the dependence of the Early voltage on Vcb because at low forward bias V_{eaf} (at $Vcb = 0$) is approximately equal to $\mathbf{E_{afl}}$ whereas at high forward bias V_{eaf} is much larger and depends on $\mathbf{E_{afv}}$[4].

In order to model the increase in r_B due to current crowding which was described in section **2.3.4**, the minority charge storage in the epitaxial base is also split into two components. These components represent charge storage under the emitter (Q_{fver} - with corresponding transit time $\mathbf{T_{fvr}}$) and charge storage between the emitter and collector (Q_{flat} - with corresponding transit time $\mathbf{T_{lat}}$). The V_{cb} dependency of the minority charge storage is modelled by adapting the sidewall component, Q_{flat}, to include the effect of the collector-base depletion width on the base width. This modulation of charge, defined in equation (**3.20**), does not apply to Q_{fver} because the charge stored in the epilayer under the emitter is obviously independent of Vcb. The result of formulating the epilayer base charge in this way is that the ratio between $\mathbf{T_{lat}}$ and $\mathbf{T_{fvr}}$ influences the dependence of the cut-off frequency f_t on Vcb and the fall off of f_t at high forward bias.

A complete description of Modella's equations and parameters can be found in Appendix A.

[3]The reciprocity theorem demands that the same saturation current $\mathbf{I_s}$ applies to the total forward current $If(= I_{flat} + I_{fver})$ and to the total reverse current $Ir(= I_{rlat} + I_{rver})$. The ratio at low injection between bottom and sidewall components is characterised by means of the parameters $\mathbf{X_{ifv}} = I_{fver}/If$ and $\mathbf{X_{irv}} = I_{rver}/Ir$. These fractions depend on geometry (see equation (**2.8**)) and are usually in the region of $10 - 20\%$.

[4]The increase in V_{eaf} at high forward bias depends of course on how much current crowding is present and consequently its value also depends on $\mathbf{X_{ifv}}$ and $\mathbf{R_{ein}}$.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

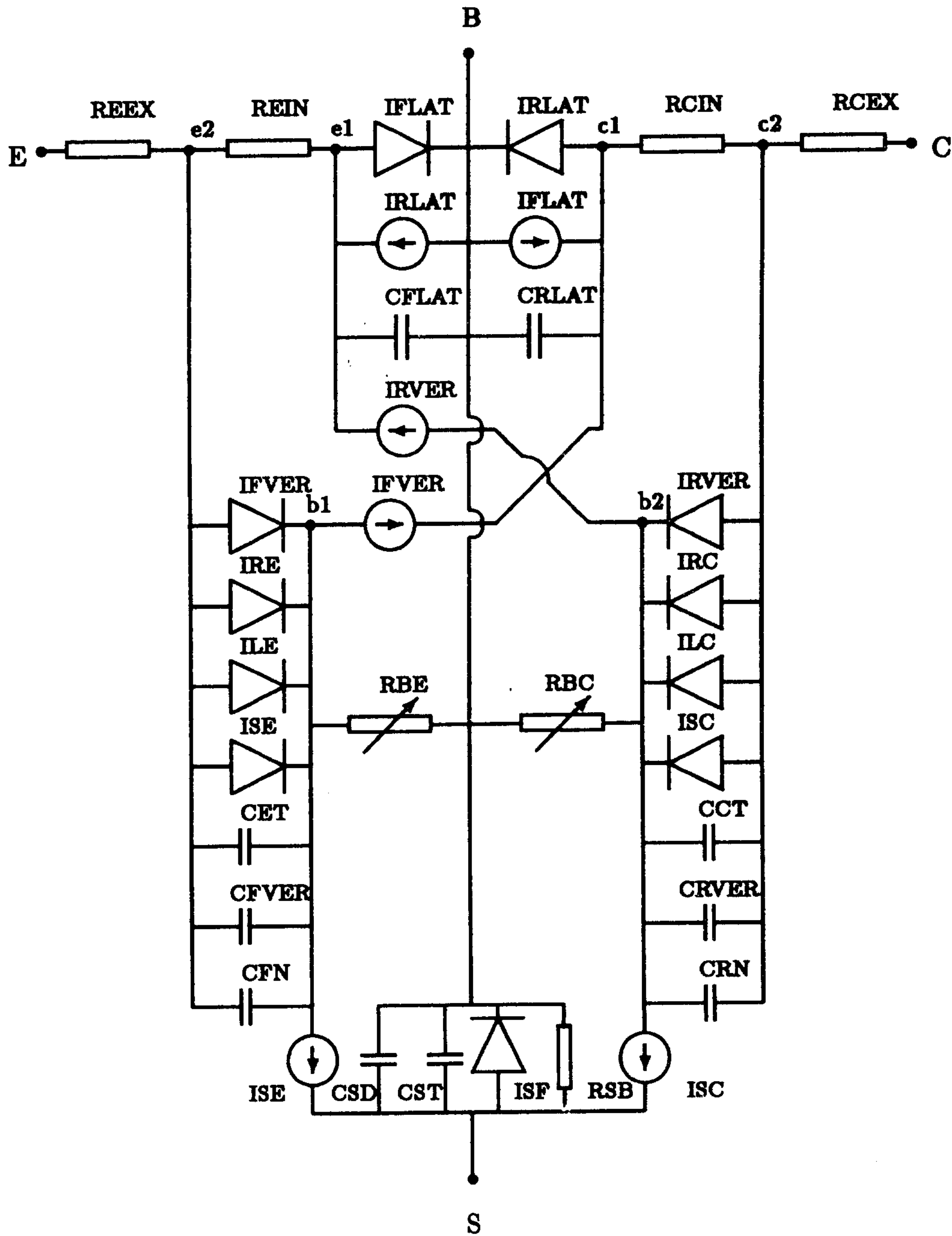


Figure 3.2: *Equivalent circuit diagram of Modella.*

© N.V. Philips' Gloeilampenfabrieken

It can be seen from these equations that one knee current parameter is used to describe the effect of high injection in the epilayer on $I_{se}, I_{sc}, I_{flat}, I_{fver}, I_{rlat}$ and I_{rver} (see section **3.2.2**). This parameter can also used to describe the conductivity modulation of the epilayer base resistances *Rbe* and *Rbc*. The resistivity ρ in the epilayer under the emitter and collector is inversely proportional to the electron concentration in these regions. Consider *Rbe*:

$$\begin{aligned} Rbe \ \propto \ \rho &= \frac{1}{q\mu n} \\ &= \frac{1}{q\mu(N_{epi} + N_{epi}p_0)} \\ &= \frac{\rho_0}{1+p_0} \\ &= \frac{2\rho_0}{1+\sqrt{1 \ + \ 16 \times \frac{If2}{\mathbf{I_k}}}} \end{aligned} \tag{3.21}$$

where $\rho_0 = \frac{1}{q\mu N_{epi}}$ and $\mu =$ electron mobility.

With regard to the substrate-base junction, an additional model parameter $\mathbf{I_{ss}}$ is used for the saturation current of the substrate-base diode current I_{sf}. This facilitates the modelling of the zero-base crossover phenomena discussed in section **2.5**. Also, an unexplained ohmic leakage current between substrate and base reported in some processes is modelled by the resistance $\mathbf{R_{sb}}$. The diffusion capacitance *Csd* has been included in Modella simply to prevent designers from unintentionally applying a forward bias to the substrate-base junction (the associated time constant *Tsd* is not a model parameter but is instead set to a large fixed value).

In summary, the equivalent circuit of Modella reflects the structural symmetry in the lateral pnp. However, the use of separate model parameters in reverse active operation allows Modella to model asymmetrical behaviour in reverse active operation with equal accuracy. The knowledge of device physics outlined in the last chapter dictated the equations and the locations chosen in the equivalent circuit for diodes, resistances, capacitances, etc. A brief description of the behaviour of this physical model follows in the next section. By combining this description with qualitative illustrations of selected parameter influences it is hoped to give an insight into its rather complex functionality.

© N.V. Philips' Gloeilampenfabrieken

3.4 Behavioural description and parameter dependencies

In order to investigate the behaviour of Modella in detail, it was incorporated into the program Biplot[42]. This is a program which plots user defined quantities as a function of terminal settings and/or model parameter values. User defined quantities refers to basically any function defined within the model or derived from information within the model. Some examples are internal currents and voltages, derivatives of model functions with respect to internal/external voltages, terminal quantities like Ic and derived quantities like Early voltages and transit times. The terminal settings which can be varied are those which define the region of operation of the device i.e. Ic, Ie, Ib, Is, Veb, Vcb and Vec.

This facility is invaluable, especially in the initial stages of model development, in that it provides a straightforward means to investigate model behaviour. Bugs in the model implementation are found quickly by observing inconsistencies in plots. The continuity of model functions and their derivatives can also be investigated using Biplot by checking graphically for the absence of singularities in user defined functions in all regions of operation. Figure 3.3 shows an example of this screening procedure. Here the depletion charge formulation is seen to be a continuous function of the voltage Veb and the parameter $\mathbf{V_{de}}$ in the range chosen. This systematic graphical testing of continuity proved to be an efficient method to confirm that Modella's equations and their derivatives did not contain singularities. These singularities, which could lead to numerical problems, are normally both time consuming to find and can result in a lack of trust in the model by designers when circuit simulations using the model fail to converge or yield meaningful results.

The use of Biplot to investigate sensitivity to model parameter values was also invaluable. This can of course only be used as a qualitative indication because changing the value of one parameter, with the rest of the parameter set unaltered, is unphysical as the correlation between parameters is disregarded. However, this qualitative information does lead to a better understanding of both the model behaviour and the sensitivity of device characteristics to parameters[5]. A few examples of the effects of parameter variations follow in order to illustrate the insight

[5]This latter point is also useful when developing the parameter determination strategy because the optimisation of parameters can be tuned to the regions of the device characteristics which exhibit the highest sensitivity to the parameters in question.

© N.V. Philips' Gloeilampenfabrieken

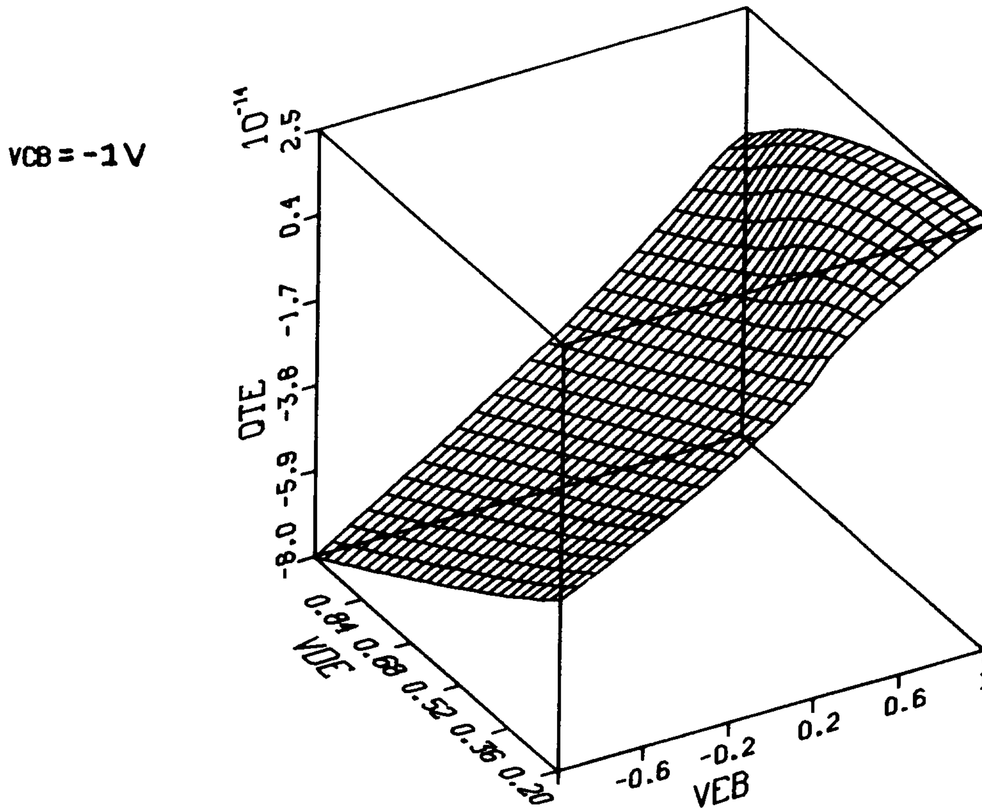


Figure 3.3: ***Emitter-base depletion charge as a function of emitter-base bias and the parameter*** $\mathbf{V_{de}}$.

afforded by this approach.

1. **Figure 3.4 shows the simple example of varying the knee current parameter $\mathbf{I_k}$ which leads directly to a shift in the onset of high injection effects in the epilayer, illustrated here using the current gain h_{fe}.**

2. **Because of ohmic voltage drop across $\mathbf{R_{eln}}$ (leading to current crowding under the emitter) the ratio between I_{flat} and I_{fver} reduces at high current levels. Figure 3.5 shows how an increase in the lateral emitter resistance increases this current crowding effect.**

3. **Figure 3.6 shows how the definition of the Early voltage influences the value of V_{eaf} at current levels where ohmic voltage drops are important. The definition presented in section 2.3.1** (i.e. $V_{eaf} = Ic\,|\frac{dVcb}{dIc}| + Vcb$) **can be evaluated keeping either Ib or Veb constant and figure 3.6 shows how this choice influences the result.**

   When Veb is kept constant in evaluating the derivative, V_{eaf} is larger because the resistance $\mathbf{R_{eex}}$ affects the interpretation of the Early effect. An increase of δVcb in Vcb leads, through the normal Early effect, to an increase in Ic which leads to increased ohmic

© N.V. Philips' Gloeilampenfabrieken

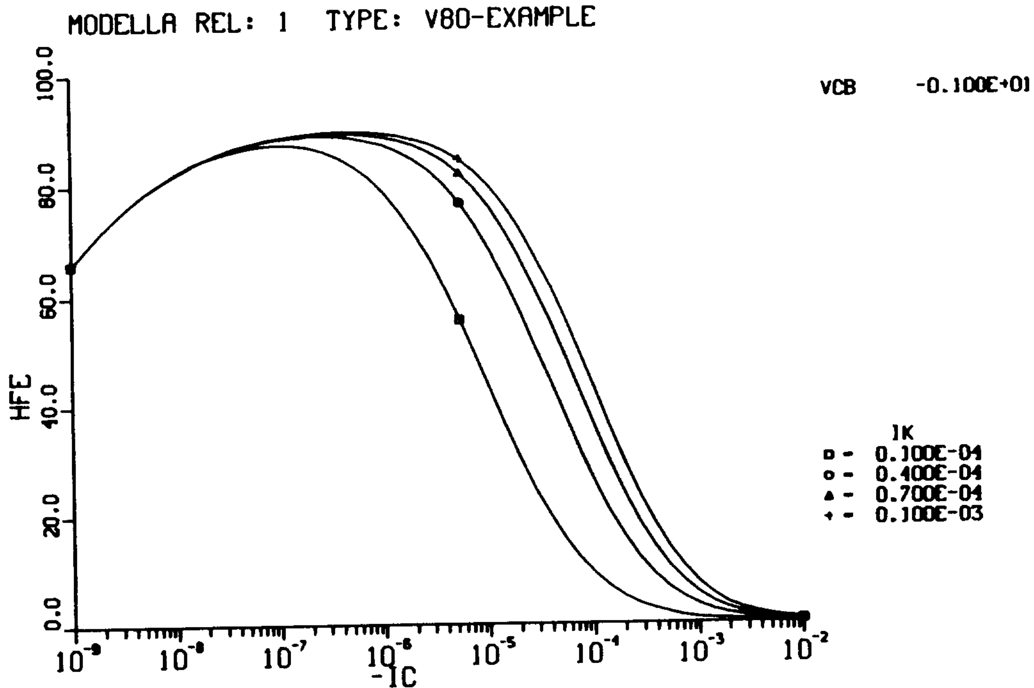


Figure 3.4: *Dependence of the current gain characteristic on the high injection parameter* $\mathbf{I_k}$.

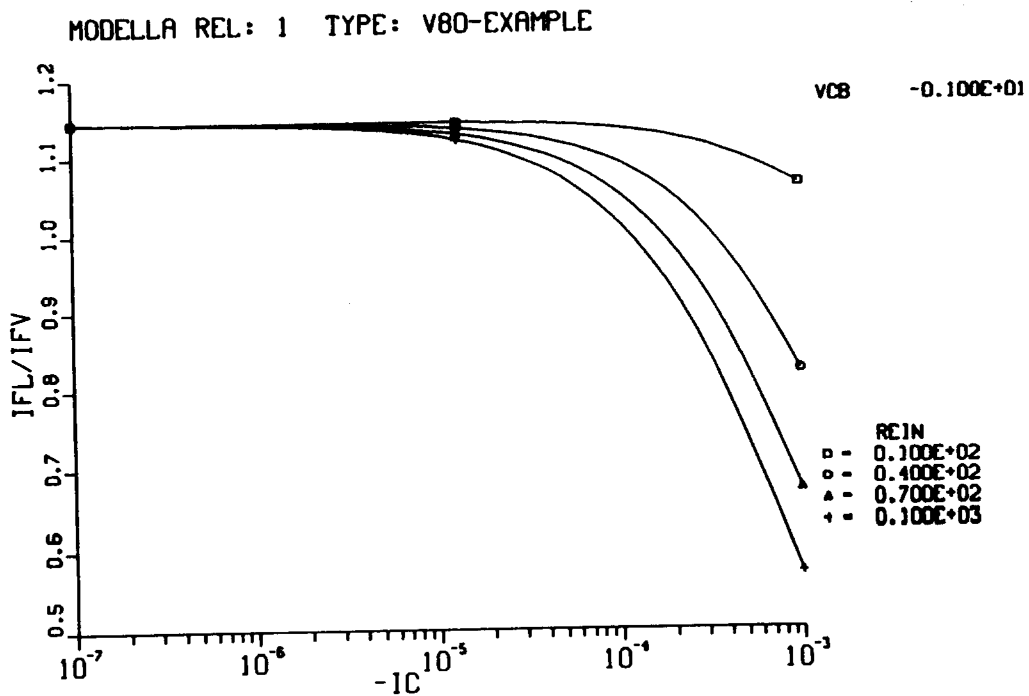


Figure 3.5: *Ratio of the sidewall and bottom components of Ic as a function of current crowding (quantified by the parameter* $\mathbf{R_{ein}}$*).*

voltage drop across $\mathbf{R_{eex}}$. The internal voltage $Ve2b1(\approx I_{flat}\mathbf{R_{ein}} + Ve1b)$ is therefore reduced leading in turn to a reduction in Ic which compensates the increase due to the normal Early effect. When Ib is kept constant, $Ve2b1$ is by definition also constant and

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

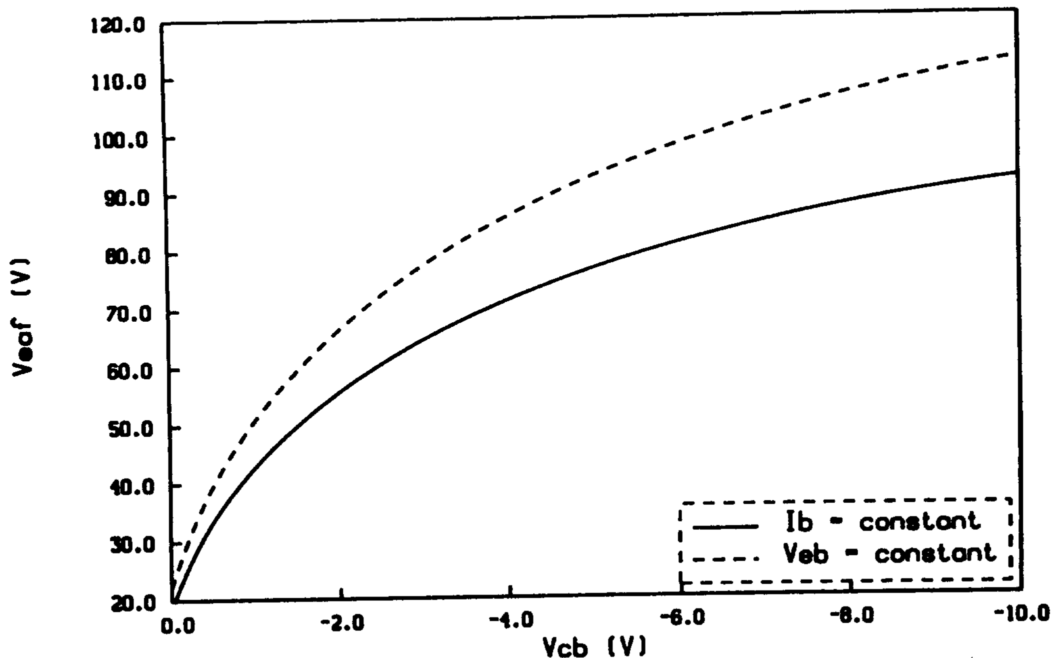


Figure 3.6: *The Early voltage V_{eaf} at a high current level as a function of Vcb. The choice of constant Veb or Ib in the definition of V_{eaf} leads to different curves.*

therefore the voltage drop across $\mathbf{R_{eex}}$ does not cause an overestimation of the Early voltage V_{eaf}[6]. The conclusion is that keeping Ib constant leads to closer estimation of the true Early effect.

4. Figure 3.7 shows the common emitter cut-off frequency versus Ic where the parameter $\mathbf{T_{fvr}}$ is varied. This variation corresponds to an increase in the charge stored in the epilayer under the emitter and its influence on f_t is clear.

3.5 Conclusions

From a knowledge of lateral pnp device physics, a new compact model has been derived which does not avail of either the Gummel concept or the charge control principle. The basic model equations were derived from analyses of a 1-D lateral pnp transistor and a 1-D parasitic vertical pnp transistor. 2-D effects were incorporated by modifying and extending the basic model equations to include the effects of series resistances, most notably current crowding. The resulting compact model has 6 internal nodes and 43 parameters. Because of the physical approach used

[6]It should be noted that by the same reasoning the ohmic voltage drop across $\mathbf{R_{ein}}$ will cause an overestimation of the true V_{eaf}. However, at these current levels, current crowding usually predominates leading to a reduction in the contribution to Ic from the current component which flows through $\mathbf{R_{ein}}$.

© N.V. Philips' Gloeilampenfabrieken

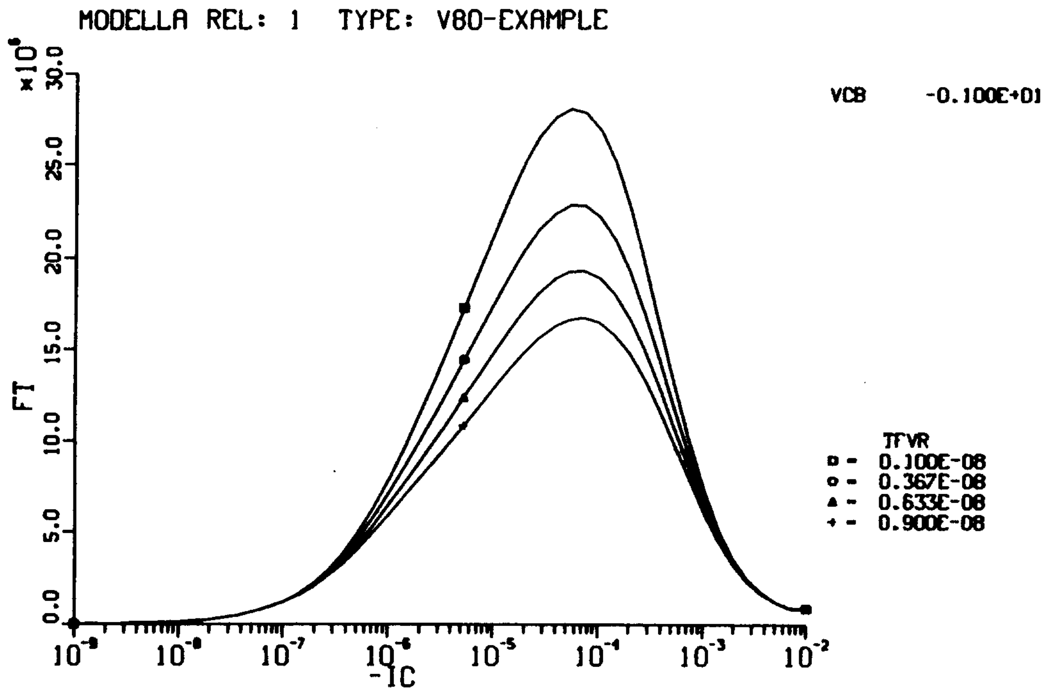


Figure 3.7: ***The common emitter cut-off frequency characteristic as a function of the parameter*** $\mathbf{T_{fvr}}$.

to derive Modella, it should possess all the advantages that physical formulations have over present semi-empirical approaches (see section 1.3).

With respect to the continuity of the equations and derivatives in Modella, the program Biplot has been used to graphically illustrate the absence of singularities in wide regions of transistor operation. This investigation has therefore helped to validate model continuity and its ability to perform 'reasonably' in extreme regions of operation - regions often encountered during the iterative solution procedure in circuit simulation programs. The Biplot package has also given an insight into the behaviour of Modella and the sensitivity of device characteristics to parameter variations. This latter aspect is important in the development of the parameter determination strategy which is the topic of the next chapter.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 4

Parameter determination

4.1 Introduction

The ability to extract good parameter sets is crucial when developing and using a compact model. Without a good parameter set a compact model is useless regardless of the extent to which it has been based on device physics. The determination of these parameter sets though is an involved task, not only because of the restrictions inherent in only being able to measure *terminal* device behaviour, but also because parameter sets are not unique. This latter problem is due to *measurement induced correlation* which means that the various device phenomena, characterised by certain model parameters, cannot always be clearly distinguished from each other in the measured characteristics. In some cases, the resulting parameter values may not be independent from each other but instead may depend on other parameter values which have similar effects on the characteristics. This type of parameter correlation is unphysical despite the fact that parameter sets containing conflicting emphases on certain device phenomena often appear to fit measured data equally well. Examples of this correlation include current crowding under the emitter and high injection effects in the epilayer which result in similar h_{fe} and f_t fall off behaviour at high current levels.

The decision as to what constitutes a 'good' parameter set cannot therefore be based exclusively on the fit to measured data. A 'good' parameter set should yield a fit to measured data with an overall error less than say 10% *and* it should contain parameter values which are physically plausible. This latter point demonstrates an advantage of a physical compact model compared to an empirical one since physical plausibility has no real meaning in relation to

© N.V. Philips' Gloeilampenfabrieken

empirical fitting parameters.

The rest of this chapter is concerned with a description of the parameter determination strategy which has yielded the best parameter sets to date. This strategy was developed from a combination of trial and error, theoretical analysis, numerical device simulations and sensitivity investigations using the program Biplot[42]. Section 4.2 introduces parameter determination methodology using the program Bipar[43] and this leads, in section 4.3, to a detailed description of the strategy now used for Modella to determine each of the model parameters. This strategy description includes a practical example using a junction-isolated lateral pnp device. The discussion in section 4.4 deals with the important topic of how Modella's physical formulation can be used to aid the parameter determination process and the chapter then ends with some conclusions.

4.2 Parameter determination with Bipar

The specific measurements used with a parameter determination strategy vary from one compact model to another but a basic approach for the lateral pnp generally involves optimising parameters to fit the following measured device characteristics:

- D.C. parameters
  - forward Gummel plot (Ic, Ib, Is .vs. Veb with $Vcb = -1V$).
  - reverse Gummel plot (Ie, Ib, Is .vs. Vcb with $Veb = -1V$).
  - forward Early plot ($Ic(Vcb)$ at constant Veb).
  - reverse Early plot ($Ie(Veb)$ at constant Vcb).
  - $Ic(Vec)$ characteristic.
- Charge parameters
  - emitter-base, collector-base and substrate-base depletion capacitances .vs. voltage measurements.
  - the cut-off frequency characteristic ($f_t(Ic)$) .

The optimisation process can be either sequential or simultaneous. In the sequential approach a small number of parameters (usually 1-3) are optimised by fitting the model response/prediction to a specific range of a measured device characteristic. Simultaneous optimisation involves optimising all parameters together to fit all the measured data at the same

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

time. The advantage of doing one major optimisation to get a set of parameters from a given measured data set is that the physical correlation between parameters is accounted for in a natural/automatic way. The disadvantages are as follows :

- convergence is a major problem as control over the optimisation process is relatively restricted. A lot of hand tuning is often required and when an optimisation goes wrong it is very difficult to debug.
- C.P.U. time is prohibitive especially when compared to a more sequential approach (hours as opposed to minutes).
- the flexibility and selectivity is quite low. For example, if one particular aspect of device behaviour requires very accurate modelling, the sequential approach can apply the optimisation of the relevant parameters directly to that region with relative ease. Because a simultaneous optimisation cannot be this selective, parameters are optimised not only in the region where they exhibit the highest sensitivity but also in other regions. This often leads to unphysical parameter values which in turn cause unphysical correlation with other parameters.

In short, a more sequential approach in parameter optimisation leads to greater automation, insight and control of the optimisation process.

In order to develop a parameter determination strategy using the sequential approach, Modella was implemented into the parameter determination program Bipar[43]. This program uses a bounded optimisation approach, based on the least squares method, to minimise the error between measurements and model predictions. The minimisation problem is solved as a function of the model parameter(s) specified by the user. Implementation of Modella into Bipar involved coding the compact model definition (as defined in appendix A) and also writing Newton-Raphson iteration procedures to solve the internal network of the equivalent circuit. This solution process involved solving for the internal junction voltages which are consistent with two independent terminal settings provided by the user. These independent settings refer to the terminal currents and/or voltages which define the operating point of the transistor. In bipolar transistors *two* independent terminal settings are required for this purpose as the branch currents are a function of the internal emitter-base and collector-base voltages. Therefore, if say Ic and Vcb are supplied by the user as the two independent terminal settings, the internal

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

junction voltages can be solved, thereby facilitating the calculation of all the other terminal settings like Ib, Is and Veb.

Optimisation of a parameter like the ideal current gain $\mathbf{B_f}$ therefore involves specifying the following:

- an initial guess for $\mathbf{B_f}$.
- maximum and minimum bounds allowed for $\mathbf{B_f}$.
- two independent settings which define each operating point.
- the terminal setting to be fitted.
- the range of measured data in which the optimisation is to take place.

In this case Ic and Vcb are the two independent settings and Ib is fitted. A range of values for the first independent setting, Ic, defines the region in which Ib is fitted. By choosing Ic as the first independent setting and not Veb, one is fitting Ib *relative* to Ic rather than the absolute value of Ib. This has a desired effect in that the *current gain* characteristic, h_{fe}, is fitted[1].

The optimisation process in Bipar requires information on the sensitivity of Ib to $\mathbf{B_f}$ in order to find the minimum in the error between measurements and model predictions. This sensitivity is given by the numerical approximation of $\frac{\partial Ib}{\partial \mathbf{B_f}}$ in the model routine. Bipar then returns a value of $\mathbf{B_f}$ which minimises the difference between the measured h_{fe} and that predicted by Modella in the region specified.

This introduction to parameter determination using Bipar should provide the necessary information to understand the use of Modella within Bipar in developing the parameter determination strategy. Appendix B contains a more detailed description of both the optimisation theory used in Bipar and the procedure required to solve the equivalent circuit of Modella.

[1]Because Ic will already have been fitted in this region, it follows that a correct prediction of h_{fe} results in a correct prediction of Ib. The difference is that here the emphasis is placed on the accuracy of the current gain prediction rather than the accuracy of the absolute value of Ib. This is because in the iterative solution of the equivalent circuit, Modella's prediction of the two independent settings are matched, within a small error margin, to the measured values. Modella's Ic will therefore be very accurate and the optimisation of $\mathbf{B_f}$ should ensure a good fit to Ib thereby providing a good current gain prediction.

© N.V. Philips' Gloeilampenfabrieken

4.3 Parameter determination strategy

An ideal parameter determination strategy should be straightforward and robust, yielding physically plausible and unique parameter sets. Although the strategy described here is not ideal in this sense, due to the possible non-uniqueness of parameter sets, it does represent a highly tuned solution to a rather difficult problem. The determination of each model parameter and the reasoning behind the approaches adopted is described in detail below. In this description the emphasis is given to those parameters determined using non-standard approaches i.e. those approaches which are not to be found in textbook descriptions of bipolar parameter determination (see e.g. [15]) but rather are specific to the determination of Modella's parameters.

This description is divided into D.C. model parameter determination and charge model parameter determination (the charge model parameters determining model behaviour in A.C. [small signal] and transient [large signal] analyses). The D.C. optimisation sequence described below is executed a number of times to ensure that physical correlation between parameters is taken into account. Experience has shown that *three*[2] optimisation loops usually result is a stable set of parameters.

4.3.1 D.C. model parameters

The low current block

$\mathbf{E_{afl}}$:
The Early voltage of the lateral forward current component at zero collector-base bias is optimised using the non-saturation region of the $Ic - Vec$ characteristic. Because $\mathbf{E_{afl}}$ characterises the Early effect in the lateral component of Ic which dominates at low current levels, it is crucial that the value of Ib chosen with the measured $Ic - Vec$ characteristic is low enough to ensure that current crowding under the emitter is not present. Ideally, a value of Ib corresponding to Veb less than about $0.65V$ is required.

[2] This number applies to the present implementation of Modella which uses a constant diffusion voltage Vd. Although not a model parameter, Vd can be derived in theory from the model parameters $\mathbf{I_k}$ and $\mathbf{I_s}$ (see equation (3.12)). Expressing Vd as a derived quantity, however, introduces an unacceptably high degree of correlation between high and low current parameters. Because Vd is required in the description of the Early effect *before* the parameters $\mathbf{I_k}$ and $\mathbf{I_s}$ have been optimised, using this relation introduced a strong coupling between the Early effect and the parameters $\mathbf{I_k}$ and $\mathbf{I_s}$. Motivation for using a constant value for Vd came also from the fact that the accuracy in modelling the Early effect in lateral pnps with different epitaxial doping concentrations was found to be rather insensitive to values of Vd in the range of 0.5V to 0.7V.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Because $\mathbf{E_{afl}}$ (and also $\mathbf{E_{afv}}$ below) only determines the slope of Ic in this region, an auxiliary parameter $\mathbf{B_f}$ is required to fit the absolute value of Ic. $\mathbf{B_f}$ is optimised in its own right later on. The terminal setting to be optimised is of course Ic and the two independent settings defining the operating conditions are Ib and Vcb. Ib is chosen instead of Veb because the use of $\mathbf{B_f}$ as auxiliary parameter to fit the absolute value of Ic requires that Ib is first matched to the measured values. If $\mathbf{I_s}$ was chosen as auxiliary parameter then Veb would have been a better choice than Ib to use as an independent setting.

The standard optimisation procedure for Early parameters using the Early measurement (Ic as a function of reverse Vcb at constant forward Veb) could in principle be used to determine $\mathbf{E_{afl}}$ since there is no difference at low current levels in the definition of the Early voltage V_{eaf} using Veb constant or Ib constant. However, for compatibility with the definition of the Early voltage used in the determination of $\mathbf{E_{afv}}$ the standard Early measurement is not used here.

$\mathbf{E_{afv}}$:
The Early voltage of the 'vertical' (i.e. bottom) forward current component at zero collector-base bias is also determined using the $Ic - Vec$ characteristic in a similar optimisation range but this time with a large value of Ib corresponding to $Veb > 0.9V$. Because $\mathbf{E_{afv}}$ describes the Early effect in the bottom component of Ic, which dominates under current crowding conditions, Ib should be large enough to ensure that Ic is mainly composed of this bottom component. Measurements of the $Ic - Vec$ characteristics are therefore done at two values of Ib: one at low forward bias which is sensitive to the lateral component of Ic and one at high forward bias which, under current crowding conditions, is sensitive to the bottom component of Ic, thereby facilitating the determination of $\mathbf{E_{afl}}$ and $\mathbf{E_{afv}}$ respectively.

The reasons for not choosing the standard Early measurements with constant Veb are as follows:

- The measurement of the Early effect is directly related to the definition of the Early voltage i.e. whether it is defined at constant Ib or constant Veb. At high current levels the definition at constant Veb yields higher values of V_{eaf} due to ohmic voltage drop across $\mathbf{R_{eex}}$. The following sequence of events leads to this overestimation of V_{eaf}: applying a greater reverse bias to the collector-base junction leads through the normal Early effect to an in-

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

crease in Ic. This increase in current causes more ohmic voltage drop across $\mathbf{R_{eex}}$ which in turn reduces the internal emitter-base junction voltage i.e. $Ve2b(= Ie2e1\ \mathbf{R_{ein}}\ +\ Ve1b)$ (see figure 3.2). Reducing the internal junction voltage means that Ic will decrease, thereby compensating the increase due to the normal Early effect. The result is an underestimation of the Early effect leading to larger values of V_{eaf}. Optimising $\mathbf{E_{afv}}$ with constant Veb is therefore unnecessarily dependent on having a correct value for $\mathbf{R_{eex}}$. Furthermore, this is inherently a resistance effect and should not be correlated to the Early effect.

- Incorrect interpretation of the Early effect as a result of the presence of internal heating caused by dissipation is much more predominant with constant Veb than constant Ib. Although internal heating is a strong function of geometry and junction depth, a rough estimate is say $0.5 - 1°C$ per milliWatt. Approximating the dissipation as $Ic.Vcb$ and taking for example a 1mA current at -5V Vcb, a significant 2.5 - 5 degree rise in device temperature results. When Veb is constant and more reverse bias is applied with Vcb, the internal heating will cause Ic to increase and this will result in an overestimation of the Early effect. This overestimation is proportional to the temperature dependence of the saturation current parameter $\mathbf{I_s}$ of the collector current. If Ib is constant, however, the internal emitter-base voltage ($Ve2b1$) will decrease when internal heating occurs in order to maintain a constant Ib value. The Early effect will be affected in this case therefore by the temperature dependence of the current gain parameter $\mathbf{B_f}$ which, relative to the temperature dependence of $\mathbf{I_s}$, is quite small.

Indeed, attempts to optimise $\mathbf{E_{afv}}$ using $Ic - Vcb$ characteristics with $Veb = 0.95V$ were very unsatisfactory yielding at best physically implausible values. The definition of V_{eaf} with constant Ib was therefore concluded to be much more appropriate for parameter determination purposes.

$\mathbf{E_{arl}}$ and $\mathbf{E_{arv}}$:

The procedure to optimise these two reverse Early parameters is identical to that used in the forward case for $\mathbf{E_{afl}}$ and $\mathbf{E_{afv}}$. Here the $Ie - Vce$ characteristics are used with $\mathbf{B_r}$ as auxiliary parameter to fit the absolute value of Ie. Veb and Ib are the two independent settings and of course Ie is the terminal setting to be fitted by optimising $\mathbf{E_{arl}}$ and $\mathbf{E_{arv}}$ in the low and high base current regions respectively.

© N.V. Philips' Gloeilampenfabrieken

$\mathbf{I_s}$:

The optimisation of the saturation current is very straightforward. Veb and Vcb are taken as independent settings with Ic as the setting to be fitted. The forward Gummel measurement is used and the low range of Veb values in which the optimisation takes place reflects the fact that the physical meaning of $\mathbf{I_s}$ is the extrapolated 0 volt intercept of Ic with the current axis in the Gummel plot. It should be noted that the optimisation of the Early effect is placed *before* the optimisation of $\mathbf{I_s}$ because the prefactor of the exponential term $(\exp(Veb/Vt))$ in Ic includes the Early term. This prefactor must therefore contain the optimised Early parameters before $\mathbf{I_s}$ can be correctly extracted.

$\mathbf{X_{es}}$:

This parameter effectively represents the saturation current of the forward substrate current and is therefore optimised in a manner similar to $\mathbf{I_s}$. Instead of fitting Ic, the forward substrate current Is (not to be confused with the bold face model parameter $\mathbf{I_s}$!) is fitted and instead of having Veb and Vcb as the two independent settings, Ic has been chosen instead of Veb. By using Ic as an independent setting and Is as the setting to be fitted, one is optimising Is *relative* to Ic and not optimising the absolute value of Is itself. This therefore optimises the gain characteristic Is/Ic in the low current region. Because the *absolute* value of Ic has already been fitted by optimising $\mathbf{I_s}$ in the low current region, it follows that the absolute value of Is will also be optimised here.

$\mathbf{X_{cs}}$:

In a similar manner to the optimisation of $\mathbf{X_{es}}$, the saturation current of the reverse substrate current is optimised by fitting the reverse Is and using Ie and Veb as the independent settings. Again the optimisation takes place in the low current range before any high injection or ohmic effects affect the device characteristics.

$\mathbf{B_f}$, $\mathbf{I_{bf}}$ and $\mathbf{V_{lf}}$:

The ideal forward current gain and the saturation current and crossover voltage of the non-ideal forward base current are the parameters which determine the low and mid current behaviour of the forward current gain characteristic. This fact is apparent in the range of Ic values chosen for optimisation in the example in section **4.3.3**. Agaiu the choice of Ic and Vcb define the

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

operating condition. Using Ib as the setting to be fitted in conjunction with Ic as one of the two independent settings means that emphasis is placed on optimising the current gain characteristic.

Optimising $\mathbf{B_f}$ independently from $\mathbf{I_{bf}}$ and $\mathbf{V_{lf}}$ showed virtually no difference in parameter values compared to optimising all three together. It should be noted that the optimisation ranges chosen can have quite a noticeable effect on the parameters and therefore on the model's predictive capabilities. For example if these parameters are optimised in a range of Ic which excludes the mid current region, the predicted maximum of the forward current gain characteristic may fall well below the measured value. On the other hand, optimisation ranges can sometimes be deliberately shortened to ensure a good prediction in that region at the expense of another region. This, however, only tends to be required with the more basic compact models and often may lead to unphysical parameter values. For this reason the optimisation ranges in this strategy are not unduly specific.

$\mathbf{B_r}$, $\mathbf{I_{br}}$ and $\mathbf{V_{lr}}$:
The ideal reverse current gain and the saturation current and crossover voltage of the non-ideal reverse base current are determined in a similar manner to the forward parameters. Independent settings are Veb and Ie; with Ib the setting to be fitted. The emphasis again lies on the accurate prediction of the reverse current gain characteristic, h_{fc}.

The high current block

$\mathbf{I_k}$:
The knee current parameter, which characterises all high injection effects in the lateral pnp, is optimised in the fall off region of the forward current gain characteristic by choosing Ic as the setting to be fitted and Ib and Vcb as the two independent settings. Because high injection begins to affect the current gain characteristic one to two decades before fall off, it is important to include this in the region of optimisation. In addition, current crowding which also contributes to the fall off of h_{fe} has a negligible influence when Veb is decreased below about $0.7V$, thereby leading to a less ambiguous determination of $\mathbf{I_k}$ in this region. This approach to $\mathbf{I_k}$ determination is however not the only one and the discussion in section **4.4** outlines other approaches and criteria for choosing the best approach.

© N.V. Philips' Gloeilampenfabrieken

$\mathbf{R_{ein}}$ and $\mathbf{X_{ifv}}$:

The internal emitter resistance and the vertical fraction of the forward current are the two parameters which characterise the amount of current crowding which takes place in the device. They are optimised in the high current region of the forward current gain characteristic by choosing Ib and Vcb as independent settings and Ic as the setting to be fitted. Assuming that the value for $\mathbf{I_k}$ has a correct physical basis, the remaining fall off of the gain characteristic has its physical basis in current crowding and can thus be described by these two parameters. In general, it has been observed that the falloff in the gain characteristic is more sensitive to $\mathbf{I_k}$ than to the current crowding but that both effects *must* be optimised in order to obtain an accurate prediction. Optimising the three parameters $\mathbf{I_k}, \mathbf{X_{ifv}}$ and $\mathbf{R_{ein}}$ together tended to to yield unphysically small values for $\mathbf{I_k}$ due to the higher sensitivity of the gain characteristic to high injection. On the other hand, when $\mathbf{I_k}$ is optimised first, a greater requirement exists for a good initial guess for $\mathbf{I_k}$. Because it is crucial to get the correct balance between the relative contributions of current crowding and high injection effects, the discussion in section **4.4** describes in more detail ways in which Modella's *physical basis* can be helpful in distinguishing the quantitative contribution from these two effects.

$\mathbf{R_{cin}}$ and $\mathbf{X_{irv}}$:

The internal collector resistance and the vertical fraction of the reverse current are optimised in the same manner as $\mathbf{R_{ein}}$ and $\mathbf{X_{ifv}}$. The high current region of the reverse current gain characteristic, h_{fc}, is used by choosing Ib and Veb as the independent settings and Ie as the setting to be fitted. The optimisation range in Ib corresponds to values of Vcb from about $0.7V$ to $1.0V$.

$\mathbf{X_{hes}}$:

The fraction of Is in the forward active mode that is subject to high injection effects is optimised by fitting Is; with Ib and Vcb defining the operating point. The region of optimisation is of course taken in the region where high injection effects, characterised by $\mathbf{I_k}$, occur.

$\mathbf{X_{hcs}}$:

The fraction of the reverse substrate current subject to high injection is optimised in a similar manner to $\mathbf{X_{hes}}$ by fitting Is; with Ib and Veb as independent settings. With junction isolation $\mathbf{X_{hcs}}$ will normally be greater than $\mathbf{X_{hes}}$, due to the proximity of the isolation, leading to a

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

greater fraction of Is passing through the epilayer directly into the isolation. This component of Is will in turn be subject to high injection effects due to the low epilayer doping and thus $\mathbf{X_{hes}}$ will increase. With some geometries $\mathbf{X_{hes}}$ will have a value of one, signifying that all of Is is composed of this component. Section 4.4 discusses the more extreme situation where a value of one is not large enough to describe the fall off in the gain characteristic Is/Ib ! With closed buried layers in oxide isolated devices, $\mathbf{X_{hcs}}$ will generally be similar to $\mathbf{X_{hes}}$ because only second order high injection effects (see section **2.3.2**) are responsible for the limited fall off in Is/Ib. With open buried layers, the situation corresponds more to the junction isolated device since in the reverse mode a greater portion of Is originates from the component which reaches the substrate without passing through the intervening buried layer.

$\mathbf{R_{bev}}$:

The variable part of the base resistance Rbe at low injection is normally calculated from a knowledge of the epilayer resistivity and geometry considerations (see section **4.3.3** for a worked example).

$\mathbf{R_{bcv}}$:

The variable part of the base resistance Rbc at low injection is also calculated (as opposed to optimised) and is generally smaller than $\mathbf{R_{bev}}$ due to the larger collector area (see section **4.3.3**).

$\mathbf{R_{eex}}$ and $\mathbf{R_{bec}}$:

The external emitter resistance and the constant part of the base resistance Rbe are determined using the method of Ning and Tang[44]. Basically this approach measures the voltage deviation δV of Ib from the ideal $Ib - Veb$ characteristic (see figure 4.1). Ib is used because, in the region of interest, δV is a function solely of the ohmic voltage drop across the base and external emitter resistances and not of high injection or current crowding as is the case with Ic. The voltage deviation can be expressed as follows.

$$\delta V = IeRe + IbRbe \tag{4.1}$$

where $Ie = Ic + Ib + Is$

$Re = \mathbf{R_{eex}}$ since Ib depends on $Ve2b1$ and is therefore unaffected by $\mathbf{R_{ein}}$

$Rbe = \mathbf{R_{bec}} + Rbv$

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Here, Rbv is the bias dependent part of Rbe (see equation (3.21)) which at low injection is equal to $\mathbf{R_{bev}}$ but at high injection decreases rapidly due to conductivity modulation of the epilayer resistivity.

Rewriting equation (4.1) yields...

$$\frac{\delta V}{Ic} = \mathbf{R_{eex}}\left(1 + \frac{Is}{Ic}\right) + \frac{Rbv}{h_{fe}} + \frac{\mathbf{R_{eex}} + \mathbf{R_{bec}}}{h_{fe}} \tag{4.2}$$

Normally $\frac{Is}{Ic} << 1$ and therefore it may be neglected.

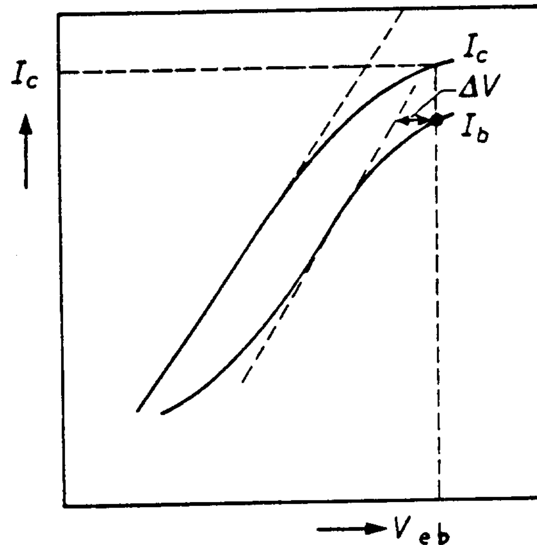


Figure 4.1: *Voltage deviation of Ib from the ideal.*

If we make the assumption that $\frac{Rbv}{h_{fe}}$ is constant at all injection levels, then a plot of $\frac{\delta V}{Ic}$ versus h_{fe}^{-1} yields a curve which has as its slope $\mathbf{R_{eex}} + \mathbf{R_{bec}}$ and as its extrapolated y-axis intercept $\mathbf{R_{eex}} + \frac{Rbv}{h_{fe}}$ (see figure 4.2). Because we have already calculated Rbv at low injection (i.e. $\mathbf{R_{bev}}$), we can easily determine $\mathbf{R_{eex}}$ and $\mathbf{R_{bec}}$ by using this curve.

The accuracy of this method to determine $\mathbf{R_{eex}}$ and $\mathbf{R_{bec}}$ depends on the accuracy with which $\mathbf{R_{bev}}$ had been calculated. In addition, the correct interpretation of the curve of figure 4.2 is dependent on the validity of the assumption that $\frac{Rbv}{h_{fe}}$ is constant. This assumption is rarely valid in npn transistors for the following reasons.

- current crowding at the emitter-base sidewall is usually present. This leads to a dramatic

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

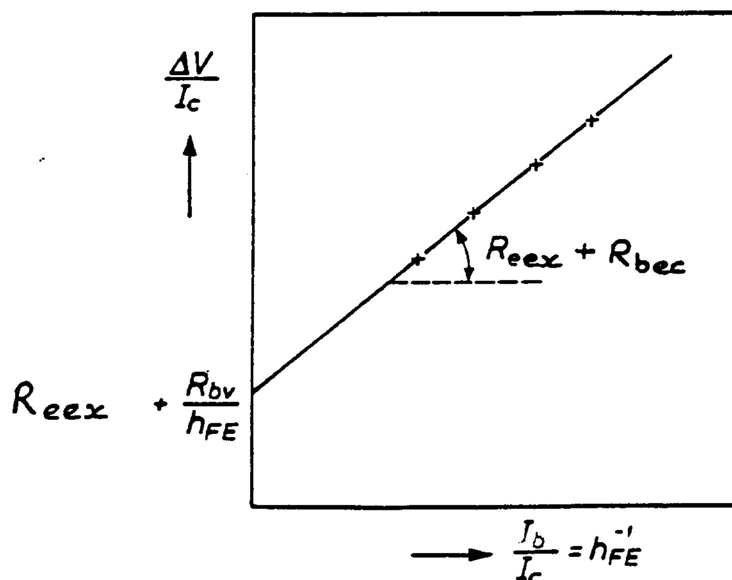


Figure 4.2: *Interpretation of the curve showing the voltage deviation of Ib divided by Ic as a function of the reciprocal of* h_{fe}.

reduction in *Rbv* but has relatively little effect on h_{fe}.

- the influence of quasi-saturation on h_{fe} increases the ratio.
- the current dependence of the minority carrier mobility in the base has been neglected. At high current levels the mobility and therefore the diffusion constant are seen to decrease by as much as a factor of two depending on the base doping profile. Again the result is a reduction in h_{fe}.

Because of these effects and the difficulty in quantifying them, the interpretation of the curve is often rather dubious in the case of the npn transistor.

With the lateral pnp, however, current crowding at the emitter-base sidewall does not occur and quasi-saturation may be neglected under normal operating regions. Neglecting for a moment current crowding under the emitter and the current dependence of the minority carrier mobility in the base, it can easily be shown (using equations 3.21, 3.5 and 2.28) that in the lateral pnp we have:

$$\frac{Rbv}{h_{fe}} \propto \frac{n_e}{p_0\,(1\,+\,p_0)} \tag{4.3}$$

which is constant at both low and high injection levels since only p_0 exhibits high injection effects. The current dependency of the hole diffusion coefficient has been derived in section **2.3.1**

© N.V. Philips' Gloeilampenfabrieken

and leads to a factor of two increase in Ic (and therefore h_{fe}) at high injection levels. This would lead to a factor of two reduction in $\frac{Rbv}{h_{fe}}$ if the effect of emitter current crowding was not present. A reduction in h_{fe} due to current crowding by as much as a factor of two is often observed and this has a desired effect in that it maintains an approximately constant ratio between Rbv and h_{fe}.

The conclusion is that the method of Ning and Tang is highly suitable for the lateral pnp, the only prerequisite being that the condition $\frac{Is}{Ic} << 1$ be satisfied. A detailed description of the procedure to extract $\mathbf{R_{eex}}$ and $\mathbf{R_{bec}}$ can be found in the example given in section **4.3.3**.

$\mathbf{R_{cex}}$ and $\mathbf{R_{bcc}}$:

In the case where $\frac{Is}{Ic} << 1$, the method of Ning and Tang can also be used here to calculate the external collector resistance and the constant part of Rbc. However, with open buried layers and junction isolated devices this condition is rarely true. An alternative method is one whereby the the absolute value of the largest current in the device (Ic in reverse active mode) is fitted by optimising $\mathbf{R_{cex}}$ and $\mathbf{R_{bcc}}$ together. Because $Ic = Ie + Is + Ib$ in the reverse active mode, fitting Ic will automatically take into account the largest currents which cause the greatest ohmic voltage drop, thereby maximising the sensitivity of optimisation to $\mathbf{R_{cex}}$ and $\mathbf{R_{bcc}}$. The independent settings are Veb and Vcb with the range of optimisation given by $0.7V < Vcb < 1.0V$. In principle a similar approach could be used to optimise $\mathbf{R_{eex}}$ and $\mathbf{R_{bec}}$ instead of using the method of Ning and Tang. However, the method described here suffers from the drawback that a substantial amount of measurement induced correlation has been introduced. The approach requires correct physical values for parameters like $\mathbf{I_k}$, $\mathbf{R_{ein}}$ and $\mathbf{R_{cin}}$ so that all other causes of deviation from the ideal $I-V$ behaviour are accurately modelled. The only source of deviation remaining is then due to the resistances in question here. Only then will correct optimisation occur. Inaccurate results will be obtained for example in the case where the model fit to the current gain curves (h_{fe} and h_{fc}) is not accurate enough. Comparisons with results from the method of Ning and Tang have shown, however, that this method does yield excellent results when the other causes of high current non-ideality are correctly modelled.

$\mathbf{I_{ss}}$:

The saturation current of the substrate-base diode can only be optimised when data is available

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

for device behaviour under *zero base current crossover* conditions. Normally this zero base current crossover, which leads to increased h_{fe} values in the low current region, is only observed under high temperature conditions. When these measurements are available, $\mathbf{I_{ss}}$ can be optimised by using the forward current gain in the low current region by fitting Ib with Ic and Vcb as independent settings. In cases where this behaviour is observed at high temperatures but not at room temperatures, the scaling of $\mathbf{I_{ss}}$ with temperature (discussed in chapter 5) should ensure that Modella's prediction of this effect reflects its reduced importance as a function of decreasing temperature.

$\mathbf{R_{sb}}$:
The substrate-base leakage resistance will not be included in subsequent versions of Modella as the requirement no longer exists to model what was basically a technological problem rather than a physical effect.

4.3.2 Charge model parameters

$\mathbf{C_{jx}}, \mathbf{V_{dx}}$ and $\mathbf{P_x}$:
Here x can be e, c, s representing the emitter-base, collector-base or substrate-base junctions. These nine parameters model the depletion capacitance of each junction and they are determined simply by fitting the appropriate depletion capacitance .vs. voltage data by optimising all three parameters together. The range of optimisation is from about $5V$ reverse junction bias to $0.5V$ forward junction bias.

$\mathbf{T_{lat}}, \mathbf{T_{fvr}}$ and $\mathbf{T_{fn}}$:
These low injection transit time parameters represent the regional charge storage in the forward mode of operation. $\mathbf{T_{lat}}$ accounts for charge stored in the epilayer between emitter and collector, $\mathbf{T_{fvr}}$ for charge stored in the epilayer under the emitter and $\mathbf{T_{fn}}$ for charge storage in the neutral emitter and the buried layer under the emitter. In order to determine these parameters, use is made of the figure of merit commonly used to characterise the high frequency behaviour of bipolar transistors - the common-emitter cut-off frequency f_t. This is defined as the frequency at which the small signal current gain, $\hat{h}_{fe}$, has dropped to a value of one and it is measured by extrapolating the 6dB/octave fall off in $\hat{h}_{fe}$ to the 0dB point. The f_t is inversely proportional to the total transit time in a device and is given by $f_t = \frac{1}{2\pi\tau_{tot}}$, where the total transit time

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

τ_{tot} is simply the sum of the individual transit times. In the quasi-static approach, the transit time can be determined from the steady state solution by taking the derivative of the charge with respect to the current. For example, the actual transit time associated with charge storage in the epilayer under the emitter, represented by low injection parameter $\mathbf{T_{fvr}}$, is determined according to $\tau(Q_{fver}) = \frac{dQ_{fver}}{dIc}|_{dVec=0}$. This is the approach used in Bipar and it allows an approximation of an A.C. figure of merit to be made by what is essentially a steady state D.C. analysis.

Figure 4.3 shows an example of the how the transit/charging times, associated with all the individual charge elements in Modella, vary as a function of the forward injection level. The charge model parameters determine the relative low current levels of the transit times, while at high current levels other effects such as high injection and current crowding alter the relative contribution to the total transit time from these individual components. The parameters $\mathbf{T_{lat}}, \mathbf{T_{fvr}}$ and $\mathbf{T_{fn}}$ are therefore determined by providing a prediction of τ_{tot} such that the bias dependence of f_t is correctly modelled. The procedure is to determine the transit times associated with each of the charges, sum these transit times to predict τ_{tot} (and therefore f_t) and then compare with the measured values of f_t. In these measurements Ic is varied and f_t is measured using Vcb as parameter (i.e. the forward f_t measurement). Normally the f_t .vs. Ic characteristic is measured at three values of Vcb but only one is used for parameter determination purposes. However, it has been found that the best results for $\mathbf{T_{lat}}, \mathbf{T_{fvr}}$ and $\mathbf{T_{fn}}$ were obtained when the parameters were all optimised together; by fitting all three measured characteristics simultaneously. In this way it is possible to avail of the dependence of the individual transit times on Vcb in order to help separate the measurement induced correlation between these three parameters.

One point that has emerged in the determination of these three charge model parameters is just how dependent the f_t predictions are on the values of certain D.C. parameters (e.g. $\mathbf{E_{afl}}, \mathbf{I_k}$ and $\mathbf{R_{eln}}$ (with $\mathbf{X_{lfv}}$)). For example, the contribution that current crowding makes to the fall off of h_{fe} relative to that of high injection, determines whether $\tau(Q_{fver})$ or $\tau(Q_{fn})$ dominates the f_t fall off at high currents. A physical parameter set for the D.C. case is therefore a prerequisite to the charge model being physically valid.

© N.V. Philips' Gloeilampenfabrieken

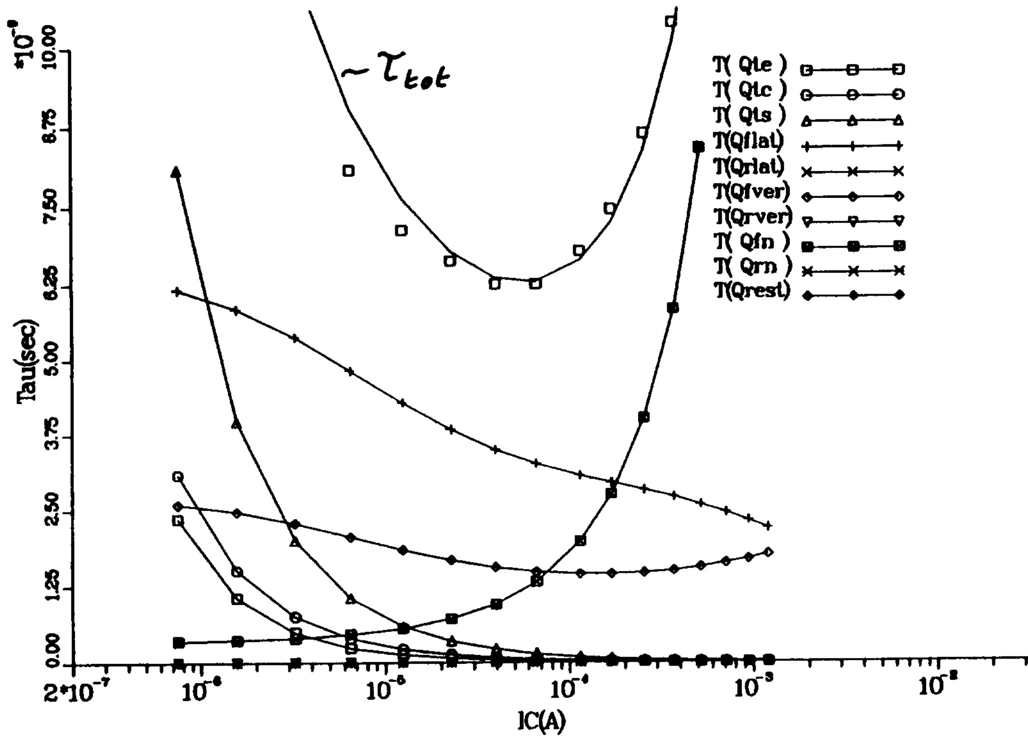


Figure 4.3: *An example of the bias dependence of Modella's transit/charging times.*

T_{rvr} and T_{rn} :

Because of symmetry in the charge storage in the epilayer between the emitter and collector in the reverse mode, T_{lat} is also used to represent this reverse charge storage. It should be noted, however, that due to charge asymmetry in the case of U-shaped and cylindrical devices, this is an approximation for these layouts. The low injection transit time parameters, T_{rvr} and T_{rn}, represent reverse mode charge storage in the epilayer under the collector and charge storage in the neutral collector and buried layer under the collector respectively. When reverse f_t measurements are available they can be optimised in a similar manner to the forward mode. However, when these measurements are not available an estimate for these parameters can be made. This estimate is based on the fact that the charge storage in both these regions scales with the collector bottom area. Because the collector-base depletion capacitance can be assumed to originate almost entirely from the collector bottom area and because similar assumptions can be made for the emitter in the forward mode, we can estimate that $T_{rvr} = T_{fvr} \frac{C_{jc}}{C_{je}}$ and $T_{rn} = T_{fn} \frac{C_{jc}}{C_{je}}$.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

4.3.3 Example

In order to illustrate the above methodology in practice, this section describes the determination of the parameter set for a junction-isolated lateral pnp transistor. This device has a square layout whereby the collector completely surrounds the $8 \times 8\mu m^2$ emitter region. A sketch of the layout and cross-section (in which lateral out diffusions are not shown) are shown in figure 4.4. The base width is given by the mask base width ($4\mu m$) minus the lateral out diffusions from emitter and collector ($0.7 \times Ye \approx 0.5\mu m$ each) yielding $Wb = 3\mu m$. Doping concentrations in the emitter/collector and epitaxial base regions are in the order of 2.0×10^{19} and 3.5×10^{15} respectively.

Table 4.1 shows the specific optimisation methodology used for each of the D.C. model parameters. This includes the measurement data used, the two independent settings chosen during the optimisation, the range of values of the first independent setting in which the optimisation takes place and of course the terminal setting which is fitted by optimising the parameter in question.

The other D.C. parameters are calculated as follows.

$\mathbf{R_{bev}}$ and $\mathbf{R_{bcv}}$: Given that the epilayer resistivity, ρ, is $1.5\Omega cm$, we can write $\mathbf{R_{bev}} = \rho \frac{l}{A}$, where l is the thickness of the epilayer under the emitter and A is the effective injection area of the electrons into the emitter region. From section **2.3.3**, $A = X''^{\,2}_{cone}$ where

$$X''_{cone} = X_{cone} + \frac{2ln2}{\pi} Ye = 4.31 \times 10^{-4}\, cm$$

Therefore $\mathbf{R_{bev}} = 565\Omega$. Using a similar calculation we obtain $\mathbf{R_{bcv}} = 195\Omega$.

$\mathbf{R_{eex}}$ and $\mathbf{R_{bec}}$: Using the method of Ning and Tang we obtain the plot shown in figure 4.5. It can easily be shown that the voltage deviation in Ib can be determined by $\delta V = \frac{kT}{q} ln(\frac{Ib0}{Ib})$ where $Ib0$ is the ideal base current free from ohmic effects. To determine $Ib0$, the base current at the maximum in the h_{fe} characteristic is taken to be ideal and this yields the ideal saturation current, $Ibf0$, to be used in $Ib0 = Ibf0 \exp \frac{Veb}{Vt}$. The resulting plot yields an extrapolated

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

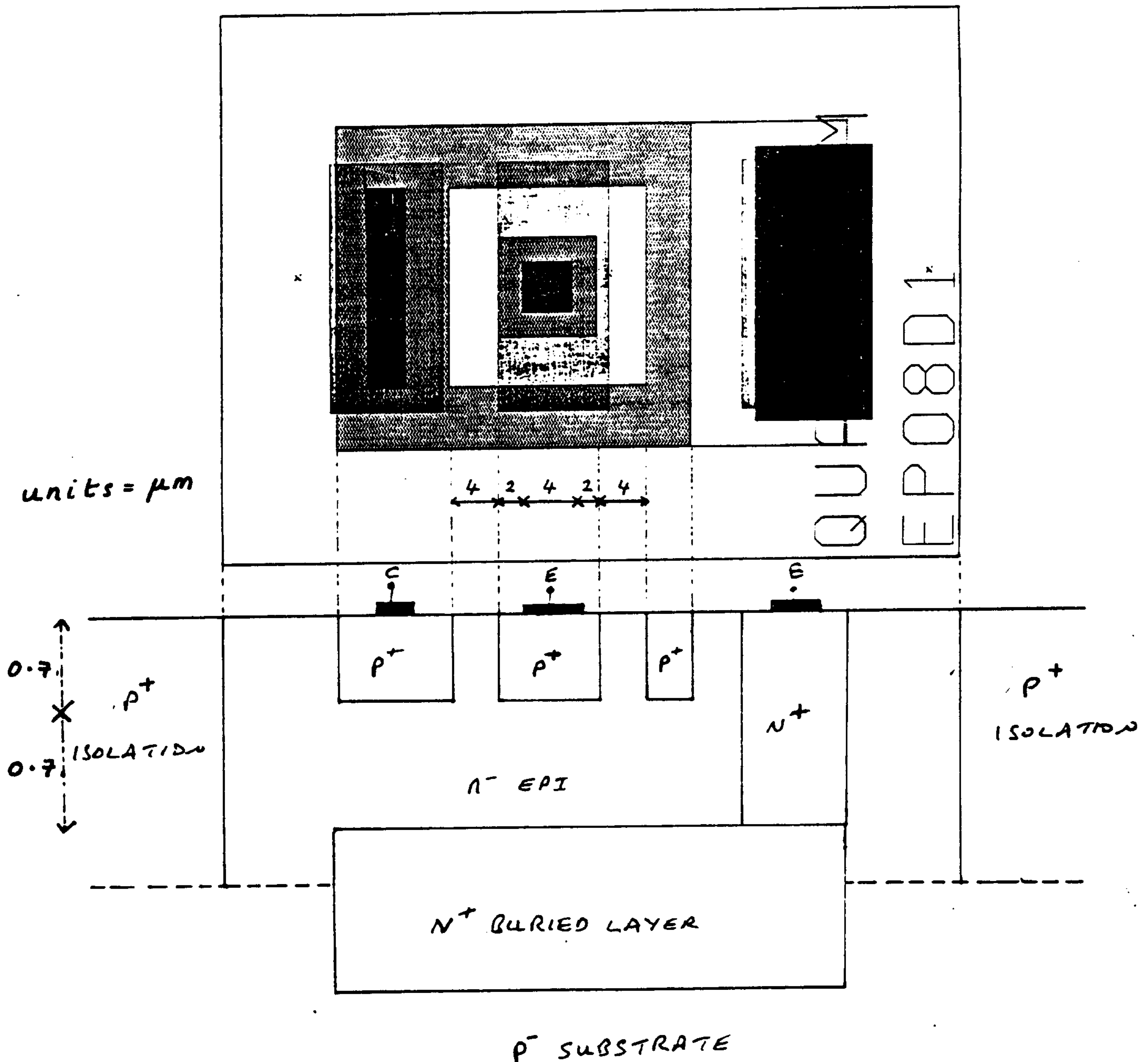


Figure 4.4: *Schematic layout and cross-section of the junction-isolated device used for parameter determination.*

y-axis intercept of 18.5 with a slope of 18.76. This gives the following two equations

$$18.5 = \mathbf{R_{eex}} + \frac{\mathbf{R_{bev}}}{h_{fe}}$$

$$18.76 = \mathbf{R_{eex}} + \mathbf{R_{bec}}$$

which are solved (using the maximum value of h_{fe}) to give $\mathbf{R_{eex}} = 12.1\Omega$ and $\mathbf{R_{bec}} = 6.66\Omega$.

© N.V. Philips' Gloeilampenfabrieken

Parameter(s)	Measurement data	Optimisation range	Settings		Fit
$\mathbf{E_{afl}}$	$Ic - Vec$ characteristic ($Ib = 1.0 \times 10^{-7}$)	$1.0 \rightarrow 10.0V$	Vec	Ib	Ic
$\mathbf{E_{afv}}$	$Ic - Vec$ characteristic ($Ib = 4.5 \times 10^{-4}$)	$1.0 \rightarrow 10.0V$	Vec	Ib	Ic
$\mathbf{E_{arl}}$	$Ie - Vce$ characteristic ($Ib = 1.0 \times 10^{-7}$)	$-10.0 \rightarrow -1.0V$	Vec	Ib	Ie
$\mathbf{E_{arv}}$	$Ie - Vce$ characteristic ($Ib = 4.5 \times 10^{-4}$)	$-10.0 \rightarrow -1.0V$	Vec	Ib	Ie
$\mathbf{I_s}$	forward Gummel plot	$0.4 \rightarrow 0.55V$	Veb	Vcb	Ic
$\mathbf{X_{es}}$	forward Gummel plot	$-2.0 \times 10^{-6} \rightarrow -1.0 \times 10^{-7}A$	Ic	Vcb	Is
$\mathbf{X_{cs}}$	reverse Gummel plot	$-5.0 \times 10^{-7} \rightarrow -1.0 \times 10^{-10}A$	Ie	Veb	Is
$\mathbf{B_f, I_{bf}, V_{lf}}$	forward Gummel plot	$-7.0 \times 10^{-7} \rightarrow -2.0 \times 10^{-9}A$	Ic	Vcb	Ib
$\mathbf{B_r, I_{br}, V_{lr}}$	reverse Gummel plot	$-6.0 \times 10^{-7} \rightarrow -1.0 \times 10^{-10}A$	Ie	Veb	Ib
$\mathbf{I_k}$	forward Gummel plot	$-8.0 \times 10^{-4} \rightarrow -6.0 \times 10^{-9}A$	Ib	Vcb	Ic
$\mathbf{R_{eln}, X_{ifv}}$	forward Gummel plot	$-1.0 \times 10^{-3} \rightarrow -1.0 \times 10^{-8}A$	Ib	Vcb	Ic
$\mathbf{R_{cin}, X_{irv}}$	reverse Gummel plot	$-5.0 \times 10^{-3} \rightarrow -2.0 \times 10^{-7}A$	Ib	Veb	Ie
$\mathbf{X_{hes}}$	forward Gummel plot	$-1.0 \times 10^{-3} \rightarrow -5.0 \times 10^{-9}A$	Ib	Vcb	Is
$\mathbf{X_{hcs}}$	reverse Gummel plot	$-5.0 \times 10^{-3} \rightarrow -1.0 \times 10^{-8}A$	Ib	Veb	Is
$\mathbf{R_{cex}}$	reverse Gummel plot	$0.7 \rightarrow 1.0V$	Vcb	Veb	Ie

Table 4.1: *The D.C. parameter determination strategy for the V-80 junction-isolated lateral pnp.*

$\mathbf{R_{bcc}}$: In this example the reverse Is is not negligible relative to Ie meaning that the method of Ning and Tang cannot be used. Therefore, $\mathbf{R_{bcc}}$ and $\mathbf{R_{cex}}$ are optimised using Bipar. From a knowledge of the device structure it is known that $\mathbf{R_{bcc}}$ should be less than $\mathbf{R_{bec}}$. However, Bipar predicts a larger value - a consequence of the fact that the accuracy of this approach is dependent on a correct prediction for the reverse current gains. $\mathbf{R_{bcc}}$ is therefore set to a smaller value (1Ω) and $\mathbf{R_{cex}}$ is optimised on its own.

© N.V. Philips' Gloeilampenfabrieken

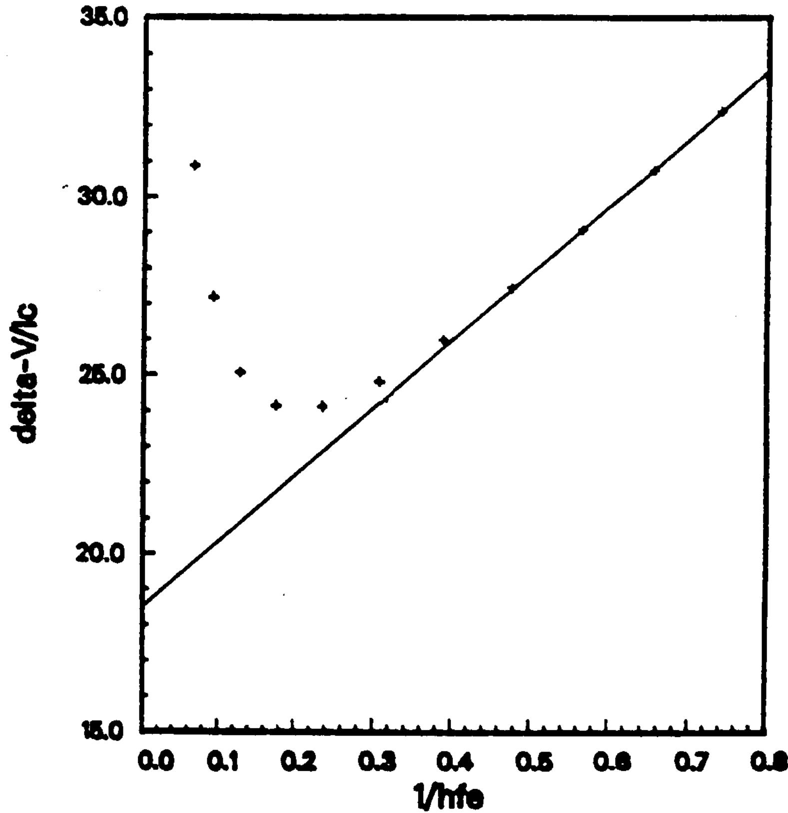


Figure 4.5: *Characteristic curve obtained using the method of Ning and Tang for the V80 device. The straight line extrapolation is indicated.*

The optimisation of the charge parameters is quite straightforward. C-V data is used to optimise the nine depletion parameters and, as can be seen in table 4.2, the resulting values are quite plausible with, for example, the substrate-base diffusion voltage having the lowest value due to the low substrate doping concentration. The three charge parameters $\mathbf{T_{lat}}, \mathbf{T_{fvr}}$ and $\mathbf{T_{fn}}$ are fitted to the f_t characteristics measured at three values of Vcb. The range of Ic chosen for the optimisation is from $10\mu A$ to $1mA$ since the accuracy of the measurements below $10\mu A$ is poor. This is due to a combination of low current values and high input impedance in the common-emitter mode, reducing the accuracy of the s-parameter determination in lateral pnp f_t measurements. The transit times shown in figure 4.3 are in fact those derived using the optimised parameter values obtained here. With regard to the reverse charge parameters $\mathbf{T_{rvr}}$ and $\mathbf{T_{rn}}$, estimates based on the ratio of collector-base and emitter-base capacitance values are used to obtain the values shown in table 4.2.

The result of the above optimisation strategy is a parameter set which is both physically plausible (see table 4.2) and fits the measured data very well (the results of this example can

© N.V. Philips' Gloeilampenfabrieken

be found in chapter 6 where Modella's prediction of device behaviour is compared to measured data and also to that of another compact model - TPL-301).

Optimised parameter values					
Level	501.0	$\mathbf{I_s}$	3.2039×10^{-17}	$\mathbf{B_f}$	80.84
$\mathbf{I_{bf}}$	8.8993×10^{-16}	$\mathbf{V_{lf}}$	4.4801×10^{-2}	$\mathbf{I_k}$	4.5522×10^{-5}
$\mathbf{X_{ifv}}$	0.5000	$\mathbf{E_{afl}}$	12.34	$\mathbf{E_{afv}}$	60.13
$\mathbf{B_r}$	11.09	$\mathbf{I_{br}}$	2.8369×10^{-15}	$\mathbf{V_{lr}}$	8.6736×10^{-19}
$\mathbf{X_{irv}}$	0.3129	$\mathbf{E_{arl}}$	11.32	$\mathbf{E_{arv}}$	13.10
$\mathbf{X_{es}}$	2.6080×10^{-3}	$\mathbf{X_{hes}}$	0.2537	$\mathbf{X_{cs}}$	2.211
$\mathbf{X_{hcs}}$	1.000	$\mathbf{I_{ss}}$	1.0000×10^{-17}	$\mathbf{R_{cex}}$	20.45
$\mathbf{R_{cin}}$	162.0	$\mathbf{R_{bcc}}$	1.000	$\mathbf{R_{bcv}}$	195.0
$\mathbf{R_{bec}}$	6.660	$\mathbf{R_{bev}}$	565.0	$\mathbf{R_{eex}}$	12.10
$\mathbf{R_{ein}}$	35.63	$\mathbf{R_{sb}}$	$1.0000 \times 10^{+16}$	$\mathbf{T_{lat}}$	9.7013×10^{-9}
$\mathbf{T_{fvr}}$	3.1972×10^{-9}	$\mathbf{T_{fn}}$	3.9472×10^{-10}	$\mathbf{C_{je}}$	3.6471×10^{-14}
$\mathbf{V_{de}}$	0.5615	$\mathbf{P_e}$	0.2166	$\mathbf{T_{rvr}}$	1.5658×10^{-8}
$\mathbf{T_{rn}}$	1.6110×10^{-9}	$\mathbf{C_{jc}}$	1.5037×10^{-13}	$\mathbf{V_{dc}}$	0.5817
$\mathbf{P_c}$	0.3546	$\mathbf{C_{js}}$	3.8223×10^{-13}	$\mathbf{V_{ds}}$	0.5170
$\mathbf{P_s}$	0.3792	$\mathbf{T_{ref}}$	21.00		

Table 4.2: *The D.C. parameter determination strategy for the V-80 junction-isolated lateral pnp.*

Physical plausibility can be determined not only by comparing parameter values to those obtained from approximate calculations but also by examining the relative values of certain parameters. A few examples of this plausibility evaluation are listed below.

- An approximate calculation of the lateral emitter resistance using geometry and sheet resistance information yields a value of about 20Ω which is in good agreement with the optimised value. Also, from the device structure a substantially larger value for $\mathbf{R_{cin}}$ is expected and this is indeed the case.

- An approximate calculation of the low current Early voltage based on a one dimensional structure (i.e. using $Veaf = \frac{Qb}{Ccb}$) yields a value at $Vcb = 0$ of $8V$ in good agreement with $\mathbf{E_{afl}}$. Also since the effective base width increases at high current levels due to current crowding, it is expected that $\mathbf{E_{afv}}$ should be larger than $\mathbf{E_{afl}}$. This is also seen to be the case although the optimised value may be too large (arising from the fact that $\mathbf{E_{afv}}$ is quite sensitive to the amount of current crowding described by $\mathbf{R_{ein}}$ and $\mathbf{X_{ifv}}$). Optimised values for $\mathbf{E_{arl}}$ and $\mathbf{E_{arv}}$ posses an even greater physical plausibility.

- Two-dimensional analytical calculations using equation **(2.6)** to estimate $\mathbf{X_{ifv}}$ yield a value, consistent with optimised values for $\mathbf{X_{ifv}}$ and $\mathbf{X_{irv}}$, of 0.43.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

- The ratio between $\mathbf{T_{lat}}$ and $\mathbf{T_{fvr}}$ should reflect the ratio of charge storage in the epilayer between emitter and collector and the charge stored in the epilayer under the emitter. A rough calculation combining geometric data with the theory developed in section **2.3.4** for approximate regional charge storage predicts a factor of three difference between the charge stored in these regions at low injection. This is reflected quite well in the ratio between the optimised values of $\mathbf{T_{lat}}$ and $\mathbf{T_{fvr}}$ shown in table 4.2.

4.4 Discussion

It was mentioned in the parameter determination example that more than one approach existed to determine the knee current parameter, $\mathbf{I_k}$. In addition, it was stated that the physical basis on which Modella was formulated could be useful in separating the quantitative contribution from high injection and current crowding effects. This discussion concentrates on both of these topics because of their importance in relation to the overall strategy.

The knee current parameter, $\mathbf{I_k}$, characterises the onset of high injection effects not only in the forward collector current but also in the reverse emitter current and the forward and reverse substrate currents. It is therefore logical that the approach described in section **4.3.3**, whereby the forward collector current was used in the determination of $\mathbf{I_k}$, should not be the only possibility when extracting $\mathbf{I_k}$ or examining the plausibility of its optimised value.

- The reverse emitter current is similar to the forward collector current in that it is subject to high injection and current crowding effects. Measurement induced correlation will therefore be present if Ie is used to determine $\mathbf{I_k}$. On the other hand, although this approach does not possess any advantages over using the forward Ic, it does constitute a control/check on the physical validity of the optimised value of $\mathbf{I_k}$.

- The forward substrate current also constitutes a method by which the physical validity of $\mathbf{I_k}$, obtained using the forward Ic, can be checked. If, for example, too much emphasis has been placed on current crowding then the fall off in Is will be impossible to model since this fall off is caused solely by high injection. In addition, $\mathbf{I_k}$ itself can be optimised by fitting the forward Is when a large enough fraction of the current is subject to high injection effects. The sensitivity of the fall off in Is to $\mathbf{I_k}$ is a function of the fraction $\mathbf{X_{hes}}$ and when this fraction is greater than about 0.5 good results can be obtained using

© N.V. Philips' Gloeilampenfabrieken

this approach for $\mathbf{I_k}$ determination. When $\mathbf{X_{hes}}$ is ascertained to be less than about 0.5, however, unreliable values for $\mathbf{I_k}$ may result. For this reason the example in section **4.3.3** did not use this approach.

- With junction isolation or 'open' buried layers, the fraction of the reverse substrate current subject to high injection effects is usually larger than in the forward Is and may even have a value of 1.0 as is the case in the example just described. Despite the apparent suitability for use with $\mathbf{I_k}$ determination, this approach is unfortunately unreliable due to the potential influence of an additional effect contributing to the fall off of the reverse Is. Examples of junction isolated devices have been found where a value of $\mathbf{X_{hes}} = 1.0$ was not sufficient to describe the fall off in the reverse Is unless the value of $\mathbf{I_k}$ was reduced to unphysically small values inconsistent with modelling the other high injection effects in the device. Two-dimensional numerical device simulations were therefore performed on these cylindrical devices and a set of approximate 2D $\rightarrow$ 3D scaling rules were developed so that these results could be compared to measured data. This study revealed that in cases such as these, the proximity of the p^{++} isolation to the collector regions was such that the reverse substrate current no longer travelled vertically downwards towards the underlying p^- substrate but instead travelled laterally to the isolation. The reverse substrate current was therefore composed of a lateral current component which was subject to the current crowding effect observed in the forward Ic and the reverse Ie. Due to a relatively large lateral collector resistance in these structures the increase in effective base width for the reverse Is under these conditions leads to considerably more fall off than can be predicted by high injection alone.

To optimise $\mathbf{I_k}$ using the reverse Is in these devices would yield a value which incorporated the effects of the current crowding and which would, therefore, be unphysically small. Because of the uncertainty in ascertaining the absence of current crowding in the reverse Is of junction isolated devices, it is therefore recommended to avoid this approach to $\mathbf{I_k}$ determination. With regard to modelling this current crowding effect in the reverse Is in Modella, an additional two parameters (one to divide the reverse Is into sidewall and bottom components and one to represent the lateral collector resistance) and an extra internal node in the equivalent circuit would be required for physical modelling purposes. Considering the cost of an extra internal node and the small number of devices in which

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

this effect plays a role it was decided, with agreement from designers, not to incorporate this effect into Modella[3].

Regardless of whether Ic, Ie or Is is fitted to optimise $\mathbf{I_k}$, it is important to validate Modella's modelling of the relative quantitative contributions from high injection and current crowding to device behaviour. The physical basis of Modella can be utilised as a form of feedback with respect to the plausibility of the respective parameter values. Consider the Early effect at low and high current levels. The increase in the Early voltage V_{eaf} (defined at constant Ib) at high current levels was shown in section **2.5** to be due to an increase in the effective base width caused by current crowding. However, it was mentioned in section **3.4** that the lateral emitter resistance $\mathbf{R_{ein}}$ could also influence the value of V_{eaf} in the same way in which $\mathbf{R_{eex}}$ increased the value of V_{eaf} when it was defined at constant Veb instead of constant Ib. Experience has shown that in cases where the amount of current crowding has been underestimated by choosing too small a value for $\mathbf{R_{ein}}$ then the increase in V_{eaf} at high current levels is unphysically accounted for by a large increase in $\mathbf{E_{afv}}$. An optimised value of say $\mathbf{E_{afv}} > 10\mathbf{E_{afl}}$ in a typical device is a clear indication that current crowding has been underestimated.

The inability to model the fall off of the $f_t - Ic$ characteristic as a function of reverse Vcb is also often a warning of an incorrect interpretation of the contribution from these two effects. Physical correlation between $\mathbf{I_k}, \mathbf{R_{ein}}$ and $\mathbf{X_{ifv}}$ on one hand and the high current behaviour (see figure 4.3) of the transit times associated with the parameters $\mathbf{T_{lat}}, \mathbf{T_{fvr}}$ and $\mathbf{T_{fn}}$ on the other, provides the link between the DC and charge models. Because the sensitivity of the transit times to Vcb variation decreases in the order $\tau(Q_{flat}), \tau(Q_{fver}), \tau(Q_{fn})$, an inability to model the increase in f_t as a function of Vcb at high current levels may indicate an overestimation of high injection. Too small a value for $\mathbf{I_k}$ means that the dominant transit time component at high current levels may be $\tau(Q_{fn})$ because both $\tau(Q_{flat})$ and $\tau(Q_{fver})$ decrease by a factor of two as a result of the Webster effect. Because the value of $\mathbf{I_k}$ also determines the point in the $f_t - Ic$ characteristic where the fall off starts to occur, an overestimation of high injection effects is often clearly visible from these curves. The situation is complicated though by the fact that current crowding reduces $\tau(Q_{flat})$ but increases $\tau(Q_{fver})$ and $\tau(Q_{fn})$. Therefore, with a large amount of current crowding, $\tau(Q_{fver})$ may be the dominant transit time at high current levels.

[3]The addition of a special knee current parameter to model both high injection *and* current crowding in the reverse Is was also dismissed as an option because it contradicted the entire physical basis on which Modella was formulated.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

In such a case the Vcb dependency of f_t may be adequately described even though $\tau(Q_{flat})$ is negligible. Whatever the situation, it is often very useful to use the parameter set with the Biplot program. This gives an insight into which DC or charge parameters are responsible for the particular form of the f_t curves predicted by Modella and may indicate, in the case of a poor f_t fit, which DC parameters are unphysical.

4.5 Conclusions

A parameter determination strategy for Modella has been developed and tested on a wide range of lateral pnp devices. In the vast majority of cases this strategy has yielded parameter sets which are both physically plausible and fit measured data extremely well. The physical plausibility of parameter values can be determined by comparisons with approximate theoretical calculations and also by examining the relative values of various parameters. In addition, the physical correlation between parameters is an invaluable asset when checking the validity of the relative contributions from the various physical effects. This is because physical correlation between parameters (which is not present in empirical models) means that the effect of having an unphysical parameter value will be noticeable in aspects of model behaviour other than that used for optimisation purposes. The best example of this is the way in which the f_t characteristic can be used to check the validity of parameter values describing high injection and current crowding.

Although this parameter determination strategy requires a certain level of skill on the part of the user to obtain the perfect balance between high injection and current crowding, it is in all other aspects robust, highly automated and reliable. The method of Ning and Tang, for example, has been used to good effect and in the determination of the Early parameters the very definition of the Early voltage has been re-interpreted to provide a more reliable solution. It can be stated therefore that the goal of providing an accurate and efficient parameter determination strategy yielding physically plausible parameter sets has been achieved. Chapter 5 deals with the next aspect of model development - the scaling of these parameter sets with temperature.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 5

Temperature Dependence Modelling

5.1 Introduction

Lateral pnp device characteristics vary as a function of temperature and therefore it is necessary to develop a set of temperature scaling rules for the parameters in Modella. By using these scaling rules, the variation in temperature of device characteristics can be modelled by one parameter set at a reference value of temperature, with parameter sets for other temperatures obtained automatically by scaling. Without temperature scaling, Modella's use in a circuit simulator, for instance, would be restricted to the original temperature at which the parameter set was obtained. The alternative would be the laborious task, involving excessive measurements and parameter extraction, of determining a different parameter set for each temperature value of interest.

Temperature scaling rules themselves contain temperature parameters which have to be fitted in order to provide an accurate description of the temperature dependence of model parameters. The values of these temperature parameters are specific to a particular process since they are a function of process quantities like the doping concentration. This means, in theory, that once the temperature parameters are determined for a single lateral pnp in a process, the same values can be used for all other lateral pnps in that process. With *empirical* compact models the scaling of model parameters with temperature tends to be problematic in this respect due to the lack of a physical basis not only for the temperature scaling but also for the model parameters themselves.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

An important aspect of temperature dependence modelling involves ascertaining which model parameters can be assumed to be temperature *independent*. When possible, it is also advantageous to scale more than one model parameter using the *same* temperature parameter. This reduces the number of temperature parameters and therefore the total number of parameters in a compact model. The approach to developing a set of scaling rules should ideally combine physical derivations with a more pragmatic approach. This chapter describes how this was attempted in the case of Modella. In section **5.2** the derivation of the scaling rules is described and this is followed in section **5.3** by a description of the procedure to determine the temperature parameters in order to verify the choice of scaling rules. This section also includes some initial results concerning the determination of temperature parameters and the variation of device characteristics with temperature. Some conclusions with regard to Modella's temperature scaling then follow in section **5.4**.

5.2 Temperature scaling rules

5.2.1 Background

The origins of the temperature dependence of device characteristics lie in the following three effects:

- the *intrinsic carrier concentration*, n_i, is a strong function of temperature which can be described according to ([15])

$$n_1^2 = 9.61 \times 10^{32}\, T^3 \,\exp \frac{-qV_{g0}}{kT} \quad (cm^{-6}) \tag{5.1}$$

  where V_{g0} is the bandgap voltage in the doping region in question. The intrinsic carrier concentration appeared in many of the expressions derived in chapter 2 and it is therefore not surprising that many device characteristics vary strongly with temperature.

- the *carrier mobility*, $\mu_{n,p}$, varies with temperature according to

$$\mu_{n,p} \propto T^{-S_{n,p}} \tag{5.2}$$

  and therefore using the Einstein relation we can also write for the *carrier diffusion coefficient*

$$D_{n,p} \propto T^{1\,-\,S_{n,p}} \tag{5.3}$$

© N.V. Philips' Gloeilampenfabrieken

These expressions, which are approximations valid for the range 250 – $450K$, have direct consequences for the temperature dependence of saturation currents, series resistances and transit times. The exponents $S_{n,p}$ are functions of the doping levels as shown in figure 5.1.

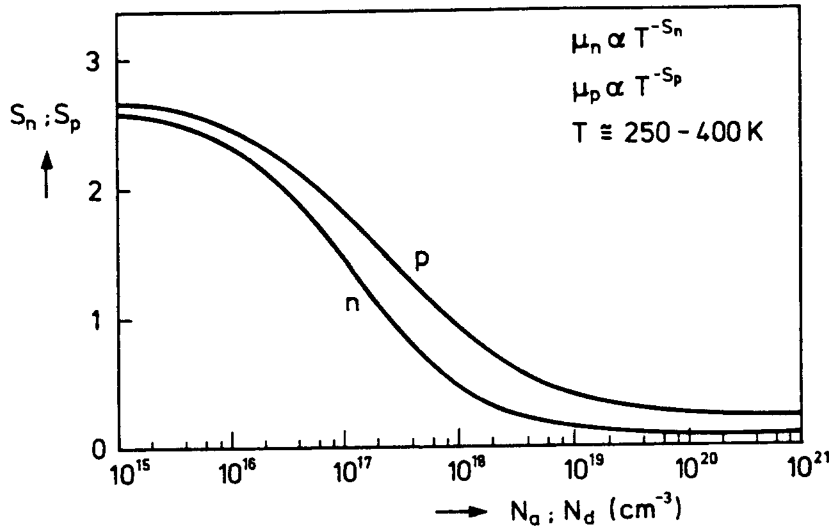


Figure 5.1: *The exponents of the power-law approximations of the mobility temperature dependence as a function of doping concentrations.*

- at certain doping levels (for example those typically found in the lateral pnp emitter and collector) the impurities may not be fully ionised. This fraction of ionised impurities, $g_i = \frac{N_a^-}{N_a}$, varies with temperature and can be roughly approximated in the range 250 - 450 K by $T^{0.3}$ to $T^{0.5}$. In the lateral pnp this only affects the ideal base current.

Given that the temperature dependencies of these basic physical quantities can be approximated as shown above, the next section describes how Modella's parameters relate to these basic quantities and therefore how the parameters are expected to vary with temperature. Parameters which have no theoretical relation to these three basic temperature variation mechanisms are assumed temperature independent.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

5.2.2 Derivation of scaling rules

The diffusion voltage Vd of a junction is dependent on n_i. For asymmetric abrupt junctions we can define[1] the diffusion voltage as

$$Vd = 2\,Vt\,ln\left(\frac{N}{n_i}\right) \tag{5.4}$$

where N is the doping level on the lower doped side. When equation (**5.1**) is substituted for n_i we obtain

$$Vd(T) = V_{g0} + Vt\,ln\left(\frac{N^2}{9.61 \times 10^{32}\,T^3}\right) \tag{5.5}$$

The formulation of a temperature scaling rule reflects the fact that we want to scale the parameter value at the reference temperature, T^r, to that at an arbitrary temperature T. Defining the diffusion voltage at the reference temperature as $Vd^{\,r}$ the above equation becomes

$$Vd(T) = \frac{T}{T^r}\,Vd^{\,r} + \left(1 - \frac{T}{T^r}\right)V_{g0} - 3\,Vt\;ln\left(\frac{T}{T^r}\right) \tag{5.6}$$

This general temperature scaling rule for the diffusion voltage has two quite different applications in Modella. Firstly, it is used to scale the diffusion voltage parameters $\mathbf{V_{de}}$, $\mathbf{V_{dc}}$ and $\mathbf{V_{ds}}$ associated with the emitter-base, collector-base and substrate-base depletion capacitances. The temperature parameters are $\mathbf{V_{geb}}$, $\mathbf{V_{gcb}}$ and $\mathbf{V_{gsb}}$ which represent the bandgap voltages in the respective depletion regions. Secondly, the model 'constant' $Vd = 0.6V$, used in the modelling of the Early effect, is scaled using this rule. A separate bandgap voltage parameter $\mathbf{V_{gb}}$ associated with the epilayer region between the emitter and collector sidewalls is required here (this temperature parameter is used in the scaling of $\mathbf{I_s}$).

The zero-voltage depletion capacitance of a junction depends on Vd according to

$$C_0 \propto Vd(T)^{-p} \tag{5.7}$$

where p is the grading coefficient (for an abrupt junction we have $C_0 = \sqrt{\frac{q\epsilon N}{2\,Vd(T)}}$). This leads

[1]Here the assumption is that the $-2Vt$ in the definition of Vd given in section **2.3.1** can be neglected.

© N.V. Philips' Gloeilampenfabrieken

directly to the scaling rules for the zero-voltage depletion capacitances used in Modella i.e.

$$\mathbf{C_{j_x}} = \mathbf{C_{j_x}^r} \times \left(\frac{\mathbf{V_{d_x}}^r}{\mathbf{V_{d_x}}}\right)^{P_x} \tag{5.8}$$

A one-dimensional approximation for the Early voltage can be given by[4]

$$\begin{aligned} V_{eaf} &= \frac{Qb_{maj}}{c_{cb}} \\ &= \frac{qN_{epi}(Xb - Wcb)}{\epsilon / Wcb} \end{aligned} \tag{5.9}$$

Here Qb_{maj} = base majority charge concentration

c_{cb} = collector-base capacitance/unit area

Since $Wcb \propto \sqrt{Vd(T)}$ and assuming $Wcb << Xb$ at $Vcb = 0V$, the Early voltage parameters (defined at $Vcb = 0V$) can be assumed to vary with temperature according to $\sqrt{Vd(T)}$. The scaling rule for $\mathbf{E_{afl}}$, for example, therefore reads

$$\mathbf{E_{afl}} = \mathbf{E_{afl}}^r \times \sqrt{Vd/Vd^r} \tag{5.10}$$

Here, Vd is the model 'constant' diffusion voltage since it is appropriate that the Early voltages should depend on the same bandgap voltage $\mathbf{V_{gb}}$ that is used with $\mathbf{I_s}$ as seen below.

The saturation current $\mathbf{I_s}$ is proportional to the product $D_p\ n_{ib}^2$ where n_{ib} is the intrinsic carrier concentration in the epitaxial base between emitter and collector. From this we obtain for the temperature dependence of $\mathbf{I_s}$

$$\begin{aligned} \mathbf{I_s}(T) &\propto T^3 \exp\left(\frac{-q\mathbf{V_{gb}}}{kT}\right) T^{1-\mathbf{S_{pb}}} \\ &= T^{4-\mathbf{S_{pb}}} \exp\frac{-q\mathbf{V_{gb}}}{kT} \end{aligned} \tag{5.11}$$

where $\mathbf{S_{pb}}$ is the temperature coefficient of the hole mobility in the epitaxial base between emitter and collector.

In order to derive a scaling rule for $\mathbf{B_f}$ we must divide the temperature dependence of $\mathbf{I_s}$ by that of the ideal base current. The latter is determined by $D_n\ n_{ie}^2\ /\ N_e$,

where D_n = minority carrier diffusion coefficient in the emitter

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

	n_{ie}	$=$	intrinsic carrier concentration in the emitter
and	N_e	$=$	emitter doping concentration

The temperature dependencies of these three quantities are $T^{1-S_{ne}}$, $T^3 \exp \frac{-q\mathbf{V_{ge}}}{kT}$ and g_i respectively. S_{ne} is the temperature coefficient of emitter electron mobility and $\mathbf{V_{ge}}$ is the emitter bandgap voltage. Combining these dependencies with that of $\mathbf{I_s}$ we obtain for $\mathbf{B_f}$

$$\begin{aligned} \mathbf{B_f}(T) &\propto \frac{g_i T^{4-\mathbf{S_{pb}}} \exp \frac{-q\mathbf{V_{gb}}}{kT}}{T^{4-S_{ne}} \exp \frac{-q\mathbf{V_{ge}}}{kT}} \\ &= T^{\mathbf{A_e} - \mathbf{S_{pb}}} \exp \frac{-q(\mathbf{V_{gb}} - \mathbf{V_{ge}})}{kT} \end{aligned} \tag{5.12}$$

Because of the uncertainty in the temperature coefficient relating to g_i the parameter $\mathbf{A_e}$ combines this coefficient with S_{ne}.

The saturation current of the non-ideal forward base current, $\mathbf{I_{bf}}$, is proportional to $\frac{n_i}{\tau}$, where τ is the temperature dependent lifetime in the emitter-base depletion region. This dependence is given by [15]

$$\mathbf{I_{bf}}(T) \propto T^2 \exp \frac{-q\mathbf{V_{gje}}}{2kT} \tag{5.13}$$

where $\mathbf{V_{gje}}$ is the bandgap voltage at the recombination peak in the emitter-base junction (which does not necessarily correspond to bandgap voltage used for emitter-base diffusion voltage).

The parameter $\mathbf{V_{lf}} = (2\ (E_t - E_i) - 2kT\ ln\ g_t)\ /\ q$ where $E_t - E_i$ is the difference between the trap and intrinsic energy levels. Considering typical doping levels in lateral pnp devices, the degeneracy factor g_t [15] can be assumed to have a value of one which means that $\mathbf{V_{lf}}$ can be taken as temperature independent.

Scaling the high injection knee current parameter, $\mathbf{I_k}$, does not require any additional temperature parameters because $\mathbf{I_k} \propto D_p$ and is scaled using $\mathbf{I_k} \propto T^{1-\mathbf{S_{pb}}}$.

The saturation current of the substrate-base diode, $\mathbf{I_{ss}}$, is used to model the zero base current phenomena - an effect normally only observed at high temperatures. Using an ideal diode equation to model what is essentially a generation current originating from the reverse-biased

© N.V. Philips' Gloeilampenfabrieken

substrate-base junction is somewhat unphysical. However, this pragmatic modelling approach only requires one parameter i.e. $\mathbf{I_{ss}}$. With regard to the temperature scaling of $\mathbf{I_{ss}}$ it is however necessary to derive a scaling rule associated with generation and therefore $\mathbf{I_{ss}}$ is taken as being proportional to n_i/τ. This leads to $\mathbf{I_{ss}} \propto T^2 \exp \frac{-q\mathbf{V_{gsb}}}{2kT}$ where the bandgap voltage of the substrate-base depletion region, $\mathbf{V_{gsb}}$, can be used as the temperature parameter. This is possible because the generation rate in a reverse-biased junction is approximately constant over the depletion region, unlike the sharp peak in the recombination rate in a forward-biased junction which tends to be located closer to the higher doped region.

Turning now to the transit times, the temperature dependencies of $\mathbf{T_{lat}}$, $\mathbf{T_{fvr}}$ and $\mathbf{T_{fn}}$ are determined by the minority carrier mobility temperature variation in the respective charge storage regions. $\mathbf{T_{lat}}$ and $\mathbf{T_{fvr}}$ are proportional to D_p^{-1} in the epilayer between emitter and collector and are scaled accordingly by $T^{\mathbf{S_{pb}}-1}$. $\mathbf{T_{fn}}$ combines emitter and buried layer minority mobilities and the scaling rule therefore uses a combined temperature coefficient $\mathbf{S_x}$ in place of $\mathbf{S_{pb}}$.

In Modella the scaling of resistances with temperature is done according to $T^{S_{n,p}}$ due to the inverse relation between resistance and majority carrier mobility (at low injection $R \propto \frac{1}{q\mu N}$). The coefficients chosen are of course different for each doping region which means that some new temperature parameters have to be introduced. By using the method of Ning and Tang to extract values for $\mathbf{R_{eex}}$ and $\mathbf{R_{cex}}$ at different temperatures (see figure 5.2), it has also been verified that these *contact* resistances are temperature independent - an assumption also made for example in Mextram. $\mathbf{R_{bec}}$ and $\mathbf{R_{bcc}}$ however, are seen to be temperature dependent because they are composed not only of the base contact resistance, but also the resistance associated with the buried layer and sometimes also the epilayer under the base contact.

A complete list of Modella's temperature scaling rules (including some of the reverse parameters which have been omitted from the above derivation because of their obvious symmetry) can be found in appendix A. These rules are of course formulated so as to scale the reference value of the parameter in question. Appendix A also contains the complete parameter list for Modella which can be seen to contain thirteen extra parameters associated with temperature scaling. Although comparable to the number of temperature parameters in the Mextram model, it would desirable to reduce this number. This can only be done however after the temperature

© N.V. Philips' Gloeilampenfabrieken

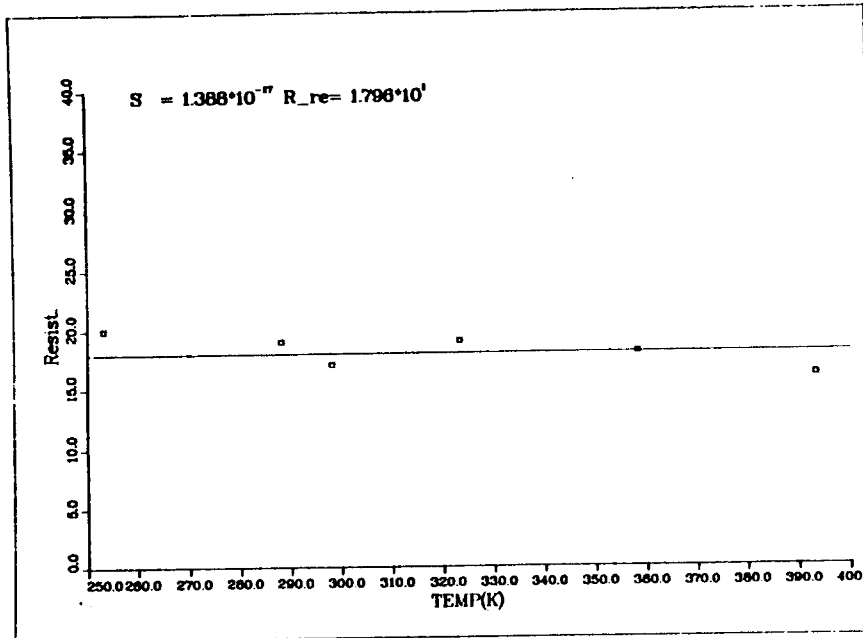


Figure 5.2: *Temperature (in)dependence of* $\mathbf{R_{eex}}$ *obtained by extracting* $\mathbf{R_{eex}}$ *using the method of Ning and Tang at different temperatures.*

parameters and scaling rules have been tested on lateral pnp transistors from the various in-house processes. The next section describes how these temperature parameters can be obtained for the purpose of scaling rule verification.

5.3 Verification

5.3.1 Introduction

The verification of temperature scaling rules in a compact model requires a great deal of measurement data and parameter determination. Because the procedure has to be repeated for each process (and variations therein), a complete verification can only be obtained with the assistance of the various process characterisation departments. This section is therefore concerned with establishing the *procedure* for temperature parameter determination and subsequent scaling rule verification. An example is used to show this procedure in operation and to provide a some insight into how well these scaling rules should perform during complete verification.

5.3.2 Procedure

The following procedure describes how to determine the temperature parameters of a device for subsequent scaling rule verification. In practice, this procedure will be performed on a variety of

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

lateral pnps within a given process to ensure that the set of temperature parameters is common to all such devices within the process. To determine the temperature parameters:

- a complete set of D.C, capacitance and f_t measurements is required at various temperatures. The particular range of temperatures chosen depends on the expected circuit operation temperatures but an example would be a choice of five to ten different temperatures in the range 250 - 400 K.
- for each of these measurement temperatures Modella's D.C. and charge parameters have to be determined using Bipar. The critical aspect of this determination is that each of these parameter sets should be physical so that the variation in extracted parameter values is solely a function of temperature.
- a least squares optimisation procedure is then used to extract values for the temperature parameters. Using the scaling rules, these temperature parameters are fitted to the 'measurement' points (i.e. the values of Modella's parameters extracted at each temperature). The order in which the temperature parameters are optimised is described in the next section.

The verification of the temperature scaling rules can be considered to consist of two stages. Firstly, a good fit to the 'measurement' points in the above procedure is a good indication that the scaling rule for the parameter in question is valid. Secondly, the combined effect of all the temperature scaling rules can be investigated by scaling the parameter set from a reference temperature to an arbitrary temperature. By comparing measured device characteristics at this arbitrary temperature to the model characteristics obtained using the scaled parameter set, the entire set of scaling rules may be verified. This latter verification stage also provides a test on the assumption that some parameters do not scale with temperature.

5.3.3 Results

Measurements were performed on a typical junction isolated lateral pnp transistor from the 'L48' process. The temperatures chosen were -20, 15, 25, 50, 85 and 120 $°C$. At each of these measurement temperatures a determination of all D.C. and charge parameters was performed according to the manner described in the last chapter. Using these model parameter values as the 'measurement' points, the temperature parameters were extracted. The strategy for

© N.V. Philips' Gloeilampenfabrieken

the determination of these temperature parameters, described below, was developed using two criteria:

1. because the scaling of some model parameters is described by the same temperature parameter, the choice of model parameter to use for fitting purposes depends on which model parameter exhibits the greatest sensitivity to the temperature parameter in question.

2. the variation in the model parameters must be solely a function of temperature if the temperature parameters are to be correctly fitted. The degree of confidence in the physical plausibility of model parameters therefore plays a role in the choice of strategy.

The bandgap voltages of the three depletion regions are firstly fitted using *both* the diffusion voltages and zero volt capacitances of the respective junctions. Figure 5.3 shows the results for the largest of these capacitances i.e. the substrate-base capacitance. The grading coefficient is assumed constant and set to its reference value of 0.35. An optimised value of $\mathbf{V_{gsb}} = 1.204V$ provides a good description of the temperature dependency of both $\mathbf{C_{js}}(T)$ and $\mathbf{V_{ds}}(T)$. Optimised values for the other two bandgap voltages $\mathbf{V_{geb}}$ and $\mathbf{V_{gcb}}$ are $1.032V$ and $1.206V$ (the theoretical maximum in silicon) respectively.

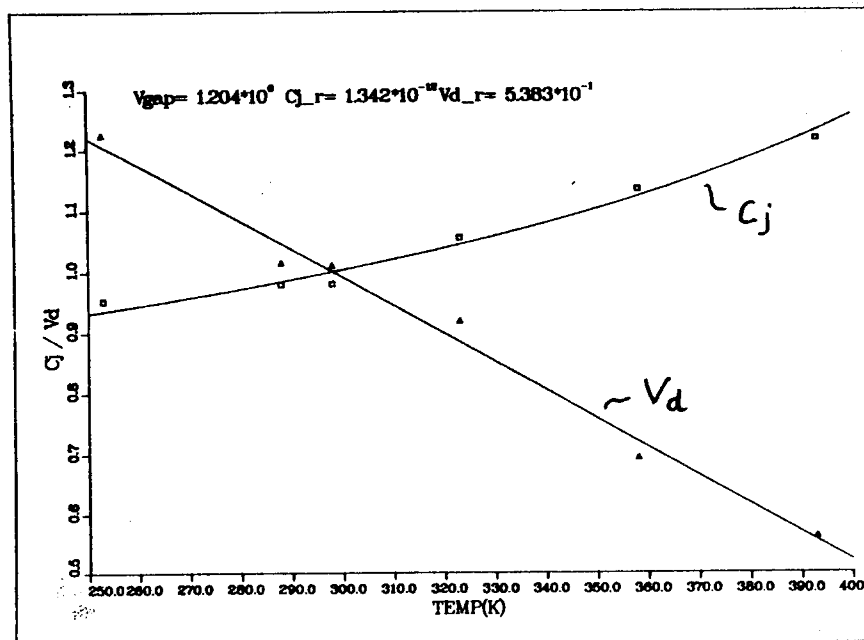


Figure 5.3: *Optimisation of* $\mathbf{V_{gsb}}$ *by fitting the temperature variation of* $\mathbf{C_{js}}$ *and* $\mathbf{V_{ds}}$*. The points are values extracted for* $\mathbf{C_{js}}$ *and* $\mathbf{V_{ds}}$ *at each temperature and the solid lines are the scaling rule predictions.*

© N.V. Philips' Gloeilampenfabrieken

The temperature dependence of $\mathbf{I_s}$ is described by $\mathbf{S_{pb}}$ and $\mathbf{V_{gb}}$[2]. Normally these two temperature parameters would be optimised by fitting the $\mathbf{I_s}(T)$ but in this case $\mathbf{S_{pb}}$ is estimated from $\mathbf{T_{lat}}(T)$ and $\mathbf{T_{fvr}}(T)$. This was because $\mathbf{S_{pb}}$ determined from $\mathbf{I_s}(T)$ overestimates the temperature variation of the transit times. Therefore, $\mathbf{S_{pb}}$ is set to 1.8 and $\mathbf{V_{gb}}$ is optimised using $\mathbf{I_s}(T)$. The result is $\mathbf{V_{gb}} = 1.173V$ and $\mathbf{I_s}(T)$ is well described as seen in figure 5.4. Moreover, this value of $\mathbf{V_{gb}}$ also provides a reasonable description of the temperature dependence of the Early parameters. Figure 5.5 shows this dependence for $\mathbf{E_{afl}}$.

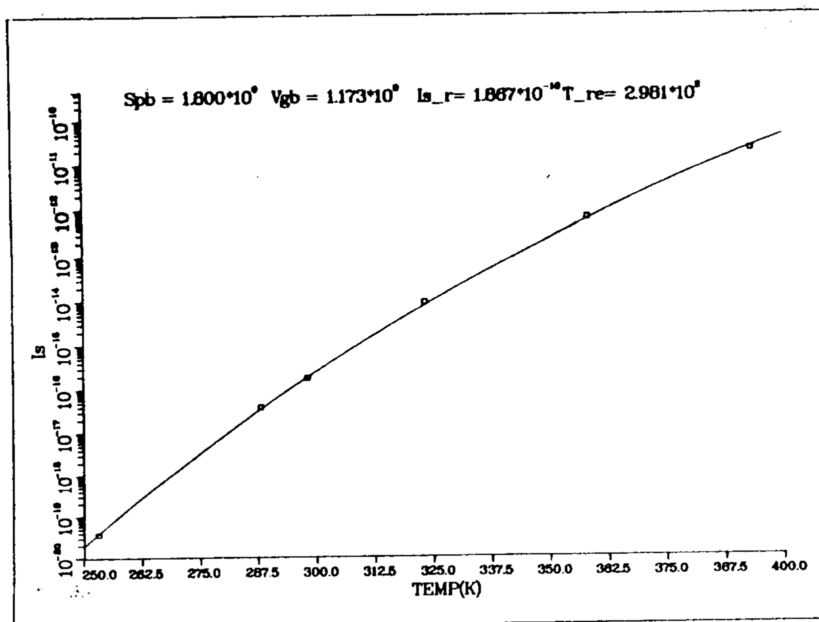


Figure 5.4: *Optimisation of* $\mathbf{V_{gb}}$ *by fitting the temperature variation of* $\mathbf{I_s}$. $\mathbf{S_{pb}}$ *was set to* 1.8 *in this optimisation.*

The temperature parameters $\mathbf{A_e}$ and $\mathbf{V_{ge}}$ are optimised using the temperature variation of $\mathbf{B_f}$ yielding $\mathbf{A_e} = 3.536$ and $\mathbf{V_{ge}} = 1.206V$ (see figure 5.6). At $T = 50°C$ the 'measured' value of $\mathbf{B_f}$ was omitted from the optimisation as it had an unphysically large value. The overall fit to $\mathbf{B_f}(T)$ is, however, reasonable.

$\mathbf{V_{gje}}$ is fitted using the temperature dependence of $\mathbf{I_{bf}}$. A value of $1.202V$ is obtained by fitting only the first four points in figure 5.7 in order to avoid large leakage currents at high

[2] In Modella the saturation current of the forward substrate current is described by $\mathbf{X_{es}}$ $\mathbf{I_s}$, where $\mathbf{X_{es}}$ is a dimensionless fraction. The temperature scaling of the substrate saturation current cannot therefore be separated from that of the collector current. Because of the different doping levels in the regions involved, the validity of this approach requires further investigation. It may yet be more appropriate to exchange $\mathbf{X_{es}}$ for a new saturation current parameter having its own temperature scaling rule.

© N.V. Philips' Gloeilampenfabrieken

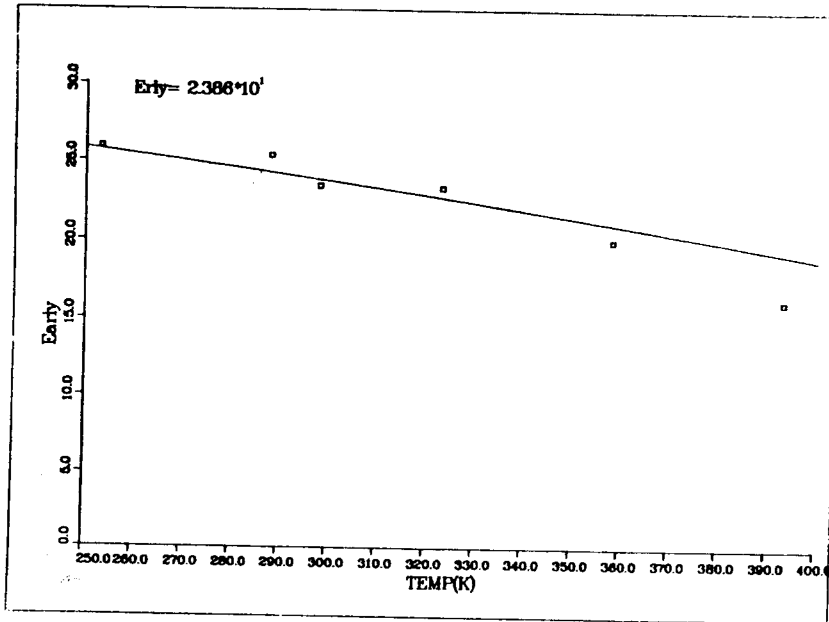


Figure 5.5: *Temperature dependence of* $\mathbf{E_{afl}}$. *The temperature scaling prediction (solid line) uses* $\mathbf{V_{gb}} = 1.173$ *obtained from optimising* $\mathbf{I_s}(T)$.

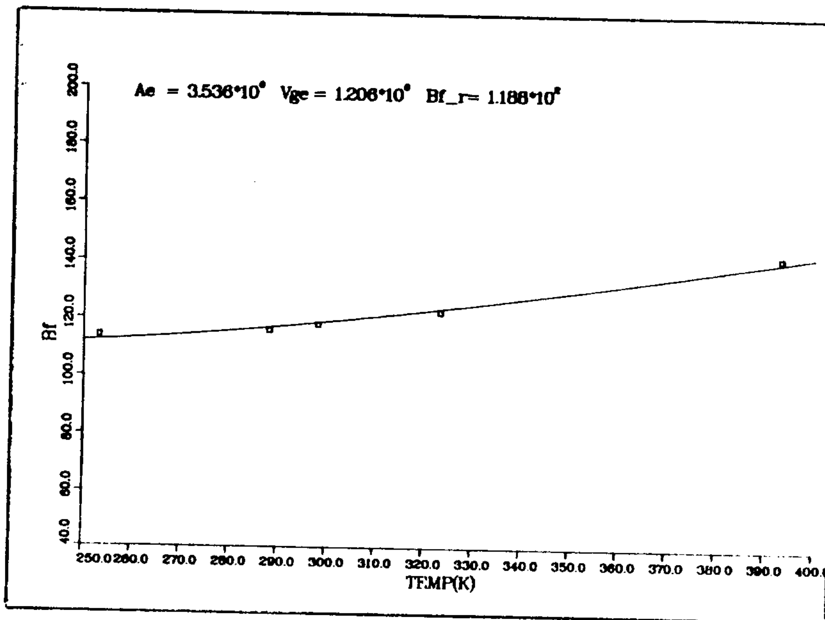


Figure 5.6: *Optimisation of* $\mathbf{A_e}$ *and* $\mathbf{V_{ge}}$ *by fitting the temperature variation of* $\mathbf{B_f}$.

temperatures. In addition, the values for $\mathbf{I_{bf}}$ at these high temperatures are rather dubious due to the difficulty of extracting $\mathbf{I_{bf}}$ in the presence of the zero base current phenomena.

The temperature coefficients of the resistances are optimised by fitting the respective temperature variations in the resistance values. Figure 5.8 shows the fit to $\mathbf{R_{bec}}$ by optimising $\mathbf{S_{nbn}}$. Note the lack of deviation in the values of $\mathbf{R_{bec}}$ obtained using the method of Ning and

© N.V. Philips' Gloeilampenfabrieken

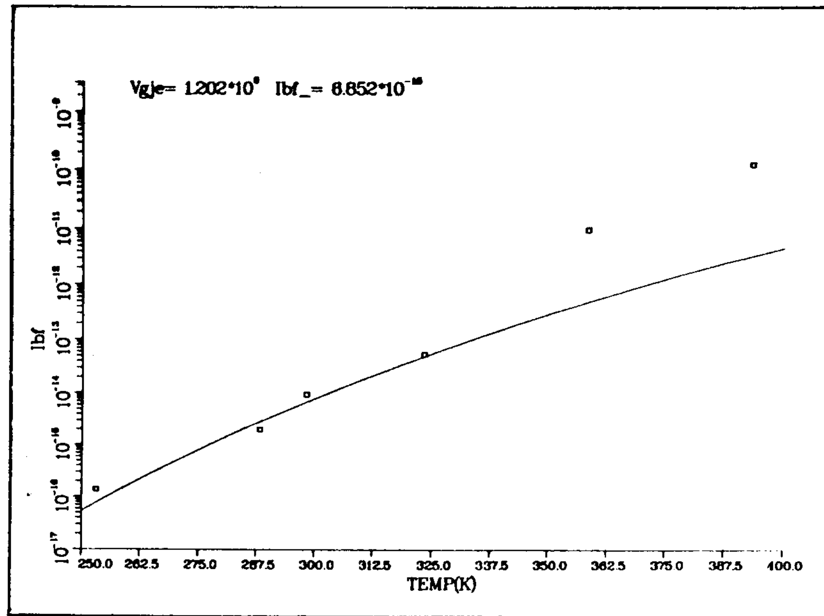


Figure 5.7: *Optimisation of* $\mathbf{V_{gje}}$ *by fitting the temperature variation of* $\mathbf{I_{bf}}$.

Tang. This contrasts to the problem of obtaining physical values for $\mathbf{R_{ein}}$ because of measurement induced correlation. Deviation is clearly visible in figure 5.9 where $\mathbf{S_{pe}}$ is optimised using $\mathbf{R_{ein}}(T)$.

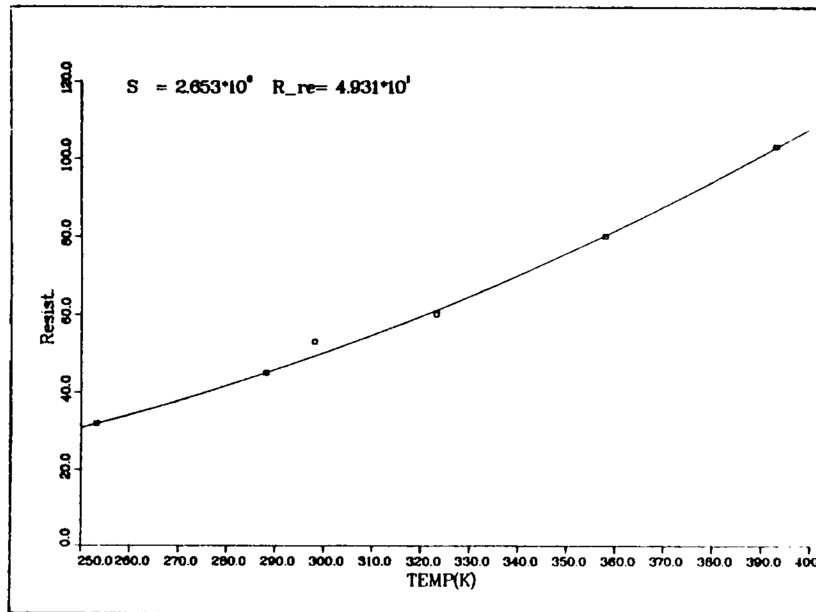


Figure 5.8: *Optimisation of* $\mathbf{S_{nbn}}$ *by fitting the temperature variation of* $\mathbf{R_{bec}}$.

As described in chapter 4, $\mathbf{R_{bev}}$ is calculated theoretically and accordingly $\mathbf{S_{nb}}$ is given a theoretical value of 2.5 in order to provide values for $\mathbf{R_{bev}}$ at temperatures other than the

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

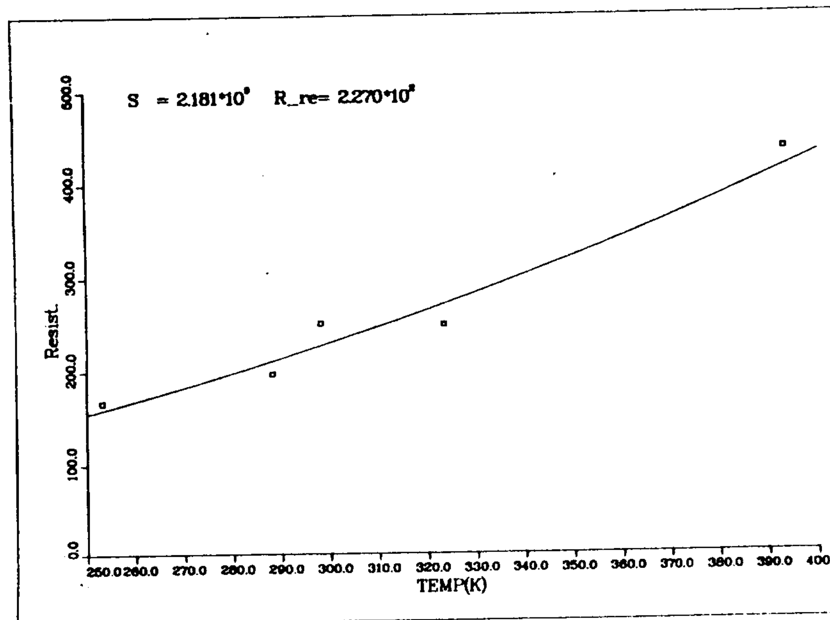


Figure 5.9: *Optimisation of* $\mathbf{S_{pe}}$ *by fitting the temperature variation of* $\mathbf{R_{ein}}$.

reference temperature. The remaining temperature parameters $\mathbf{S_{pc}}$ and $\mathbf{S_x}$ are determined by fitting $\mathbf{R_{cin}}(T)$ and $\mathbf{T_{fn}}(T)$ respectively, yielding $\mathbf{S_{pc}} = 1.86$ and $\mathbf{S_x} = 1.3$.

Having determined the temperature parameters, the device characteristics at $-20°C$ and $120°C$ were modelled by scaling the reference parameter set from $25°C$. Figure 5.10 shows the agreement between the model predictions and measurements for (a) the forward Gummel plot (b) the forward current gain h_{fe} and (c) the f_t characteristic. As can be seen, the scaling rule predictions agree very well with measured values. This unexpectedly good description of the temperature dependence of device characteristics in the first ever verification example was, however, not achieved without a price. Close examination of the optimised temperature parameter values reveals that, even though some values agree well with theory, others were somewhat unphysical[3]. As stated in section **5.3**, a prerequisite to obtaining physical temperature parameter values is to have physical sets of model parameters at each measurement temperature. Unfortunately, due to measurement difficulties and time constraints (this example was done only to illustrate the verification procedure) this prerequisite was not always met. As seen with $\mathbf{R_{ein}}$ (figure 5.9), some parameter values exhibited unexpected deviation as a function of temperature which therefore made the extraction of physical values for certain temperature parameters more difficult.

[3]For example, the parameters $\mathbf{S_{pb}} = 1.8$ and $\mathbf{V_{gb}} = 1.173V$ should in theory be close to their maximum values of about 2.7 and 1.206 respectively because of the low doping concentration in the epilayer. This can be seen from figure 5.1 and the fact that bandgap narrowing is not expected at such low doping levels.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

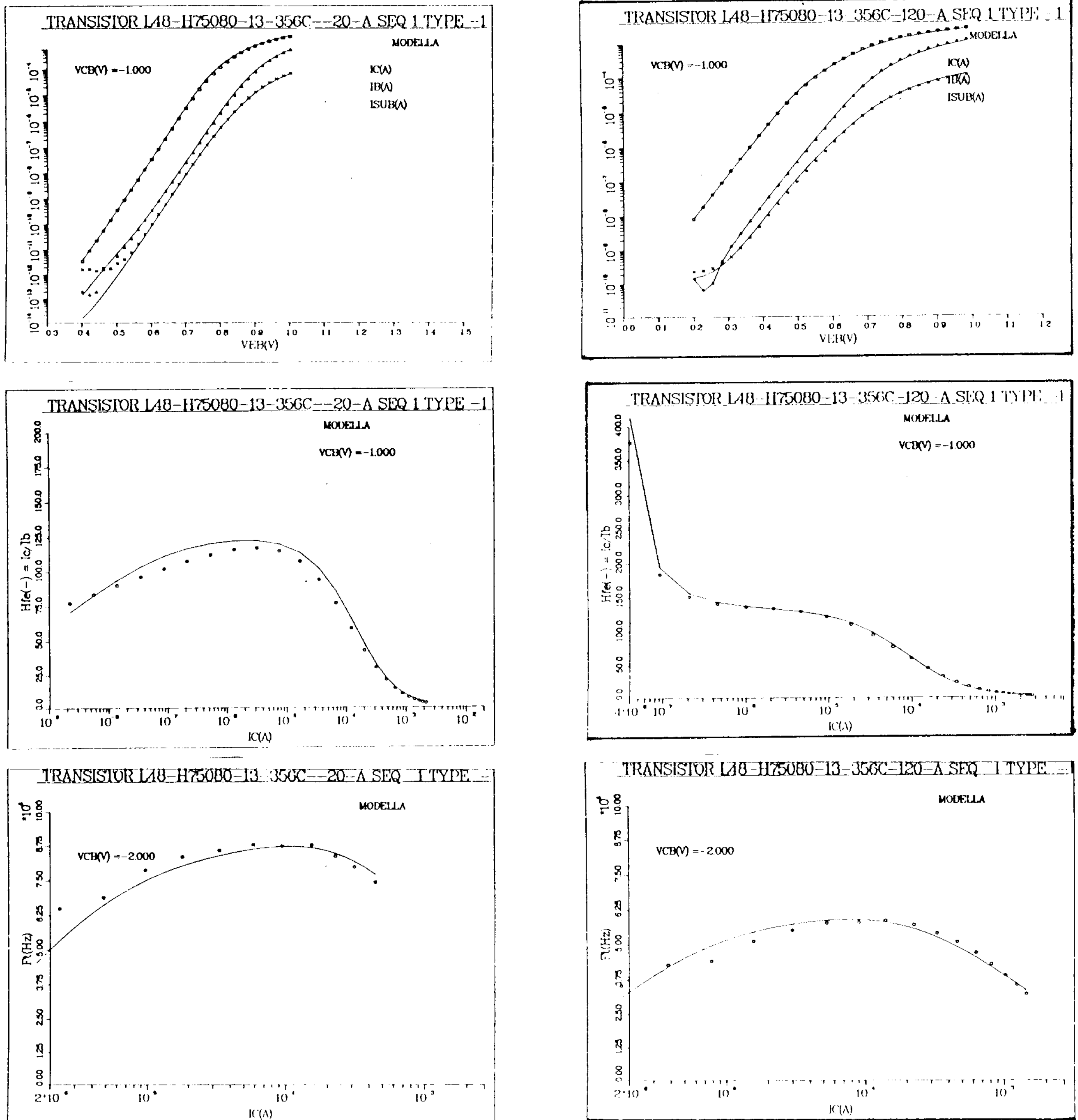


Figure 5.10: *Plots of (a) the forward Gummel, (b) the forward h_{fe} and (c) the f_t characteristics at $-20°C$ and $120°C$. The points are measured values and the solid lines are model predictions obtained by scaling the reference parameter set from $25°C$ to $-20°C$ and $120°C$ using the optimised temperature parameters. Measurement inaccuracy is responsible for the disturbance in the low current region of the $-20°C$ Gummel plot.*

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

5.4 Conclusions

A complete set of temperature scaling rules for Modella has been derived from the temperature variation of three fundamental quantities - the intrinsic carrier concentration, the carrier mobility and the presence of non-ionised impurities. By relating model parameters to these mechanisms in a physical way, a compact set of scaling rules with thirteen temperature parameters was obtained. The procedure to determine these temperature parameters has been described and a worked example to illustrate this procedure has been presented. In this example, two stages in the verification of temperature scaling rules were evident - namely the ability to scale the individual parameters and also the combined influence of all scaling rules describing the variation of device characteristics with temperature. This latter stage also involved verifying the assumed temperature *independence* of certain model parameters like $\mathbf{R_{eex}}$ and $\mathbf{X_{ifv}}$. Although this example was only used to illustrate the verification procedure, the results were quite encouraging.

It may therefore be concluded that the aims of deriving a physical set of temperature scaling rules for Modella's parameters and of establishing a procedure to determine the associated temperature parameters have been achieved. Initial results are good, paving the way for a complete verification using devices from different processes.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 6

Model comparisons

6.1 Introduction

Before Modella can be implemented in a circuit simulation program it is necessary to prove that this new compact model constitutes a significant improvement over present models. To this end, a qualitative and quantitative comparison between Modella and the best presently available lateral pnp compact model will be presented. Inherent in this comparison is a test of the validity of all the work presented in the previous chapters. Furthermore, this comparison highlights the problems associated with present day lateral pnp compact modelling and shows whether or not Modella is the solution to these problems.

At present, the best lateral pnp compact model available is the Extended Gummel-Poon model, referred to here as the E-GP model[1]. The E-GP model is described in [15] and [10], but for the sake of continuity a brief description of the model is given in the next section. Section 6.3 outlines the main improvements in Modella's formulation compared to that of the E-GP model. These improvements are expected to provide benefits not only in the accuracy with which measured device characteristics are modelled but also in terms of the advantages which a physically based formulation possesses over a semi-empirical model (see section **1.3**). The example chosen in chapter 4 to illustrate the parameter determination strategy is then used in section **6.4** to show how Modella actually compares with the E-GP model in practice. Some conclusions on Modella's performance follow in section **6.5**.

[1]Within Philips, the E-GP model is known as TPL-301.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

6.2 The extended Gummel-Poon model

The E-GP model provides an extensive description of the integrated circuit lateral pnp transistor (junction and oxide isolated). It is intended for use with DC, transient and AC analyses at all current levels (i.e. at low injection and high injection) and represents a significant improvement over lateral pnp compact models of the Ebers-Moll type[15]. The model has 3 internal nodes and 32 model parameters[2] (excluding temperature parameters) and incorporates high injection effects (in the forward Ic, reverse Ie and in the substrate currents), ideal and non-ideal base currents, a current dependent base resistance, the Early effect and charge storage effects. The equivalent circuit for the E-GP model is shown in figure 6.1. A first glance at the equivalent circuit reveals a structure not unlike Modella. However, closer examination of the circuit and the model equations reveals some fundamental differences. The main currents I_1 and I_2 are given by

$$I_1 = \frac{I_F}{q_B}$$
$$I_2 = \frac{I_R}{q_B}$$

where

$$I_F = IS \exp\left(\frac{V_{E'B'}}{Vt}\right) - 1$$
$$I_R = IS \exp\left(\frac{V_{C'B'}}{Vt}\right) - 1$$
$$q_B = \frac{q_1}{2}\left(\sqrt{1 + 4q_2} + 1\right)$$

and where

$$q_1 = \left(1 - \frac{V_{C'B'}}{VEAF} - \frac{V_{E'B'}}{VEAF}\right)^{-1}$$
$$q_2 = \left(\frac{I_F}{IKF}\right)^{NHF} + \frac{I_R}{IKR}$$

The parameters in these equations are the saturation current, IS, the Early voltage related to the collector-base junction, $VEAF$, the high injection knee current of the forward beta, IKF, the base-widening exponent(!), NHF, and the high injection knee current of the reverse beta, IKR.

[2]In MODELLA there are 6 internal nodes and 42 model parameters.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

The forward ideal and non-ideal base currents are given by

$$\begin{aligned} I_{RE} &= \frac{I_F}{BF} \\ I_{LE} &= ILF\left(\exp\left(\frac{V_{E'B'}}{NLF\,Vt}\right) - 1\right) \end{aligned}$$

and the forward substrate current is given by

$$I_4 = XES\,I_F\left(\frac{2}{1 + \sqrt{1 + 4\frac{I_F}{IKS}}}\right)$$

The parameters in these equations are the ideal forward common-emitter current gain, BF, the saturation current, ILF, and the non-ideality factor, NHF, of the forward non-ideal base current, the current fraction of the e-b-s transistor, XES, and the high injection knee current of the forward substrate current, IKS. In the reverse active case, similar equations with reverse parameters are used to model the reverse base and substrate currents (the exception being the reverse substrate current, I_3, which uses the forward knee current parameter, IKS).

The saturation current of the substrate-base diode (I_{SF}) is taken from the reverse substrate current i.e. $I_{SF} = XCS\ IS\ (\exp(\frac{V_{sb}}{Vt})\ -\ 1)$.

The forward stored charge is modelled as

$$Q_{ED} = \tau_B\,I_1\,{q_1}^2 + \tau_{NE}\,IS\left(\frac{I_F + IS}{IS}\right)^{1/M_T}$$

where the parameters τ_B and τ_{NE} are the base and neutral emitter transit times respectively and M_T is a fit parameter.

From this brief description of the E-GP model a few points should be noted.

- one emitter/collector resistance is used and therefore current crowding is not modelled. Sidewall and bottom components of currents and charges cannot, therefore, be correctly distinguished.
- *three* knee current parameters are used in the modelling of high injection in order to

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

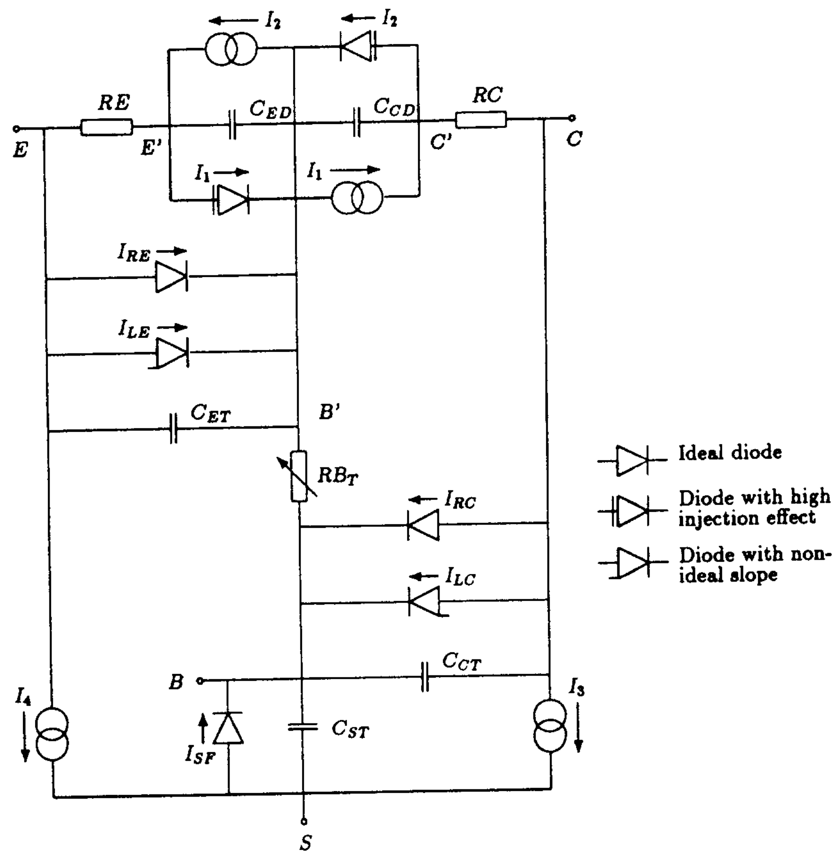


Figure 6.1: *E-GP equivalent circuit diagram.*

compensate for the simplified resistive network in the equivalent circuit.

- only one base resistance is used for both forward and reverse active cases.
- the reverse active case sometimes uses the model parameters from the forward active case e.g. VEAF and IKS.
- the forward and reverse base (and substrate) currents depend on $V_{E'B'}$ and $V_{C'B'}$ respectively, despite the fact that they are not positioned between these nodes in the equivalent circuit. The idea is to save computational time by reducing the number of exponentials to be evaluated.
- *all* of the substrate current is subject to the high injection effect.
- the charge control principle is used to describe the stored charge.
- *all* of the stored base charge is modulated by the collector-base voltage.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

6.3 Modella improvements

This section provides a brief qualitative comparison between Modella and the E-GP model by highlighting the differences between the two models in relation to the equivalent circuits and the formulation of equations and parameters.

When compared to the E-GP equivalent circuit shown in figure 6.1, Modella's equivalent circuit has a more complex resistive network with the associated increase in the number of internal nodes and model parameters. The increase in complexity will increase the C.P.U. time required by the circuit simulator to solve the internal network and may adversely affect convergence properties in a circuit simulation environment. These points are investigated in the next chapter. The benefits arising from this more complex resistive network in Modella are, however, very significant. It allows for the separation of sidewall and bottom components of currents and charges in a physical way thereby reflecting the 2-D nature of the lateral pnp. In turn, the equations and model parameters have a clear physical meaning which results, for example, in more reliable parameter determination. In the E-GP model, compromises relating to an inadequate network are evident. For example, because current crowding is not modelled, three different knee current parameters are used to describe the high injection effects in the forward Ic, the reverse Ie and the substrate currents. Physically, only one parameter is required, but in the E-GP model the influences of current crowding on the current gain characteristics have to be accounted for by altering the modelling of high injection effects in an unphysical way. Also, the symmetry in Modella's equivalent circuit which reflects the physical structure of the lateral pnp is lacking in the E-GP model due to the compromise of using one base resistance for both forward and reverse active cases. This inadequacy is often observed in the inaccurate modelling of the 'on' resistance in the saturated part of the $Ic - Vec$ characteristic.

Many of the equations in the E-GP model are adaptations of npn formulations rather than derivations from lateral pnp device physics. The resulting loss of accuracy can be seen for example by comparing, in the one dimensional case, the collector current formulation with the exact analytical solution[28]. The E-GP model deviates by up to 13% from this solution, compared with a 4% deviation using Modella's formulation. A lack of physical significance in some model parameters is also a consequence of adapting npn formulations and some fitting parameters, like M_T and NHF, restrict the model's flexibility with regard to procmodel formation and geomet-

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

rical scaling.

In theory, therefore, Modella possesses greater accuracy and physical validity in its parameters, equations and equivalent circuit than the E-GP model and it also models some physical effects not previously incorporated in the E-GP model. The following list summarises the major improvements in Modella with respect to the E-GP model:

- same accuracy for reverse active as for forward active operation.
- more accurate modelling of high injection by taking the Webster effect into account i.e. a reduction by a factor of two in the base transit time at high injection. The base transit time in the E-GP model is bias independent due to the use of the charge control principle.
- current crowding under the emitter is modelled.
- the forward Early voltage is dependent on Vcb.
- the forward Early voltage is dependent on Veb.
- more accurate modelling of the depletion capacitances around the diffusion voltage.
- more accurate (and physical) modelling of the diffusion capacitances.
- fall off of f_t and h_{fe} is not only due to high injection but also to ohmic voltage drop across the emitter.
- more accurate modelling of the Vcb dependence of f_t.
- charge storage when the substrate-base junction is forward biased.
- separate saturation current for the substrate-base diode (facilitating h_{fe} modelling under zero base current conditions at high temperatures).

6.4 Results

In order to investigate Modella's performance on real devices, a measurement database of lateral pnp devices from the various bipolar processes within Philips was established. This measurement database incorporated devices with many variations in geometrical dimensions, doping levels, junction depths, etc. Furthermore, both types of isolation (oxide and junction) were included as

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

were the different forms of layout design (stripe, U-shaped and cylindrical). Parameters sets for these devices were extracted using Bipar and extensive comparisons between the performance of Modella and the E-GP model were made. In these comparisons the performance of Modella was seen to be superior to that of the E-GP model, providing excellent agreement with measured device characteristics over wide regions of operation.

In order to illustrate some important aspects of the improved performance, a comparison between the models using the measured characteristics of a typical junction isolated device is shown below. This device from the V80 process was described in section **4.3.3** where it was used to show the parameter determination strategy in practice. The parameter set for Modella used in the comparison below is identical to that shown in table 4.2. The results are shown in figures 6.2 to 6.7 and one can see the improved performance of Modella especially with respect to

- reverse active behaviour (figure 6.3),
- the bias dependence of the cut-off frequency f_t (figure 6.6),
- the bias dependence of the Early effect (figure 6.7).

The improvement in reverse active modelling by Modella is mainly due, in this case, to the improved resistive network and the way in which a *fraction* of the substrate currents are subject to high injection. In the E-GP model *all* of the forward and reverse substrate currents are subject to high injection and therefore fitting the knee current parameter IKS cannot provide a good description of both forward and reverse active cases.

The improved modelling of the f_t characteristic by Modella is due to 1) the inclusion of the Webster effect, 2) the fact that the f_t fall off at high current levels in Modella is a function not only of high injection but also of current crowding and 3) the way in which the stored base charge is divided into two components with differing Vcb dependencies. This improved f_t modelling is significant in itself, but it may also be reasonable to assume that a consequence of an improved f_t prediction will be an improved prediction of the small-signal parameters (e.g. admittance parameters). This will be investigated in the future by comparing measured small-signal parameters with those predicted under a small-signal analysis of Modella in a circuit simulation environment.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

With the E-GP model the Early voltage, V_{eaf}, has a negligible dependence on Veb and is seen to be completely independent of Vcb. MODELLA, on the other hand, shows excellent agreement with measured data due to the physical modelling of both base width modulation by Vcb and current crowding under the emitter. This bias dependence of V_{eaf} is often a critical aspect in lateral pnp compact modeling e.g. in circuit applications involving lateral pnp current sources.

In figure 6.8 a comparison between the h_{fe} predictions of Modella and the E-GP model is shown for another junction isolated device from the L48 process. These measurements were performed at 120°C and the zero base current phenomena is clearly visible in the low current region. As can be seen Modella provides a good description of this effect whereas the E-GP model provides no prediction whatsoever of this increase in gain.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

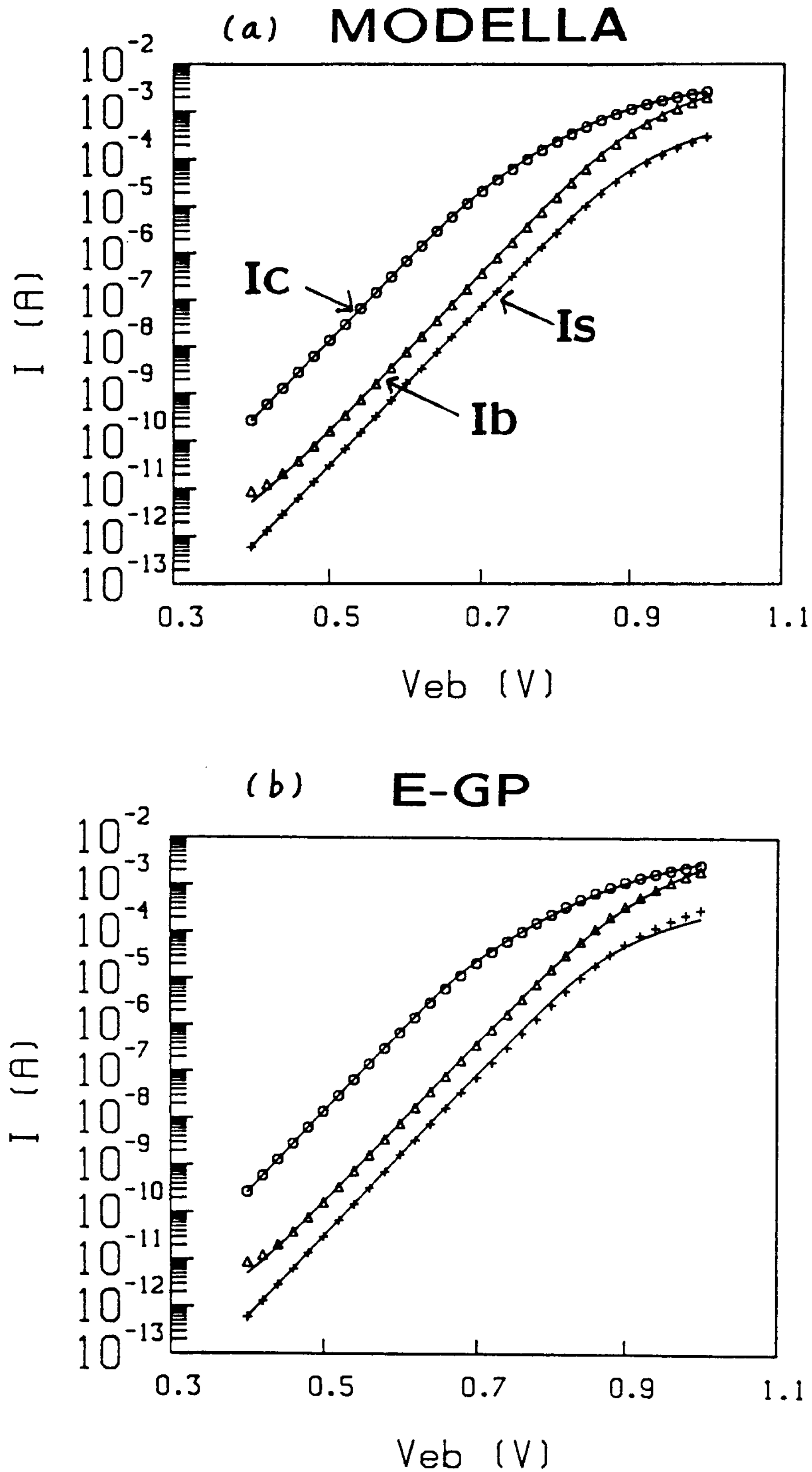


Figure 6.2: *Forward active Gummel plots comparing (a) MODELLA and (b) the E-GP model with measured characteristics.*

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

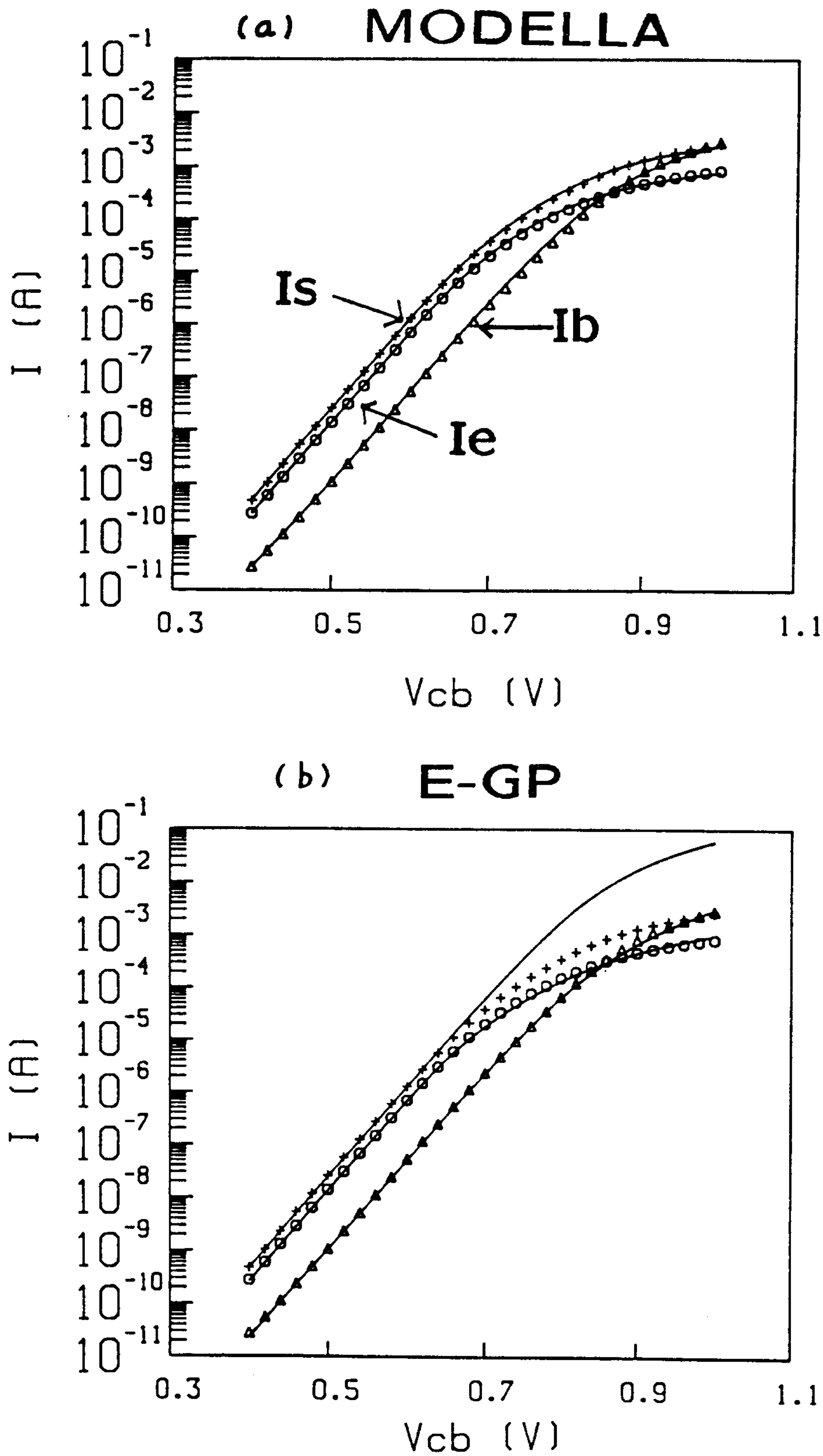


Figure 6.3: *Reverse active Gummel plots comparing (a) MODELLA and (b) the E-GP model with measured characteristics.*

© N.V. Philips' Gloeilampenfabrieken

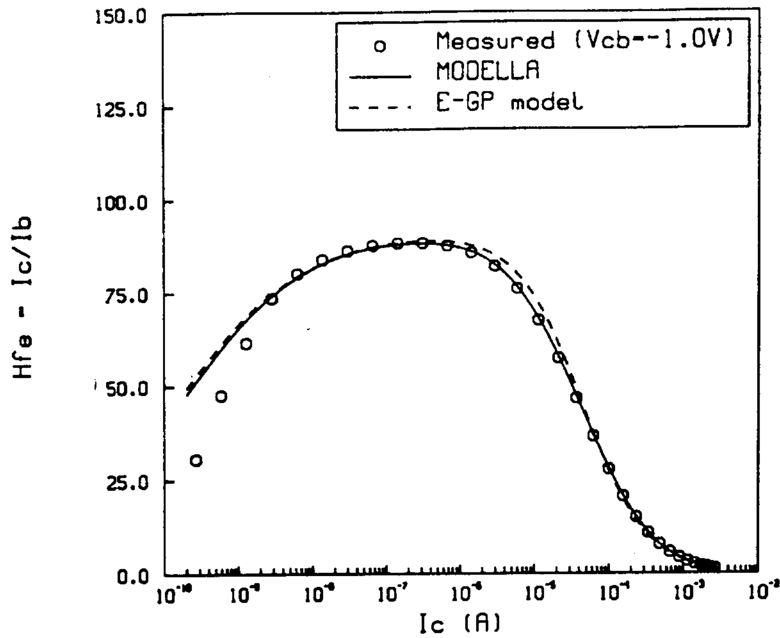


Figure 6.4: *Comparison between MODELLA and the E-GP model predictions of the measured forward current gain* $h_{fe} = \frac{Ic}{Ib}$.

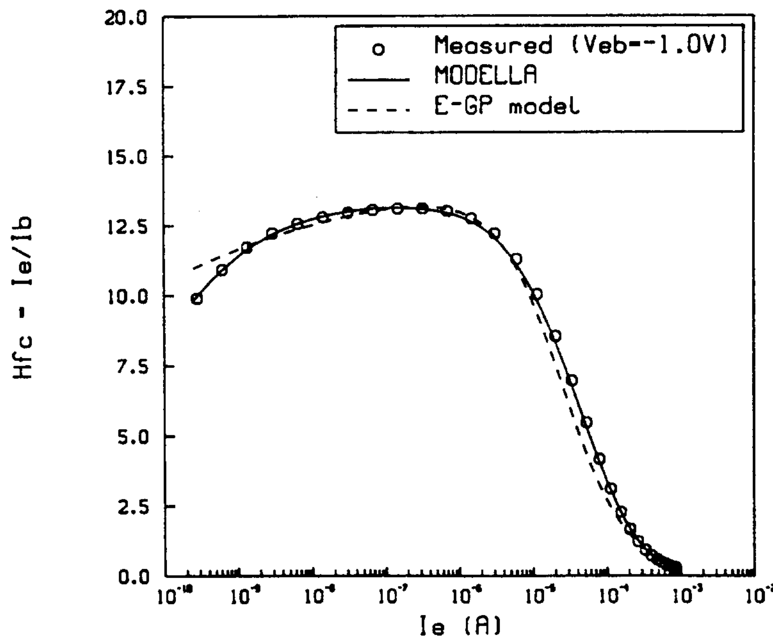


Figure 6.5: *Comparison between MODELLA and the E-GP model predictions of the measured reverse current gain* $h_{fc} = \frac{Ie}{Ib}$.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

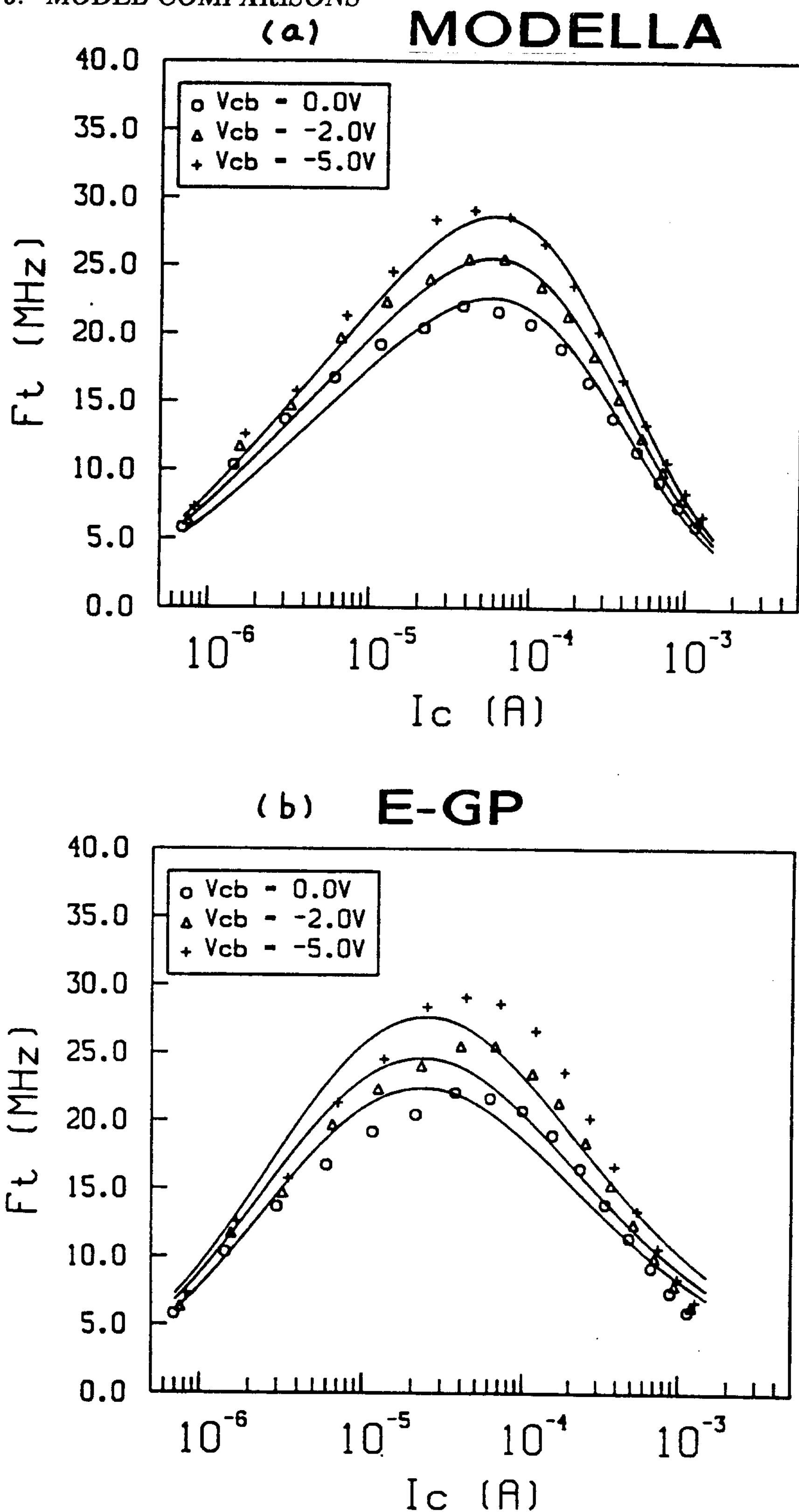


Figure 6.6: *Comparison between (a) MODELLA and (b) E-GP model predictions of the measured cut-off frequency, f_t, as a function of I_c at different V_{cb} values.*

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

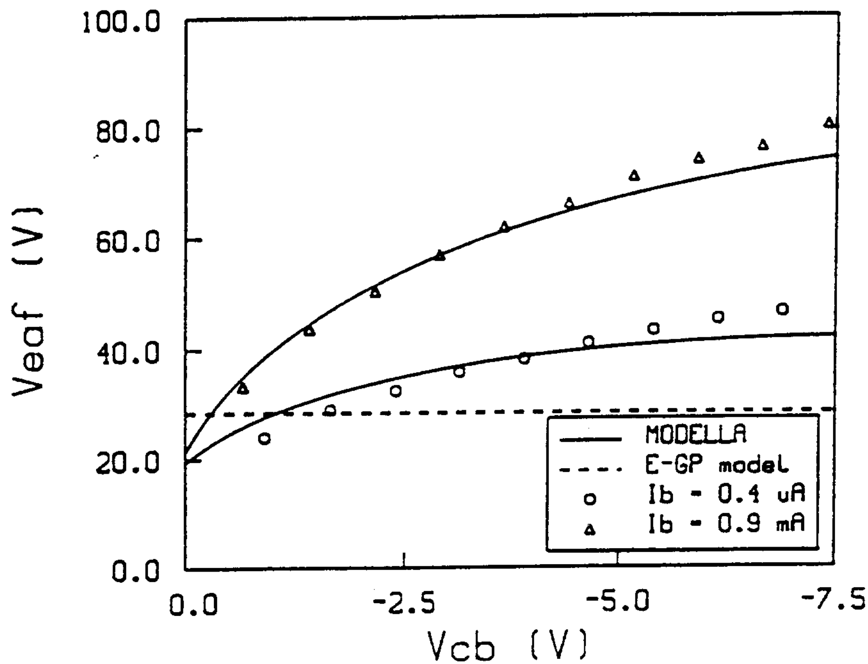


Figure 6.7: *Comparison between MODELLA and the E-GP model predictions of the measured forward Early voltage, V_{eaf} as a function of V_{cb}. The Early voltage is defined as $V_{eaf} = I_c|\frac{dV_{cb}}{dI_c}|_{I_b=const.} - V_{cb}$. The two I_b values correspond to $V_{eb} = 0.7V$ and $V_{eb} = 0.94V$, respectively.*

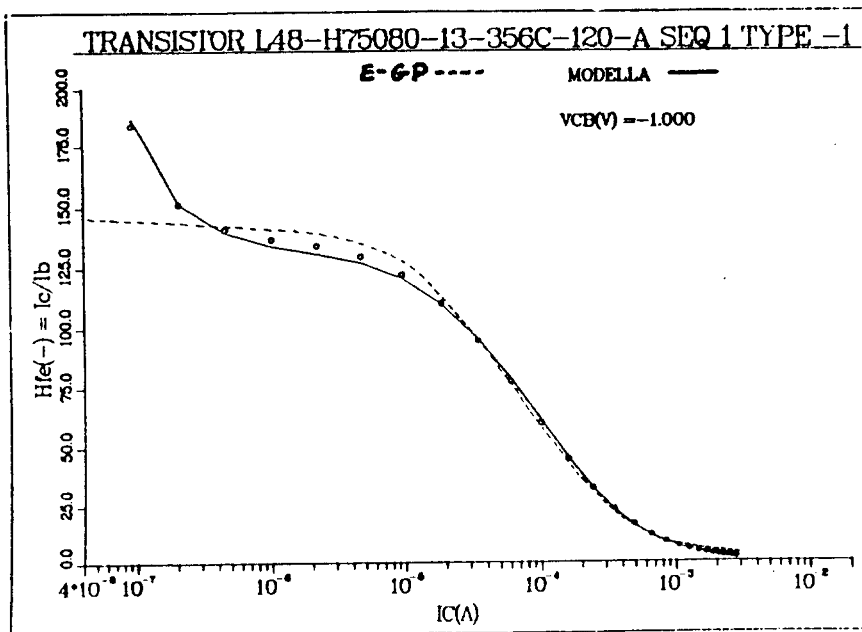


Figure 6.8: *Comparison between MODELLA and the E-GP model predictions of the measured current gain h_{fe} at 120°C of a junction isolated device.*

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

6.5 Conclusions

Despite being the best lateral pnp compact model presently available, the extended Gummel-Poon model is visibly lacking in many respects. The modelling of certain physical effects (e.g. current crowding) is absent from the model and the equations describing the remaining physical effects have been adapted from one-dimensional derivations on vertical npn transistors. As described in section **6.3**, it is obvious from a theoretical point of view that Modella's physical formulation provides a more accurate and comprehensive description of lateral pnp device physics.

The most important test of a compact model, though, is its ability to provide good predictions of measured device characteristics. In numerous comparisons between Modella and the E-GP model using measurements, the theoretical improvements did indeed translate into an improved performance in practice. This improvement was especially significant in the modelling of reverse active behaviour and the bias dependence of both the Early voltage and the f_t characteristics. All these aspects are crucial to designers requiring accurate simulation results in circuits containing lateral pnps.

It may be concluded that Modella constitutes a considerable improvement in the compact modelling of lateral pnp transistors. Compared to the E-GP model it provides greater modelling accuracy. Furthermore, because it is a *physical* model it possesses advantages in terms of parameter determination, forecasting abilities, geometrical scaling, statistical modelling and it facilitates an intuitive understanding of device behaviour due to its close link with device physics. Given these favourable conclusions, a final crucial aspect in the development of Modella requires investigation - the circuit simulation environment. This is the subject of the next chapter.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 7

Circuit simulation environment

7.1 Introduction

The ultimate goal of this project is the acceptance of Modella by the design community for use in circuit simulations involving lateral pnp transistors. In demonstrating Modella's advantages over the E-GP model in the last chapter this goal would appear to have been achieved. However, the suitability of a compact model to the circuit simulation environment is never assured until the model has been correctly implemented into a circuit simulation program and its behaviour thoroughly tested in this environment. This chapter investigates Modella's performance in relation to this critical aspect of model development.

The procedure required for this investigation is as follows:

- The implementation in a circuit simulator necessitates the solution of Modella's equivalent circuit by a general circuit solver. It must firstly be shown that Modella's network and equations are suited to this general solution approach so that convergence can be obtained for single transistor circuits.
- Verification of these single transistor DC, AC (small-signal) and Transient (large-signal) analyses solutions must then be obtained. Suitable *benchmarks* must be found in order to test each of these aspects of model implementation. This verification is vital to ensure a 'bug' free implementation which could otherwise result in inaccurate results or poor convergence behaviour. Both of these scenarios would lead to a lack of confidence in Modella on the part of circuit designers - an initial impression which could prove very

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

difficult to disspell.

- Larger test circuits will then be supplied by designers in order to examine Modella's convergence behaviour and this will be compared to the convergence behaviour of the E-GP model. The main emphasis here will be on the ability to converge but the speed of convergence will also be evaluated to provide designers with an estimate of Modella's C.P.U. intensity.

In the next section the implementation procedure is outlined and this is followed in section **7.3** by a description of how this implementation was verified. Results are presented in section **7.4** and the chapter ends with some conclusions on the circuit simulation environment.

7.2 Implementation procedure

Modella was implemented into the circuit simulation program Panacea [3]. Unfortunately no *link* facility for user-defined compact models existed in Panacea, so a special version of the entire Panacea program was required for the implementation. The choice of machine on which to run this Panacea version was an Apollo workstation. This choice was determined partly by the graphics capabilities but mostly by the availability of sophisticated software management tools (e.g. DSEE [45]) which were necessary due to the large amount of code (> 100,000 lines) and the distributed nature of coding alterations. Initially this version of Panacea ran in batch mode on an Apollo DN 10000 but was then ported to an Apollo DN 4000 which, despite being slower, facilitated interactive debugging.

The implementation was basically divided into two parts - the description of Modella and the interface definition. The description of Modella involved the following

- translating all the Fortran functional subroutines (describing currents, charges, etc.) of Bipar into 'C' functions. Extensive use was made of *structures* (i.e. a collection of variables grouped together under a single name) facilitating efficient and convenient data handling.
- the circuit solver in Panacea required derivatives of charges and currents with respect to the five independent voltages $Ve2b1, Ve1b, Vc2b2, Vc1b$ and Vsb and not two voltages as was the case with the solution procedure in Bipar (i.e. $Ve2b1$ and $Vc2b2$ - see Appendix B).

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

- initial estimates of internal junction voltages were required. The values which were seen to yield the best results in terms of convergence were $Ve2b1 = 0.75V, Ve1b = 0.7V, Vsb = -1.0V, Vb2b = -0.1V, Vb1b = 0.1V$ with all other internal voltages set to zero.

The interface definition[1] involved supplying extra routines required for implementation into a circuit simulator. These ranged from simple internal voltage limiting functions to routines which supplied the complete admittance matrix (specific for the topology of Modella) required by the circuit solver.

7.3 Implementation verification

Having implemented Modella into Panacea, the next stage was to *verify* that this implementation was correct and 'bug' free. This verification was not as straightforward as one might think because this was the only implementation of Modella which performed certain modes of analyses. The main problem, therefore, was the choice of suitable benchmarks and this section describes the approaches taken with each of the three analyses modes, namely D.C, A.C. (small-signal) and Transient (large-signal) analyses.

- **D.C. analysis** : this was the most straightforward to test since a direct comparison could be made with Bipar. Identical parameter sets were supplied to Panacea D.C. analysis and to Bipar and a comparison of forward and reverse Gummel plots and Ic-Vec characteristics was made. Agreement between currents from Panacea and Bipar is now within 0.1% for D.C. analysis.

- **A.C. analysis** : in Panacea the A.C. analysis calculates the circuit response in the frequency domain. It is a small-signal analysis; the circuit is treated as a linear circuit. Non-linear circuits are linearised at their D.C. operating point (the operating point being calculated prior to the A.C. analysis by an implicit D.C. analysis). This linearisation makes use of the derivatives of the currents and charges with respect to the independent internal voltages in Modella. The A.C. output consists of small-signal parameters like admittance parameters, hybrid parameters, etc. from which real A.C. figures of merit like

[1]This aspect of the model implementation procedure was done in co-operation with the Panacea support team. All other aspects of the implementation were done by the author in order to ensure a speedy implementation and to minimise implementation errors which could affect accuracy and convergence behaviour in initial releases of Modella in Panacea.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

the cut-off frequency, f_t, can be calculated. Because of the importance of f_t as a figure of merit and because a great deal of information is actually contained in this value, it became an obvious candidate for verification purposes.

As described in section **2.3.4**, f_t can be determined by extrapolating the high frequency 6dB/octave fall off of the small-signal current gain, $\hat{h}_{fe}$. The existence of a 6dB/octave fall off region can be derived directly the current continuity equations by using the quasi-static and charge control approximations. If $\hat{h}_{fe}(0)$ is less than about ten however, no 6dB/octave region can be found and the definition fails. Another definition which does not rely on this 6dB/octave fall off region is the Gummel definition[46]:

$$f_t^G = \frac{f}{Im(\hat{h}_{fe}(f)^{-1})}$$

Here Im denotes the *imaginary* part and f is a frequency in the constant part of the $\hat{h}_{fe}$ versus frequency curve (i.e. the low frequency range).[2] This definition is much simpler to use with a circuit simulator because no curve extrapolation is needed and indeed a 6dB/octave fall off is not even required.

The f_t is inversely proportional to the total transit time in a device and is given by $f_t = \frac{1}{2\pi\tau_{tot}}$. In the quasi-static approach the transit time can be determined from the steady state solution by taking the derivative of the charge with respect to the current i.e. $\tau = \frac{dQ}{dI_C}|_{dV_{ec}=0}$. This is the approach which is taken with Bipar and it allows an approximation of an A.C. figure of merit to be made by what is essentially a steady state D.C. analysis. The assumptions made with charge control also apply to this definition and therefore Bipar will predict an f_t which is based on a first order system (i.e. a system which predicts a 6dB/octave fall off in $\hat{h}_{fe}$ as described above). It is to be expected therefore that differences will arise in values predicted by Bipar for f_t and those predicted by Panacea for f_t^G. These differences will occur when the small-signal current gain, $\hat{h}_{fe}$, is not much greater than one or when RC parasitics play a role in the Panacea simulation.

[2]To show that this definition agrees with the first order definition (i.e. a definition which assumes a 6dB/octave fall off in $\hat{h}_{fe}$), consider the first order equation for $\hat{h}_{fe}$ i.e. $\hat{h}_{fe} = \frac{\beta_0}{1+j\omega\beta_0\tau}$. Using this definition $Im(\hat{h}_{fe}(f)^{-1}) = \omega\tau$ and therefore $f_t^G = \frac{1}{2\pi\tau}$ which is the normal f_t formulation.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

The testing of the implementation with respect to A.C. analysis was based largely on this comparison between the Bipar f_t and the Panacea f_t^G. The agreement between these values was exactly as expected. They agreed to within about 0.5% over the whole D.C. operating range except at high forward bias. In this region the small-signal gain, $\hat{h}_{fe}$, became close to unity leading for example to a difference of about 5% in the values at a forward bias of 1V. Manipulation of Modella's parameters to increase $\hat{h}_{fe}$ showed that perfect agreement between the two definitions occurred when $\hat{h}_{fe} >> 1$.

A qualitative test on y- and h-parameters was made by comparing the frequency dependence of the magnitude and phase predicted by Modella to that of the E-GP model. A comparison with actual small-signal measurements should be made in the future but this relates to model verification and not to implementation testing.

- **Transient analysis** : In a circuit simulator the transient analysis calculates the circuit response in the time domain. Unlike an A.C. analysis, a transient analysis makes use of the actual values of the charges in Modella and not their analytical derivatives. Quantities like $i = \frac{dQ}{dt}$ are calculated by dividing the input voltage/current (i.e. V(t) or I(t)) into small discrete time steps and performing a D.C. analysis at each step. The derivative of charge with respect to time is therefore obtained by simply dividing the charge difference by the time step size.

Verification of correct implementation with regard to transient analysis cannot really be done in a direct manner since no obvious benchmark exists. Instead the following procedure was carried out. A transient analysis was performed with a time-varying input voltage applied to the base of the lateral pnp. The amplitude of this input signal was 1μV. After an appropriate delay (1 milli sec) to ensure no 'switch-on' transients affected the output current, a Fourier analysis was done over several periods of the output current (i_c). This analysis yielded amplitude and phase values for the first and higher order harmonics. Because the amplitude of the input signal is so small, the circuit behaves as a linear circuit. This is verified by observing that the second and higher order harmonics in the output signal, which normally cause distortion of the sinusoid, are in fact negligible. The principal of this approach therefore is that the amplitude and phase of this perfectly

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

sinusoidal output should be the same as that obtained from an actual small-signal (A.C.) analysis (assuming of course that the periodic time of the transient input is equivalent to the frequency used in the A.C. analysis).

Tests using this approach to verify correct transient analysis solutions have shown excellent agreement between i_c obtained from the A.C. analysis and i_c derived from this 'small-signal' transient analysis. This agreement was obtained over the whole D.C. operating range.

Having completed all these tests it can be stated with a reasonable degree of confidence, that a correct model implementation has been verified. Furthermore, the test files used for the verification of D.C., A.C. and transient modes can automatically form the basis of a test implementation kit or set of benchmarks to verify future implementations of Modella.

7.4 Results

Assuming correct implementation in Panacea, the behaviour of Modella in the circuit simulation environment could then be investigated. This section summarises the findings of this investigation.

Firstly, the convergence behaviour of Modella for single transistor circuits was investigated and no convergence difficulties were observed for all modes of operation. When compared to the E-GP model for these single transistor circuits, Modella required an equivalent or smaller number of iterations to obtain a solution. The C.P.U. time per iteration was only a factor of 1.5 slower than that of the E-GP model which compares with a factor of two between Mextram and the Gummel-Poon based vertical transistor compact model. The reason for Modella's speed is that, despite the apparent complexity (six internal nodes), its equations are quite straightforward and no extra iterative internal solution procedures are required (as is the case with Mextram).

The next aspect to be investigated was Modella's performance when used in larger circuits with many devices. Example circuits, varying in size from two transistors to thirty transistors

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

(the majority of transistors in each circuit were lateral pnps), were supplied by designers. These tests investigated Modella's convergence behaviour for both the straightforward cases and the more problematic cases because it was known that the E-GP model exhibited convergence difficulties with some of these circuits. The results of these tests are shown in table 7.1 where the performance and convergence behaviour of Modella and the E-GP model are compared for seven circuits. The number of iterations required for each circuit was of course a function of the analysis carried out. In two examples, the maximum allowed number of iterations was increased in order to increase the possibility of convergence for both Modella and E-GP. The overall result of these initial investigations into Modella's convergence behaviour was extremely encouraging. *All* examples converged when Modella was used, whereas with the E-GP model two examples did not converge and another gave an unreliable result. The relative C.P.U. time between Modella and E-GP (in those cases where convergence was obtained using E-GP) varied between 2.8 and 1.2 depending on the circuit.

In addition to the above quantitative comparisons between Modella and the E-GP model, qualitative comparisons were also made between Modella and the hybrid E-GP models. These hybrid or user-defined models incorporate variations on the E-GP model in order to compensate specific inadequacies of E-GP. The hybrid models TPLHEA[11] and TPLHLEAK[12], for example, exhibited qualitative behaviour similar to Modella in terms of the Vcb dependence of the Early effect (TPLHEA) and the increase in current gain at high temperatures (TPLHLEAK). Modella, however, has the advantage that it incorporates, into one compact model, effects which at present require the use of different hybrid models. Furthermore, these hybrid models are known to possess convergence problems which have not been observed in Modella.

7.5 Conclusions

Modella has been implemented into the circuit simulation program Panacea and correct implementation has been verified for all modes of analyses. Because of the emphasis which was placed on verification, it is expected that Modella will be introduced into the design community with a minimum of difficulty. This verification process has also produced a set of benchmarks which will be useful when verifying future implementations of Modella in different circuit simulation programs.

© N.V. Philips' Gloeilampenfabrieken

Circuit	No. of iterations	CPU (secs)	Rel. CPU
Current mirror (2 transistors)			
.......Modella	77	103	1
.......E-GP	*	*	*
V - I converter (4 transistors)			
.......Modella	45	2.5	1.7
.......E-GP	45	1.5	1
Comparator (9 transistors)			
.......Modella	80	8.6	2.8
.......E-GP ???	46	3.1	1
Level detector (13 transistors)			
.......Modella	101	20.2	2
.......E-GP	108	10.1	1
Bandgap reference (13 transistors)			
.......Modella	53	7.4	1.3
.......E-GP	67	5.6	1
Voltage follower (17 transistors)			
.......Modella	322	71.2	1.2
.......E-GP	348	59.8	1
Bandgap + follower (30 transistors)			
.......Modella	330	280.4	1
.......E-GP	*	*	*

Table 7.1: *Comparison between Modella and the E-GP model in terms of convergence behaviour and speed using typical circuits incorporating lateral pnp devices. A * means that convergence could not be obtained and ??? means that the solution obtained was indicated by Panacea to be unreliable.*

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

With regard to the suitability to circuit simulation, it may be concluded that Modella is highly suited to this environment. Its speed of evaluation for single transistor circuits is only 1.5 times slower that that of the E-GP model and its convergence behaviour for typical circuits was seen, in initial tests, to be superior to that of the simpler E-GP model! Therefore, despite having six internal nodes, Modella's straightforward approach to the formulation of model equations provides both a highly accurate prediction of device behaviour while at the same time maintaining excellent convergence behaviour with acceptable evaluation times.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Chapter 8

Conclusions

In the field of bipolar transistor compact modelling the emphasis has always been placed on vertical npn compact models, with lateral pnp compact models receiving relatively little attention. Present day lateral pnp compact models therefore lack a physical basis and are, at best, semi-empirical formulations. Consequently, the integration of these models into process blocks has been problematic and the reliability of circuit simulations involving lateral pnp devices has been quite inadequate. The work described in this thesis provides a solution to these problems by developing the first truly *physics* based lateral pnp compact model.

The research on lateral pnp device physics, outlined in chapter 2, provided the basic knowledge required to physically model present day lateral pnp devices. This theory of lateral pnp device physics was verified by the agreement obtained between scaled one-dimensional analytical expressions and the results of two-dimensional numerical device simulations of typical device structures. Furthermore, this work led to the conclusion that when ohmic effects were excluded, the behaviour of lateral pnp devices could be modelled by means of one-dimensional expressions. This principle forms the basis of Modella, the new compact model; the basic equations in Modella are one-dimensional derivations and the two-dimensional effects are incorporated by modifying and extending the basic equations to include the effects of series resistances, most notably current crowding. The incorporation of this current crowding effect is crucial for a physical model as it influences the current gain, h_{fe}, the cut-off frequency, f_t, and the Early voltage, V_{eaf}. The result of this modelling approach is a highly physical compact model with six internal nodes in the equivalent circuit and forty two model parameters.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

An efficient and reliable parameter determination strategy was developed in order to extract physically plausible values for these model parameters. Physical intuition and sensitivity analysis played important roles in the choice of procedures to determine parameters. Aspects of the strategy which are of a model specific nature include the following:

- the method of Ning and Tang[44] was shown to be highly suitable in the determination of certain emitter and base resistance parameters.
- the Early voltage parameters were determined using a method which involved redefining the Early voltage itself.
- the determination of certain charge model parameters availed of all three f_t characteristics to make maximum use of the relative sensitivity of the parameter values to the collector-base voltage variation.

This final strategy to determine Modella's parameters represents a highly tuned solution to a rather difficult problem and can be described as highly automated and reliable, yielding physically plausible values.

Because it is desirable in a circuit simulation environment to use Modella at temperatures other than the specific reference temperature for which the parameter set is valid, a physical set of temperature scaling rules for the model parameters was derived. Because the temperature scaling rules themselves contain temperature parameters, a procedure was developed to determine these temperature parameters and in turn to verify the entire set of scaling rules. Initial results comparing measurements at different temperatures with simulated behaviour obtained from temperature scaled parameter sets were very encouraging. This description of the temperature dependence (and in some cases independence) of the model parameters has provided an excellent basis from which to undertake process specific temperature dependence verification.

In order to demonstrate the superiority of Modella, comparisons were made to the best presently available lateral pnp compact model (E-GP), using measured device characteristics. Modella's performance was significantly better especially with respect to reverse active behaviour and the bias dependencies of the cut-off frequency[1] and the Early voltage. In addition, Modella

[1]Recently, Modella was compared with the E-GP model using measured values of the small-signal admittance parameters as a function of frequency and again Modella was much more accurate.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

incorporated effects presently only available in hybrid versions of E-GP, thereby providing *one* model for *all* applications.

The investigation described in chapter 7 verified the correct implementation of Modella in the circuit simulator Panacea (thereby providing a set of benchmarks to check future implementations of Modella) and, furthermore, demonstrated that Modella was highly suited to the circuit simulation environment. In initial tests comparing Modella to the E-GP model, Modella required a negligible increase in evaluation time and, despite its apparent complexity, exhibited better convergence behaviour.

The work described in this thesis has charted the development of Modella from a virtually untested prototype to the stage where it is already available to designers and constitutes a realistic and significantly better alternative to present day lateral pnp compact models. In global terms, the contribution of this research to the field of lateral pnp compact modelling is even more significant. This is a result of the fact that SPICE, the most widely used circuit simulation program in the world, only offers designers the SPICE Gummel-Poon model for lateral pnp transistors. This model is inferior to the E-GP model in many respects:

- it gives no prediction for the forward (or reverse) substrate current, Is; a fatal omission for example with junction isolated devices.
- the substrate is actually connected to the collector and not the base as, of course, it should be for a lateral pnp device. This leads to an inconsistent charge model and poor f_t and small-signal behaviour modelling.

By combining three of these SPICE-GP models to form a sub-circuit, it is possible to obtain a prediction for the substrate current, but this is inefficient and still not as accurate as Modella. It is clear, thus, that the proven superiority of Modella above the E-GP model will be even greater when it is compared to this 'standard' SPICE pnp model. The distinct possibility therefore exists that Modella may follow the path being taken by Mextram and be implemented into the standard SPICE release thereby constituting a significant improvement over Gummel-Poon models.

This possibility of Modella becoming a world standard is enhanced by the fact that the bipolar community is realising the advantages of using *physical* models in more integrated sim-

© N.V. Philips' Gloeilampenfabrieken

ulation systems based on a process block like approach. The scaling of *physical* parameters as a function of geometry and process quantities in a process block is inherently more reliable than *empirical* parameters. Furthermore, this approach to the generation of parameters is highly efficient compared to parameter optimisation from measurements. This approach also facilitates reliable device/circuit *optimisation* to be carried out more efficiently in the simulation environment compared to using the costly procedure of trial and error production. At the circuit design stage, physical model parameters can also function as diagnostic tools since unacceptable parameter values can be traced to the offending process quantity.

In conclusion, this project has been highly successful in achieving the aims outlined in chapter 1 and has resulted in a new compact model which is poised to become the most accepted lateral pnp compact model within Philips and possibly also in the field of bipolar compact modelling.

Future work

Although the development of Modella has reached quite an advanced stage, certain enhancements would be desirable before describing Modella as a fully mature compact model. Indeed, it is often said that the development of a compact model is *never* really completed. Even so, it is necessary to set priorities and the following areas are those which should receive attention in future development work:

- the modelling of excess phase shift at high frequencies should be incorporated into Modella as this effect can cause circuit instability.
- a demonstration of the systematic scaling of Modella's parameters as a function of geometry and process quantities would prove the inherent advantages of a physical rather than empirical formulation.
- verification of the noise model (which is easily derived from the equivalent circuit by incorporating thermal and shot noise sources) by comparison with noise measurements.
- further investigation of Modella's small signal behaviour compared to the E-GP model using measured device behaviour.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Bibliography

[1] F.G. O'Hara, J.J.H. van den Biesen, H.C. de Graaff, J.B. Foley ; *A new physical compact model for lateral pnp transistors*, Proc. I.E.E.E. Bipolar Circuits and Technology Meeting (BCTM), Minneapolis, p102, Sept. 1990.

[2] Philips Components ; *ESPICE Reference manual*, Release 03.00.00, 1989.

[3] Corporate CAD centre (Philips) ; *PANACEA Users manual*, Release 1, 1988.

[4] G.A.M. Hurkx, F.G. O'Hara ; *A comparative study of lateral pnp concepts*, Philips Technical Note Nr. 350/87.

[5] K. Hart, A. Slob ; *Integrated Injection Logic: A new approach to LSI* I.E.E.E. J. Solid State Circuits, SC-7, 346, 1972.

[6] J.J. Ebers, J.L. Moll ; *Large Signal Behaviour of Junction Transistors*, Proc. I.R.E. 42, 1761, 1954.

[7] H.K. Gummel, H.C. Poon ; *An integral charge control model for bipolar transistors*, Bell Systems Technical Journal 49, 827, 1970.

[8] H. de Graaff, W. Kloosterman ; *New Formulation of the Current and Charge Relations in Bipolar Transistor Modelling for CACD Purposes*, I.E.E.E. Trans. on Electron Devices, Vol.ED-32, No.11, Nov. 1985.

[9] H.C. de Graaff, W. Kloosterman, T.N. Jansen ; *Compact Bipolar Transistor Model for CACD, with accurate description of collector behaviour*, Ext. Abstr. 18th Conf. Solid St. Devices and Materials, Tokyo, 287, 1986.

[10] C. van Velthooven ; *Model equations for TPL level 301 and 401*, (Philips) RNR 23/84-41-II-CvV/MB, Nijmegen, 1984.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

[11] J.O. Voorman ; *High-Q interpolation, Early effect and computer modelling and delay-cell design formulas for gyrator circuit design*, Philips Technical Note Nr. 053/86.

[12] M. Stoutjesdijk ; *Base-substrate current effects for the lateral pnp model TPL-301 at different temperatures*, (Philips) RNR 23/86-91-IV-MS, Nijmegen 1986.

[13] J.M. Early ; *Effects of space-charge widening in junction transistors*, Proc. I.R.E. 40, 1401, 1952.

[14] J.J.H. van den Biesen ; *Physical working and compact modelling of lateral pnp transistors*, Nat. Lab. (Philips) Report Nr. 6131.

[15] H.C. de Graaff, F.M. Klaassen ; *Compact transistor modelling for circuit design*, Springer-Verlag, 1990.

[16] H.C. Lin, T.B. Tan, C.Y. Chang, B. van der Leest, N. Formigoni ; *Lateral complementary transistor structure for the simultaneous fabrication of functional blocks*, Proc. I.E.E.E. 52, 1491, 1964.

[17] S. Chou ; *An investigation of lateral transistors - DC characteristics*, Solid St. Electr. 14, p811, 1971.

[18] I. Kidron ; *Integrated circuit model for lateral pnp transistors including isolation junction interaction*, Int. J. Electr. 31, p421, 1971.

[19] S. Chou ; *Small-signal characteristics of lateral transistors*, Solid St. Electr. 15, p27, 1972.

[20] H.H. Berger, U. Drekmann ; *The lateral pnp transistor - A practical investigation of the d.c. characteristics*, I.E.E.E. Trans. Electr. Dev. ED-26, p1038, 1979.

[21] K-S. Seo, C-K. Kim ; *On the geometrical factor of lateral pnp transistors*, I.E.E.E. Trans. Electr. Dev. ED-27, p295, Jan. 1980.

[22] A.A. Eltoukhy, D.J. Roulston ; *A complete analytic model for the base and collector current in lateral pnp transistors*, Solid St. Electr. 27, p69, 1984.

[23] M.Schneider and J.Bodinaud ; *A new formulation of the Early effect in lateral pnp transistors*, Physica 129B, p.327, 1985.

[24] L. Deferm, S. Decoutere, C. Claeys, G. Declerck ; *The influence of the interface trap density on the performance of bipolar devices*, Proc. I.E.E.E. BCTM, Minneapolis, p136, Sept. '89.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

[25] T.B. Hook, M.E. Johnson, A.V. Ferris-Prabhu ; *Electron trapping, nitride conduction and forward gain instability in a lateral pnp device*, I.E.E.E. Trans. Electr. Dev. ED-37, p755, Mar. 1990.

[26] Corporate CAD centre ; *Curry user manual*, Version 7, 1987.

[27] F.G. O'Hara, G.M.M. Majoor ; *Manual for the bipolar transistor analysis program Trap*, Nat. Lab. Report Nr. 6221, 1987.

[28] A. van der Ziel ; *Solid state physical electronics*, Prentice-Hall, Englewood Cliffs, eqn. 16.71, 1968.

[29] W.M. Webster ; Proc. I.R.E., 42, p914, 1954.

[30] M. Stoutjesdijk ; *Private communication*, Philips Nijmegen.

[31] J. Lindmayer, C.Y. Wrigley ; *Fundamentals of semiconductor devices*, Van Nostrand, Princeton, NJ 1965.

[32] W. Shockley, W.T. Read ; *Statistics of the recombination of holes and electrons*, Phy. Rev. 87, p835, 1952.
R.N. Hall ; *Electron-Hole recombination in Germanium*, Phy. Rev 87, p387, 1952.

[33] J.C. Maxwell ; *A treatise on electricity and magnetism*, 3rd Edn, Dover, New York, 1954.

[34] G.A.M. Hurkx ; *Private communication.*

[35] M.Y. Ghannan, R.P. Mertens ; *Surface recombination current with a nonideality factor greater than 2*, Electr. Dev. Letters 10, p242, June 1989.

[36] J.D. Last, D.W. Lucas, G.W. Sumerling ; *A numerical analysis of the dc performance of small geometry lateral transistors*, Solid St. Electr. 17, p1111, 1974.

[37] J.J.H. van den Biesen ; *A simple regional analysis of transit times in bipolar transistors*, Solid St. Electr. 29, p529, 1986.

[38] C.T. Kirk ; *A theory of transistor cut-off frequency (f_t) fall-off at high current densities*, I.R.E. Trans. Electr. Dev. ED-9, p164, 1962.

[39] H.C. de Graaff ; *Electrical behaviour of lightly doped collectors in bipolar transistors*, Thesis, University of Technology, Eindhoven, 1975.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

[40] J.J.H. van den Biesen ; *P-N junction capacitances Part 1 : The depletion capacitance*, Philips J. Res. 40, p88, 1985.

[41] H.C. Poon, H.K. Gummel ; *Modelling of the emitter capacitance*, Proc. I.E.E.E. 57, p2181, 1969.

[42] T. Jansen, M. Versleijen ; *Biplot program*, Nat. Lab., 1988.

[43] W.J. Kloosterman ; *Userguide to BIPAR*, Nat. Lab. Technical Note Nr. 245/83.

[44] T.H. Ning, D.D. Tang ; *Method for determining the emitter and base series resistances of bipolar transistors*, I.E.E.E. Trans. Electr. Dev. ED-31, p409, 1984.

[45] Apollo Domain ; *Engineering in the DSEE (Domain Software Engineering Environment) environment*, 1986.

[46] H.K. Gummel ; Proc. I.E.E.E. 57, p2159, 1969.

[47] E. Kreyszig ; *Advanced engineering mathematics*, 5th Edn, Wiley, 1983.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Appendix A

Model equations and parameters

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Model constants

$$\begin{aligned} Temp &= 273.16 + Tnom + Dta \\ q &= 1.602 \times 10^{-19}\ \text{C} \\ k/q &= 0.86171 \times 10^{-4}\ \text{V/K} \\ \delta &= 0.01 \\ Tsd &= 1.0 \times 10^{-6}\ \text{s (fixed transit time for Qsd)} \\ Vd &= 0.6\ \text{V (the base diffusion voltage)} \end{aligned}$$

The default reference temperature $\mathbf{T_{ref}}$ for parameter determination is 25 °C.

Early factors

The Early factors for the components of the main current Ip are derived from the variation of the depletion widths in the base relative to the base width itself.

Early factor of the lateral current components

$$Flat = 1 - \frac{\sqrt[4]{\left(1 - \frac{Ve1b}{Vd}\right)^2 + \delta}}{1 + \frac{\mathbf{E_{arl}}}{2Vd}} - \frac{\sqrt[4]{\left(1 - \frac{Vc1b}{Vd}\right)^2 + \delta}}{1 + \frac{\mathbf{E_{afl}}}{2Vd}} \tag{A.1}$$

Early factor of the forward vertical current component

$$Ffver = 1 - \frac{\sqrt[4]{\left(1 - \frac{Ve2b1}{Vd}\right)^2 + \delta}}{1 + \frac{\mathbf{E_{arv}}}{2Vd}} - \frac{\sqrt[4]{\left(1 - \frac{Vc1b}{Vd}\right)^2 + \delta}}{1 + \frac{\mathbf{E_{afv}}}{2Vd}} \tag{A.2}$$

Early factor of the reverse vertical current component

$$Frver = 1 - \frac{\sqrt[4]{\left(1 - \frac{Ve1b}{Vd}\right)^2 + \delta}}{1 + \frac{\mathbf{E_{arv}}}{2Vd}} - \frac{\sqrt[4]{\left(1 - \frac{Vc2b2}{Vd}\right)^2 + \delta}}{1 + \frac{\mathbf{E_{afv}}}{2Vd}} \tag{A.3}$$

Model parameters : $\mathbf{E_{afl}}$, $\mathbf{E_{afv}}$, $\mathbf{E_{arl}}$, $\mathbf{E_{arv}}$.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Ideal diodes

$$Vt = \frac{k \times Temp}{q}$$

The ideal diode equations are as follows...

$$If1 = \mathbf{I_s} \times \left(e^{Ve1b/Vt} - 1\right) \quad (A.4)$$

$$If2 = \mathbf{I_s} \times \left(e^{Ve2b1/Vt} - 1\right) \quad (A.5)$$

$$Ir1 = \mathbf{I_s} \times \left(e^{Vc1b/Vt} - 1\right) \quad (A.6)$$

$$Ir2 = \mathbf{I_s} \times \left(e^{Vc2b2/Vt} - 1\right) \quad (A.7)$$

Model parameter : $\mathbf{I_s}$.

The main current I_p

$$I_p = I_{flat} + I_{fver} - I_{rlat} - I_{rver} \quad (A.8)$$

Forward currents I_{flat} and I_{fver}

The main forward current is separated into lateral and vertical components originating from the emitter-base junction sidewall and bottom respectively. These formulations include Early and high injection effects and because the two currents depend on different internal emitter-base junction voltages, emitter current crowding is also modelled.

The lateral forward current component $Iflat$

$$I_{flat} = \left(\frac{4 \times (1 - \mathbf{X_{ifv}}) \times If1}{3 + \sqrt{1 + 16 \times \frac{If1}{\mathbf{I_k}}}}\right) / \, Flat \quad (A.9)$$

The vertical forward current component $Ifver$

$$I_{fver} = \left(\frac{4 \times \mathbf{X_{ifv}} \times If2}{3 + \sqrt{1 + 16 \times \frac{If2}{\mathbf{I_k}}}}\right) / \, Ffver \quad (A.10)$$

Model parameters : $\mathbf{X_{ifv}}$, $\mathbf{I_k}$.

© N.V. Philips' Gloeilampenfabrieken

Reverse currents I_{rlat} and I_{rver}

The main reverse current is separated into lateral and vertical components originating from the collector-base junction sidewall and bottom respectively. These formulations include Early and high injection effects and because the two currents depend on different internal collector-base junction voltages, collector current crowding is also modelled.

The lateral reverse current component *Irlat*

$$I_{rlat} = \left(\frac{4 \times (1 - \mathbf{X_{lrv}}) \times Ir1}{3 + \sqrt{1 + 16 \times \frac{Ir1}{\mathbf{I_k}}}} \right) / \, Flat \tag{A.11}$$

The vertical reverse current component *Irver*

$$I_{rver} = \left(\frac{4 \times \mathbf{X_{lrv}} \times Ir2}{3 + \sqrt{1 + 16 \times \frac{Ir2}{\mathbf{I_k}}}} \right) / \, Frver \tag{A.12}$$

Model parameters : $\mathbf{X_{lrv}}$.

The Base current

Forward components

The total forward base current is composed of an ideal and a non-ideal component. Both components depend on the bottom part of the emitter-base junction.

Ideal component :

$$I_{re} = \frac{If2}{\mathbf{B_f}} \tag{A.13}$$

Non-ideal component :

$$I_{le} = \frac{\mathbf{I_{bf}} \times \left(e^{Ve2b1/Vt} - 1 \right)}{e^{Ve2b1/2\,Vt} + e^{\mathbf{V_{lf}}/2\,Vt}} \tag{A.14}$$

Model parameters : $\mathbf{B_f}$, $\mathbf{I_{bf}}$, $\mathbf{V_{lf}}$.

Reverse components

The total reverse base current is composed of an ideal and a non-ideal component. Both components depend on the bottom part of the collector-base junction.

© N.V. Philips' Gloeilampenfabrieken

Ideal component :

$$I_{rc} = \frac{Ir2}{\mathbf{B_r}} \tag{A.15}$$

Non-ideal component :

$$I_{lc} = \frac{\mathbf{I_{br}} \times \left(e^{Vc2b2/Vt} - 1\right)}{e^{Vc2b2/2\,Vt} + e^{\mathbf{V_{lr}}/2\,Vt}} \tag{A.16}$$

Model parameters : $\mathbf{B_r}$, $\mathbf{I_{br}}$, $\mathbf{V_{lr}}$.

The substrate current

Forward component

The forward substrate component depends on the bottom part of the emitter-base junction and consists of an ideal component and a component subject to high injection effects. The parameter $\mathbf{X_{hes}}$ determines the fraction subject to high injection.

$$I_{se} = (1 - \mathbf{X_{hes}}) \times \mathbf{X_{es}} \times If2 \quad + \quad \frac{4 \times \mathbf{X_{hes}} \times \mathbf{X_{es}} \times If2}{3 + \sqrt{1 + 16 \times \frac{If2}{\mathbf{I_k}}}} \tag{A.17}$$

Model parameters : $\mathbf{X_{es}}$, $\mathbf{X_{hes}}$.

Reverse component

The reverse substrate component depends on the bottom part of the collector-base junction and consists of an ideal component and a component subject to high injection effects. The parameter $\mathbf{X_{hcs}}$ determines the fraction subject to high injection.

$$I_{sc} = (1 - \mathbf{X_{hcs}}) \times \mathbf{X_{cs}} \times Ir2 \quad + \quad \frac{4 \times \mathbf{X_{hcs}} \times \mathbf{X_{cs}} \times Ir2}{3 + \sqrt{1 + 16 \times \frac{Ir2}{\mathbf{I_k}}}} \tag{A.18}$$

Model parameters : $\mathbf{X_{cs}}$, $\mathbf{X_{hcs}}$.

Additional substrate and base current

An ideal diode models the substrate-base junction. The reverse leakage current of this junction can be used to model the zero-crossover phenomena sometimes observed in the base current at

© N.V. Philips' Gloeilampenfabrieken

low bias conditions and high temperatures.

$$I_{sf} = \mathbf{I_{ss}} \times \left(e^{Vsb/Vt} - 1\right) \tag{A.19}$$

Model parameters : $\mathbf{I_{ss}}$.

Depletion charges

The Poon-Gummel formulation is used in the modelling of the depletion charges.

Emitter-base depletion charge

$$Q_{te} = \frac{-\mathbf{C_{je}}}{1-\mathbf{P_e}} \times \left\{ \frac{\mathbf{V_{de}} - Ve2b1}{\left((1 - \frac{Ve2b1}{\mathbf{V_{de}}})^2 + \delta\right)^{\frac{\mathbf{P_e}}{2}}} \right\} \tag{A.20}$$

Model parameters : $\mathbf{C_{je}}$, $\mathbf{V_{de}}$, $\mathbf{P_e}$.

Collector-base depletion charge

$$Q_{tc} = \frac{-\mathbf{C_{jc}}}{1-\mathbf{P_c}} \times \left\{ \frac{\mathbf{V_{dc}} - Vc2b2}{\left((1 - \frac{Vc2b2}{\mathbf{V_{dc}}})^2 + \delta\right)^{\frac{\mathbf{P_c}}{2}}} \right\} \tag{A.21}$$

Model parameters : $\mathbf{C_{jc}}$, $\mathbf{V_{dc}}$, $\mathbf{P_c}$.

Substrate-base depletion charge

$$Q_{ts} = \frac{-\mathbf{C_{js}}}{1-\mathbf{P_s}} \times \left\{ \frac{\mathbf{V_{ds}} - Vsb}{\left((1 - \frac{Vsb}{\mathbf{V_{ds}}})^2 + \delta\right)^{\frac{\mathbf{P_s}}{2}}} \right\} \tag{A.22}$$

Model parameters : $\mathbf{C_{js}}$, $\mathbf{V_{ds}}$, $\mathbf{P_s}$.

© N.V. Philips' Gloeilampenfabrieken

Charges

Forward stored charges

The storage of charge in the forward active case is divided into three main components. The first component represents charge storage in the epilayer between emitter and collector. Charge storage in the epilayer under the emitter is another component and the storage of charge in the neutral regions forms the third component. The neutral charge formulation is obtained simply from the charge control principle. The epilayer charge storage formulations, however, are obtained by relating the charge storage to the injected minority concentration, p', in the epilayer. In the epilayer between emitter and collector p' is assumed to have a linear profile for all injection levels.

Charge stored in epitaxial base region between emitter and collector:

$$Qflat = \mathbf{T_{lat}} \times \mathbf{I_k} \times \left(\sqrt{1 + 16 \times \frac{If1}{\mathbf{I_k}}} - 1\right) \times \frac{Flat}{8} \tag{A.23}$$

Charge stored in epitaxial base region under emitter:

$$Qfver = \mathbf{T_{fvr}} \times \mathbf{I_k} \times \left(\sqrt{1 + 16 \times \frac{If2}{\mathbf{I_k}}} - 1\right) \ / \ 8 \tag{A.24}$$

Charge stored in emitter and buried layer under emitter:

$$Qfn = \mathbf{T_{fn}} \times If2 \tag{A.25}$$

Model parameters : $\mathbf{T_{lat}}$, $\mathbf{T_{fvr}}$, $\mathbf{T_{fn}}$.

Reverse stored charges

The storage of charge in the reverse active case is divided into three main components. The first component represents charge storage in the epilayer between emitter and collector. Charge storage in the epilayer under the collector is another component and the storage of charge in the neutral regions forms the third component. Charge formulations are obtained in a similar manner to the forward case.

Charge stored in epitaxial base region between emitter and collector:

$$Qrlat = \mathbf{T_{lat}} \times \mathbf{I_k} \times \left(\sqrt{1 + 16 \times \frac{Ir1}{\mathbf{I_k}}} - 1\right) \times \frac{Flat}{8} \tag{A.26}$$

Charge stored in epitaxial base region under collector:

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

$$Qrver = \mathbf{T_{rvr}} \times \mathbf{I_k} \times \left(\sqrt{1 + 16 \times \frac{Ir2}{\mathbf{I_k}}} - 1\right) \; / \; 8 \tag{A.27}$$

Charge stored in collector and buried layer under collector:

$$Qrn = \mathbf{T_{rn}} \times Ir2 \tag{A.28}$$

Model parameters : $\mathbf{T_{rvr}}$, $\mathbf{T_{rn}}$.

Substrate-base stored charge

Charge stored in substrate and base due to the substrate-base junction. This charge storage *only* occurs when the substrate-base junction is forward biased (note that Tsd is a constant):

$$Qsd = Tsd \times I_{sf} \tag{A.29}$$

Series resistances

$$\begin{array}{rlcl} \text{emitter:} & \mathbf{R_{eex}} & = & \text{constant} \\ & \mathbf{R_{ein}} & = & \text{constant} \\ \text{collector:} & \mathbf{R_{cex}} & = & \text{constant} \\ & \mathbf{R_{cin}} & = & \text{constant} \end{array}$$

The conductivity modulation of the base resistances is derived from the fact that the voltage drop across the epitaxial layer is inversely proportional to the electron concentration under the emitter and collector.

Base resistance under the emitter:

$$Rbe = \mathbf{R_{bec}} + \frac{2 \times \mathbf{R_{bev}}}{1 + \sqrt{1 + 16 \times \frac{If2}{\mathbf{I_k}}}} \tag{A.30}$$

Base resistance under the collector:

$$Rbc = \mathbf{R_{bcc}} + \frac{2 \times \mathbf{R_{bcv}}}{1 + \sqrt{1 + 16 \times \frac{Ir2}{\mathbf{I_k}}}} \tag{A.31}$$

The resistance $\mathbf{R_{sb}}$ models ohmic leakage across the substrate-base junction.

Model parameters : $\mathbf{R_{eex}}$, $\mathbf{R_{ein}}$, $\mathbf{R_{cex}}$, $\mathbf{R_{cin}}$, $\mathbf{R_{bec}}$, $\mathbf{R_{bev}}$, $\mathbf{R_{bcc}}$, $\mathbf{R_{bcv}}$, $\mathbf{R_{sb}}$.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Temperature dependence of the parameters

In this section the superscript 'r' denotes the parameter value at the reference temperature $\mathbf{T_{ref}}$.

$$T_k = \mathbf{T_{ref}} + 273.16 \quad (A.32)$$

$$T_n = \frac{Temp}{\mathbf{T_{ref}} + 273.16} \quad (A.33)$$

$$T_i = \frac{1}{\mathbf{T_{ref}} + 273.16} - \frac{1}{Temp} \quad (A.34)$$

Series resistances:

$$\mathbf{R_{cin}} = \mathbf{R_{cin}}^r \times T_n^{\mathbf{S_{pc}}} \quad (A.35)$$

$$\mathbf{R_{bcc}} = \mathbf{R_{bcc}}^r \times T_n^{\mathbf{S_{nbn}}} \quad (A.36)$$

$$\mathbf{R_{bcv}} = \mathbf{R_{bcv}}^r \times T_n^{\mathbf{S_{nb}}} \quad (A.37)$$

$$\mathbf{R_{bec}} = \mathbf{R_{bec}}^r \times T_n^{\mathbf{S_{nbn}}} \quad (A.38)$$

$$\mathbf{R_{bev}} = \mathbf{R_{bev}}^r \times T_n^{\mathbf{S_{nb}}} \quad (A.39)$$

$$\mathbf{R_{ein}} = \mathbf{R_{ein}}^r \times T_n^{\mathbf{S_{pe}}} \quad (A.40)$$

$\mathbf{R_{eex}}$ and $\mathbf{R_{cex}}$ are assumed temperature independent.

Depletion capacitances:

$$\mathbf{V_{d_x}} = -3k\frac{Temp}{q} \times \ln(T_n) + \mathbf{V_{d_x}}^r \times T_n + (1 - T_n) \times V_{gap} \quad (A.41)$$

$$\mathbf{C_{j_x}} = \mathbf{C_{j_x}}^r \times \left(\frac{\mathbf{V_{d_x}}^r}{\mathbf{V_{d_x}}}\right)^{P_x} \quad (A.42)$$

with:

for the emitter–base junction: $\begin{cases} V_{gap} = \mathbf{V_{geb}} \\ x = e \end{cases}$

for the collector–base junction: $\begin{cases} V_{gap} = \mathbf{V_{gcb}} \\ x = c \end{cases}$

for the substrate–base junction: $\begin{cases} V_{gap} = \mathbf{V_{gsb}} \\ x = s \end{cases}$

The internal diffusion voltage Vd:

$$Vd = -3k\frac{Temp}{q} \times \ln(T_n) + Vd^r \times T_n + (1 - T_n) \times \mathbf{V_{gb}} \quad (A.43)$$

The Early voltages:

$$\mathbf{E_{afl}} = \mathbf{E_{afl}}^r \times \sqrt{Vd/Vd^r} \quad (A.44)$$

© N.V. Philips' Gloeilampenfabrieken

The parameters $\mathbf{E_{arl}}$, $\mathbf{E_{afv}}$ and $\mathbf{E_{arv}}$ are subject to the same scaling rule.

$$\mathbf{I_s} = \mathbf{I_s}^r \times T_n^{4.0-\mathbf{S_{pb}}} \times e^{\left(q \times \mathbf{V_{gb}} \times T_i/k\right)} \tag{A.45}$$

$$\mathbf{B_f} = \mathbf{B_f}^r \times T_n^{\mathbf{A_e}-\mathbf{S_{pb}}} \times e^{\left(q \times \left(\mathbf{V_{gb}} - \mathbf{V_{ge}}\right) \times T_i/k\right)} \tag{A.46}$$

$$\mathbf{I_{bf}} = \mathbf{I_{bf}}^r \times T_n^2 \times e^{\left(q \times \left(\mathbf{V_{gje}}/2\right) \times T_i/k\right)} \tag{A.47}$$

$$\mathbf{I_k} = \mathbf{I_k}^r \times T_n^{\left(1-\mathbf{S_{pb}}\right)} \tag{A.48}$$

$$\mathbf{B_r} = \mathbf{B_r}^r \times \frac{\mathbf{B_f}}{\mathbf{B_f}^r} \tag{A.49}$$

$$\mathbf{I_{br}} = \mathbf{I_{br}}^r \times \frac{\mathbf{I_{bf}}}{\mathbf{I_{bf}}^r} \tag{A.50}$$

$$\mathbf{I_{ss}} = \mathbf{I_{ss}}^r \times T_n^2 \times e^{\left(q \times \mathbf{V_{gsb}} \times T_i/k\right)} \tag{A.51}$$

The transit times:

$$\mathbf{T_{lat}} = \mathbf{T_{lat}}^r \times T_n^{\left(\mathbf{S_{pb}}-1.0\right)} \tag{A.52}$$

$$\mathbf{T_{fvr}} = \mathbf{T_{fvr}}^r \times \frac{\mathbf{T_{lat}}}{\mathbf{T_{lat}}^r} \tag{A.53}$$

$$\mathbf{T_{fn}} = \mathbf{T_{fn}}^r \times T_n^{(\mathbf{S_x}-1.0)} \tag{A.54}$$

$$\mathbf{T_{rvr}} = \mathbf{T_{rvr}}^r \times \frac{\mathbf{T_{lat}}}{\mathbf{T_{lat}}^r} \tag{A.55}$$

$$\mathbf{T_{rn}} = \mathbf{T_{rn}}^r \times \frac{\mathbf{T_{fn}}}{\mathbf{T_{fn}}^r} \tag{A.56}$$

All other model parameters are assumed to be temperature independent.

Temperature parameters: $\mathbf{V_{geb}}$, $\mathbf{V_{gcb}}$, $\mathbf{V_{gsb}}$, $\mathbf{V_{gb}}$, $\mathbf{V_{ge}}$, $\mathbf{V_{gje}}$, $\mathbf{A_e}$, $\mathbf{S_{pb}}$, $\mathbf{S_{nb}}$, $\mathbf{S_{nbn}}$, $\mathbf{S_{pe}}$, $\mathbf{S_{pc}}$, $\mathbf{S_x}$.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Parameter list

	Name	Description of the parameter	Example
1	**Level**	Level of the model (500 or Modella)	
2	$\mathbf{I_s}$	Collector-emitter saturation current (A)	1.8×10^{-16}
3	$\mathbf{B_f}$	Ideal forward common-emitter current gain	131
4	$\mathbf{I_{bf}}$	Saturation current of non-ideal forward base current (A).	2.6×10^{-14}
5	$\mathbf{V_{lf}}$	Cross-over voltage of non-ideal forward base current (V)	0.54
6	$\mathbf{I_k}$	High injection knee current (A)	1.1×10^{-4}
7	$\mathbf{X_{ifv}}$	Vertical fraction of forward current	0.43
8	$\mathbf{E_{afl}}$	Early voltage of the lateral forward current component at zero collector-base bias (V)	20.5
9	$\mathbf{E_{afv}}$	Early voltage of the vertical forward current component at zero collector-base bias (V)	75.0
10	$\mathbf{B_r}$	Ideal reverse common-emitter current gain	25.0
11	$\mathbf{I_{br}}$	Saturation current of non-ideal reverse base current (A)	1.2×10^{-13}
12	$\mathbf{V_{lr}}$	Cross-over voltage of non–ideal reverse base current (V)	0.48
13	$\mathbf{X_{irv}}$	Vertical fraction of reverse current	0.43
14	$\mathbf{E_{arl}}$	Early voltage of the lateral reverse current component at zero emitter-base bias (V)	13.1
15	$\mathbf{E_{arv}}$	Early voltage of the vertical reverse current component at zero emitter-base bias (V)	104.0
16	$\mathbf{X_{es}}$	Ratio between saturation current of e-b-s transistor and e-b-c transistor	2.7×10^{-3}
17	$\mathbf{X_{hes}}$	Fraction of substrate current of e-b-s transistor subject to high injection	0.7
18	$\mathbf{X_{cs}}$	Ratio between the saturation current of c-b-s transistor and c-b-e transistor	3.0
19	$\mathbf{X_{hcs}}$	Fraction of substrate current of e-b-s transistor subject to high injection	1.0
20	$\mathbf{I_{ss}}$	Saturation current of substrate-base diode (A)	4.0×10^{-13}
21	$\mathbf{R_{cex}}$	External part of the collector resistance (Ω)	5.0
22	$\mathbf{R_{cin}}$	Internal part of the collector resistance (Ω)	47.0
23	$\mathbf{R_{bcc}}$	Constant part of the base resistance RBC (Ω).	10.0
24	$\mathbf{R_{bcv}}$	Variable part of the base resistance RBC (Ω).	10.0

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

	Name	Description of the parameter	Default
25	$\mathbf{R_{bec}}$	Constant part of the base resistance RBE (Ω)	10.0
26	$\mathbf{R_{bev}}$	Variable part of the base resistance RBE (Ω)	50.0
27	$\mathbf{R_{eex}}$	External part of the emitter resistance (Ω)	27.0
28	$\mathbf{R_{ein}}$	Internal part of the emitter resistance (Ω)	66.0
29	$\mathbf{R_{sb}}$	Substrate-base leakage resistance (Ω)	1.0×10^{15}
30	$\mathbf{T_{lat}}$	Low injection (forward and reverse) transit time of charge stored in the epilayer between emitter and collector (s)	2.4×10^{-9}
31	$\mathbf{T_{fvr}}$	Low injection forward transit time due to charge stored in the epilayer under the emitter (s)	3.0×10^{-8}
32	$\mathbf{T_{fn}}$	Low injection forward transit time due to charge stored in the emitter and the buried layer under the emitter (s)	2.0×10^{-10}
33	$\mathbf{C_{je}}$	Zero-bias emitter-base depletion capacitance (F)	6.1×10^{-14}
34	$\mathbf{V_{de}}$	Emitter-base diffusion voltage (V)	0.52
35	$\mathbf{P_{e}}$	Emitter-base grading coefficient	0.3
36	$\mathbf{T_{rvr}}$	Low injection reverse transit time due to charge stored in the epilayer under the collector (s)	1.0×10^{-9}
37	$\mathbf{T_{rn}}$	Low injection reverse transit time due to charge stored in the collector and the buried layer under the collector (s)	3.0×10^{-9}
38	$\mathbf{C_{jc}}$	Zero-bias collector-base depletion capacitance (F)	3.9×10^{-13}
39	$\mathbf{V_{dc}}$	Collector-base diffusion voltage (V)	0.57
40	$\mathbf{P_{c}}$	Collector-base grading coefficient	0.36
41	$\mathbf{C_{js}}$	Zero-bias substrate-base depletion capacitance (F)	1.3×10^{-12}
42	$\mathbf{V_{ds}}$	Substrate-base diffusion voltage (V)	0.52
43	$\mathbf{P_{s}}$	Substrate-base grading coefficient	0.35
44	$\mathbf{T_{ref}}$	Reference temperature of the parameter set ($^\circ C$)	25.0
45	Dta	Difference between the device temperature and the ambient analysis temperature ($^\circ C$)	0.0
46	$\mathbf{V_{geb}}$	Bandgap voltage of the emitter-base depletion region (V)	1.206
47	$\mathbf{V_{gcb}}$	Bandgap voltage of the collector-base depletion region (V)	1.206

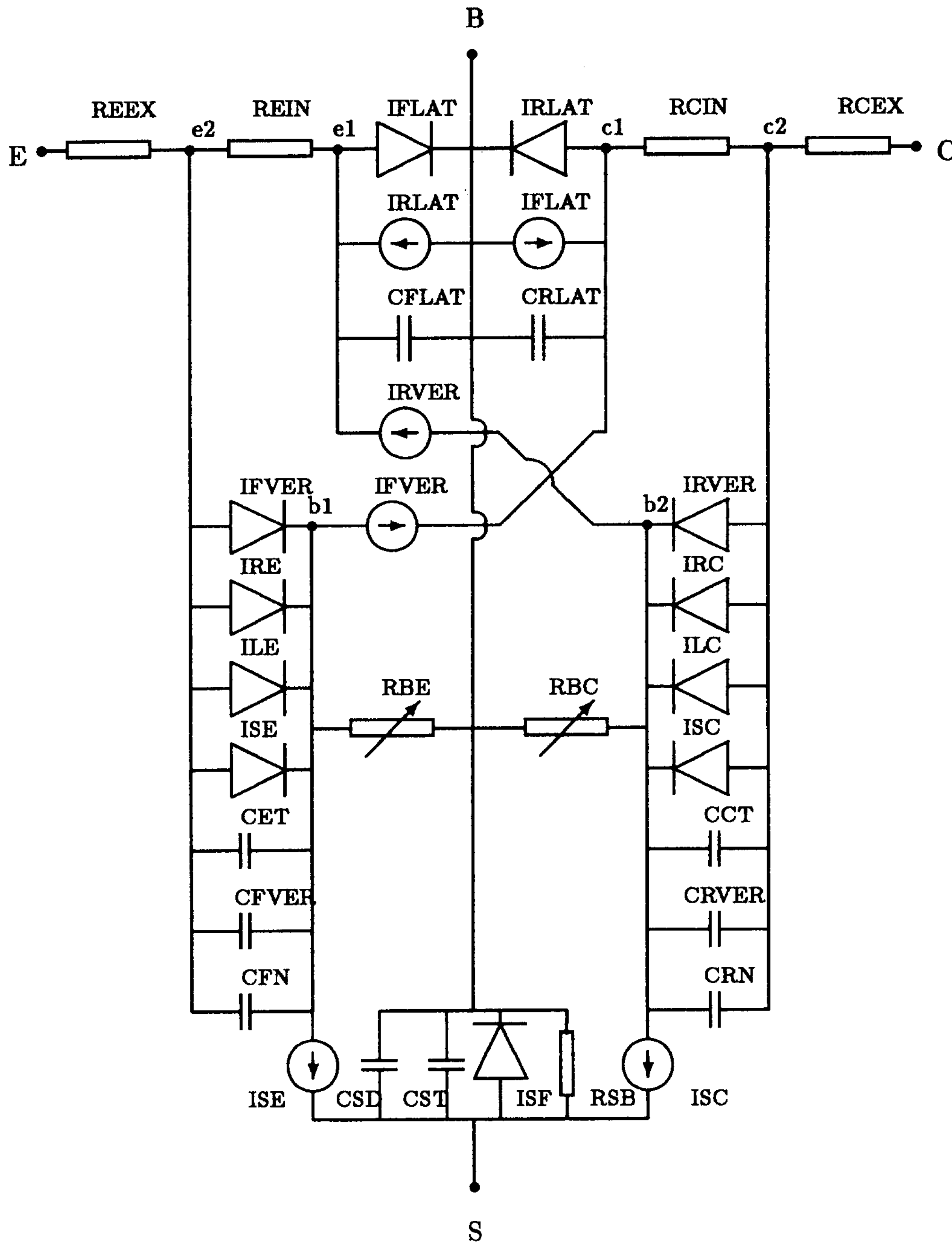
© N.V. Philips' Gloeilampenfabrieken

	Name	Description of the parameter	Default
48	$\mathbf{V_{gsb}}$	Bandgap voltage of the substrate-base depletion region (V)	1.206
49	$\mathbf{V_{gb}}$	Bandgap voltage of the base between emitter and collector (V)	1.206
50	$\mathbf{V_{ge}}$	Bandgap voltage of the emitter (V)	1.206
51	$\mathbf{V_{gje}}$	Bandgap voltage recombination emitter-base junction (V)	1.123
52	$\mathbf{A_e}$	Temperature coefficient of $\mathbf{B_f}$	4.48
53	$\mathbf{S_{pb}}$	Temperature coefficient of the epitaxial base hole mobility	2.853
54	$\mathbf{S_{nb}}$	Temperature coefficient of the epitaxial base electron mobility	2.6
55	$\mathbf{S_{nbn}}$	Temperature coefficient of buried layer electron mobility	0.3
56	$\mathbf{S_{pe}}$	Temperature coefficient of emitter hole mobility	0.73
57	$\mathbf{S_{pc}}$	Temperature coefficient of collector hole mobility	0.73
58	$\mathbf{S_x}$	Temperature coefficient of combined minority carrier mobilities in emitter and buried layer	1.0

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

MODELLA



© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Appendix B

Parameter optimisation theory

The use of BIPAR to determine optimised parameter sets has been treated to a limited extent in chapter 4. In this appendix, however, a more complete description of the parameter optimisation process and the associated model evaluation routine will be presented.

B.1 The optimisation process

In BIPAR the optimisation process is based on the least squares method. This involves solving the least squares problem...

$$minimise\ F = \sum_{i=1}^{n} (w_i \times \sigma_i)^2 \tag{B.1}$$

as a function of the model parameters,
where the deviation σ_i is given by

$$\sigma_i = f_i(\vec{ZC}, \vec{P}) - y_i \tag{B.2}$$

with...

f_i	=	model response or prediction
$\vec{ZC}$	=	vector of experimentally accessible variables. These are the terminal quantities, termed *settings,* like Ic, Ib, Vcb, etc. This vector contains thus the total set of independent variables which characterises one measuring condition (or operating point).
$\vec{P}$	=	vector of model parameters.
y_i	=	the measured output response i.e. measured values for Ic, Ib, etc.

Also

w_i	=	a weighting factor which is equal to the reciprocal

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

of the measured values. This ensures that the *relative* deviation is minimised.

n = number of measurement points over which the optimisation is to be carried out.

The procedure which solves this least squares problem needs, for each point i, the following information...

- y_i - the measured data
- f_i - the model prediction which of course is a function not only of the model parameters but also of the measuring condition.

Given this information at each of the measurement points $i = 1$ to n, the procedure can evaluate F. The *minimisation* of F as a function of the model parameters however, requires knowledge of the sensitivity (i.e. derivative) of the model response f_i to the specified parameter(s). The function of the model evaluation routine, which is called repeatedly during the optimisation process, is to provide not only the model response f_i but also the derivatives of f_i with respect to the parameters $\vec{P}$.

To illustrate the above process it may be useful to look briefly at the optimisation of one parameter, say the saturation current $\mathbf{I_s}$. The user specifies that $\mathbf{I_s}$ is to be optimised by fitting the model prediction for Ic (f_i) to the measured values for Ic (y_i).

The optimisation range in Ic is selected along with the two independent settings (contained in $\vec{ZC}$) which define the measuring condition i.e. in this case Veb and Vcb. The model evaluation routine provides the prediction for Ic which is a function not only of the value of parameter $\mathbf{I_s}$ but also of the measuring condition. Also the derivative of Ic with respect to $\mathbf{I_s}$ is made available by the model evaluation routine.

With access to all this information the least squares procedure solves eqn.B.1 and provides an optimised value for $\mathbf{I_s}$.

The procedure which solves the least squares problem is a standard NAG subroutine called CRLESQ. The model evaluation routine used by CRLESQ was developed from first principles and is described in the next section.

B.2 The model evaluation routine

As stated above, the function of the model evaluation routine is to provide the model response $f_i(\vec{ZC}, \vec{P})$. This model response is directly related to the parameter values. Initially these parameter values are estimated and because of this the model prediction will often deviate greatly from the measured data y_i. However, when the parameters are optimised the model prediction should deviate only slightly (ideally not at all) from y_i.

The model response is also a function of the settings $\vec{ZC}$ because $\vec{ZC}$ contains the two independent settings which characterise one measuring condition. If we say...

© N.V. Philips' Gloeilampenfabrieken

$\vec{Z}_i$ = the independent subset of $\vec{ZC}$ i.e. the two independent settings
$\vec{Z}_d$ = the dependent subset of $\vec{ZC}$ i.e. all other settings

then the primary aim of the model evaluation routine could be stated as: calculate $\vec{Z}_d$ given $\vec{Z}_i$ and $\vec{P}$. The model response $f_i(\vec{ZC}, \vec{P})$ will be contained in $\vec{Z}_d$. In the example given above of $\mathbf{I_s}$ optimisation, $\vec{Z}_i$ would be Veb and Vcb, $\vec{Z}_d$ would be the set ($Ic, Ie, Ib, Isub$ and Vec) where in particular Ic is the model response f_i to be fitted to measured data by optimising $\mathbf{I_s}$.

The reason why *two* independent settings ($\vec{Z}_i$) are required to characterise one measuring condition is that for bipolar transistors the branch currents are given as a function of the internal emitter-base and collector-base voltages. To determine the two unknowns therefore requires knowledge of two terminal settings which specify the operating point. Knowing two terminal settings allows the equivalent circuit to be solved. Because an iterative procedure is required in this solution process the model is termed *implicit*.

A second purpose of the model evaluation routine is to provide the derivatives of the settings $\vec{ZC}$ (and therefore the model response f_i) to the parameters $\vec{P}$. This is done numerically at present by calling the model evaluation routine a second time with a perturbation in model parameter value(s). The difference in $\vec{ZC}$ leads to estimates for the derivatives.

The actual solution of the equivalent circuit constitutes the bulk of the model evaluation. This solution process is outlined in the next section.

B.2.1 The solution process

In circuit simulation the procedure often adopted to solve the internal network of a circuit is to solve for those values of node voltages which drive the summation of currents at each node to zero. These currents are functions of one or more of the internal node voltages, the latter of which are assumed independent of each other. The solution of the internal network of MODELLA as implemented in the model evaluation routine uses however a different procedure. Two terminal settings are specified as the independent settings and then an iterative procedure is used to find those internal voltages which lead to model predictions of the two terminal settings which are consistent with measured values. This iterative procedure is a Newton-Raphson iteration which converges when the difference between measured values for the two settings and values predicted by the model are within a specified error limit. When convergence has been obtained all internal voltages are known and it is possible to calculate all the other dependent terminal settings.

In TPL-301 for example there are two internal voltages ($V_{e'b'}$ and $V_{c'b'}$) found in the model equations and it is the solution of these two internal voltages which comprises the solution process. By an examination of the component dependency scheme (which shows the dependency of currents and charges on the independent voltages) it can be seen that in MODELLA there are four internal junction voltages required to evaluate the model equations. They are

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

$Ve2b1, Ve1b, Vc2b2$ and $Vc1b$. However, they are not independent of each other because...

$$Ve2b1 + Irbe \times Rbe = Ie2e1 \times \mathbf{R_{ein}} + Ve1b \quad (B.3)$$
$$Vc2b2 + Irbc \times Rbc = Ic2c1 \times \mathbf{R_{cin}} + Vc1b \quad (B.4)$$

where $Irbe$ is the current through Rbe, $Ie2e1$ is the current flowing from node $e2$ to $e1$, etc. Because we only have two independent settings we can only solve for two of these four voltages in one Newton-Raphson process. A second Newton-Raphson process is therefore required to solve for the other two voltages to facilitate the calculation of all model equations. By examining the equivalent circuit and model equations it is evident that the logical choice of voltages to solve in the outer loop is $Ve2b1$ and $Vc2b2$. The inner loop in this nested iteration procedure then solves $Ve1b$ and $Vc1b$. A more detailed description of this nested solution procedure now follows.

The variables x and y are chosen to denote $Ve2b1$ and $Vc2b2$ respectively and w and z to denote $Ve1b$ and $Vc1b$ respectively.

If we take for example Ib and Vcb as the two independent settings [1] we may write the ideal aim of the solution process as

$$S1 = \frac{Ib_{model} - Ib_{meas}}{Ib_{meas}} = 0 \quad (B.5)$$
$$S2 = Vcb_{model} - Vcb_{meas} = 0 \quad (B.6)$$

The subscripts *model* and *meas* denote the model prediction and measured values respectively. Using the relative variation for the current represents the fact that with convergence criterion in the solution process, the absolute error ϵ for currents is a much smaller value than that for voltages. Note that if other elements of $\vec{ZC}$ were chosen as independent settings then they would appear instead of Ib and Vcb.

The method, which is sketched in general terms in figure B.1, now proceeds as follows :

1. x,y,w and z are estimated for each measurement condition. This is done once before the optimisation begins. Because the model routine is called repeatedly during the least squares optimisation the same conditions reoccur. Therefore the final values of x,y,w and z from the previous calculation are used to speed up convergence.

2. Using the estimates for x and y the currents $Irbe(= Ire + Ile)$ and $Irbc(= Irc + Ilc)$ are determined, thus allowing the voltages $Ve2b$ and $Vc2b$ to be calculated directly from the relations

$$Ve2b = Ve2b1 + Irbe \times Rbe \quad (B.7)$$
$$Vc2b = Vc2b2 + Irbc \times Rbc \quad (B.8)$$

3. With the estimated values for $Ve1b$ and $Vc1b$ it is now possible to calculate the currents $Ie2e1$ and $Ic2c1$ (i.e. the currents flowing between nodes $e2$ and $e1$ and between $c2$ and $c1$ respectively).

[1] In the forward active case Ib is determined by the internal emitter-base voltage $Ve2b1$. The choice of Ib and Vcb therefore defines the operation point since both sides of the equivalent circuit are taken into account.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

4. By taking Ve2b and Vc2b as 'internal settings' just as Ib and Vcb were taken as 'external settings' above, we can write :

$$S1' = (Ie2e1 \times \mathbf{R_{ein}} + Ve1b) - Ve2b = 0$$
$$S2' = (Ic2c1 \times \mathbf{R_{cin}} + Vc1b) - Vc2b = 0$$

where the prime denotes the fact that we are now dealing with the internal loop. $\mathbf{R_{ein}}$ and $\mathbf{R_{cin}}$ are the internal emitter and collector resistances as shown in figure 3.2. $S1'$ and $S2'$ are now calculated as are their derivatives to w and z i.e.

$\frac{\partial S1'}{\partial w}, \frac{\partial S1'}{\partial z}, \frac{\partial S2'}{\partial w}, \frac{\partial S2'}{\partial z}$.

5. w and z are varied according to the Newton-Raphson method [2] to drive $S1'$ and $S2'$ to zero. This leads to the equations :

$$\frac{\partial S1'}{\partial w} dw + \frac{\partial S1'}{\partial z} dz = -S1'$$

$$\frac{\partial S2'}{\partial w} dw + \frac{\partial S2'}{\partial z} dz = -S2'$$

This system of equations is solved to yield dw and dz and then w and z are updated according to

$$w_{new} = w + dw$$
$$z_{new} = z + dz$$

6. The process is repeated from step 3 with the new values for w and z until $S1'$ and $S2'$ are small enough i.e. within the specified error limits.

7. Returning to the outer loop it is now possible to calculate all the terminal settings using the estimated values of x and y and converged values of w and z.

8. The derivatives of the terminal settings to x and y are calculated. In equation form :

[2] See for example [47]. In short the idea is to determine the variables x_i , $i = 1, 2.., N$ which drive the function value $g_i(x_1, x_2, ..., x_N)$, $i = 1, 2, .., .N$ to zero. If we denote X as the entire vector of values x_i, then in the neighbourhood of X each of the functions g_i can be expanded in a first order Taylor series....

$$g_i(X + dX) = g_i(X) + \sum_{j=1}^{N} \frac{\partial g_i}{\partial x_j} dx_j \qquad (B.9)$$

This leads to a set of linear equations for the corrections dX that move each function closer to zero simultaneously....

$$\sum_{j=1}^{N} \frac{\partial g_i}{\partial x_j} dx_j = -g_i(X) \qquad (B.10)$$

This equation is solved and the corrections are added to the solution vector..

$$x_i^{new} = x_i^{old} + dx_i \quad i = 1, ..., N \qquad (B.11)$$

and the process is iterated to convergence.

© N.V. Philips' Gloeilampenfabrieken

$$\frac{d\vec{ZC}}{dx} = \frac{\partial \vec{ZC}}{\partial x} + \frac{\partial \vec{ZC}}{\partial w}\frac{\partial w}{\partial x} + \frac{\partial \vec{ZC}}{\partial z}\frac{\partial z}{\partial x}$$

$$\frac{d\vec{ZC}}{dy} = \frac{\partial \vec{ZC}}{\partial y} + \frac{\partial \vec{ZC}}{\partial w}\frac{\partial w}{\partial y} + \frac{\partial \vec{ZC}}{\partial z}\frac{\partial z}{\partial y}$$

The partial derivatives with respect to w and z have already been calculated and $\frac{\partial \vec{ZC}}{\partial x}$ and $\frac{\partial \vec{ZC}}{\partial y}$ are easily obtained. The other partial derivatives however are derived as follows : Consider

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{pmatrix} \begin{pmatrix} dw \\ dz \end{pmatrix} \tag{B.12}$$

Denoting the 2x2 matrix as A and noting that :

$$\begin{pmatrix} dw \\ dz \end{pmatrix} = \begin{pmatrix} \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \tag{B.13}$$

If we premultiply equation **(B.12)** by A^{-1} and then compare to equation **(B.13)** we can write the *unknown* derivatives with respect to x and y in terms of the *known* derivatives with respect to w and z. For example

$$\frac{\partial w}{\partial x} = \frac{\frac{\partial y}{\partial z}}{\frac{\partial x}{\partial w}\frac{\partial y}{\partial z} - \frac{\partial x}{\partial z}\frac{\partial y}{\partial w}}$$

etc.

9. $S1$ and $S2$ are now calculated together with their derivatives to x and y i.e.

   $\frac{\partial S1}{\partial x}, \frac{\partial S1}{\partial y}, \frac{\partial S2}{\partial x}, \frac{\partial S2}{\partial y}.$

10. x and y are varied according to the Newton-Raphson method to drive $S1$ and $S2$ to zero. This leads to the equations :

$$\frac{\partial S1}{\partial x}dx + \frac{\partial S1}{\partial y}dy = -S1$$

$$\frac{\partial S2}{\partial x}dx + \frac{\partial S2}{\partial y}dy = -S2$$

   This system of equations is solved to yield dx and dy and then x and y are updated according to

$$x_{new} = x + dx$$
$$y_{new} = y + dy$$

11. The process is repeated from step 2 with the new values for x and y until $S1$ and $S2$ are small enough.

Compared to a general circuit solver, this solution procedure is limited in that it is only suited to Modella's equivalent circuit. However, it is highly efficient in terms of C.P.U. time and has exhibited excellent convergence properties - two essentials for a parameter optimisation environment.

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

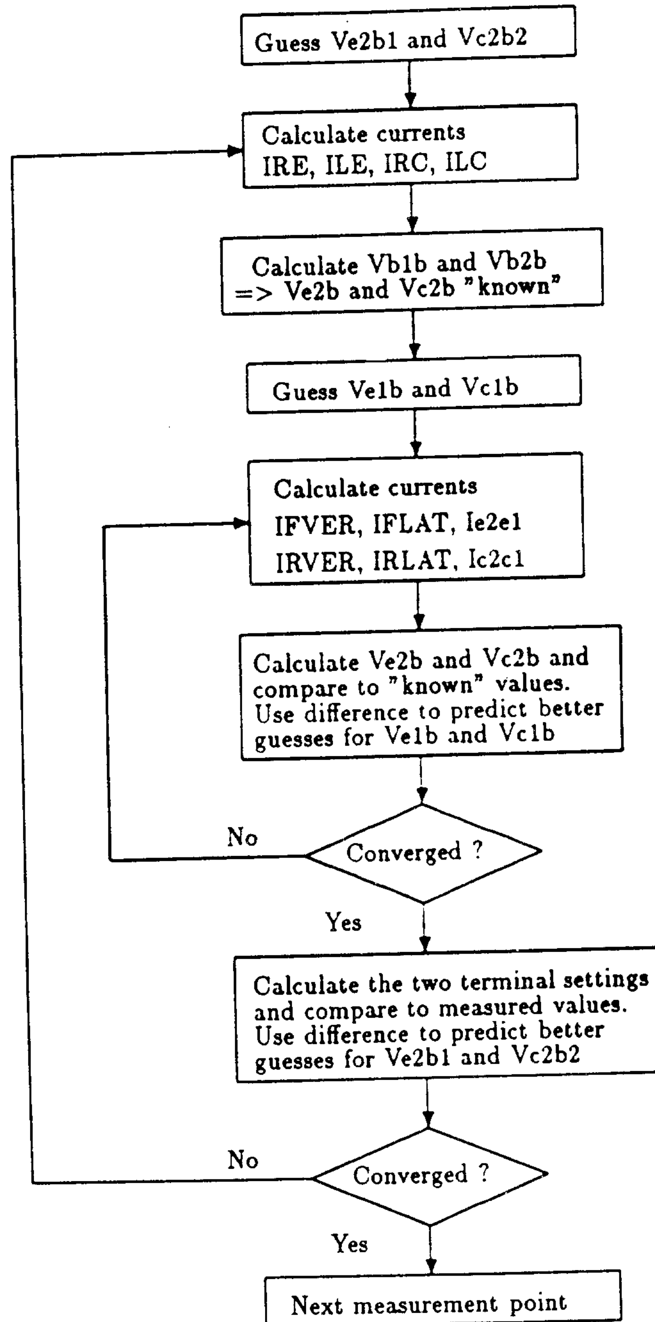


Figure B.1: *General outline of solution process for Modella in Bipar.*

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

Appendix C

Proceedings from the BCTM '90

© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

A NEW PHYSICAL COMPACT MODEL FOR LATERAL PNP TRANSISTORS

F.G. O'Hara, J.J.H. van den Biesen, H.C. de Graaff and J.B. Foley.#

Philips Research Laboratories
P.O. Box 80.000
5600 JA Eindhoven - The Netherlands

Dept. of Microelectonics & Elec. Eng.
Trinity College Dublin
Dublin 2 - Ireland

Abstract

A new lateral pnp compact model, suitable for computer-aided circuit design purposes, is presented. In this new formulation, called MODELLA, the equivalent circuit, analytical equations and model parameters are derived directly from the physics and structure of the lateral pnp. In comparisons with measured device characteristics the performance of this model is shown to be superior to the extended Gummel-Poon model.

Introduction

In the design of bipolar analog integrated circuits, greater flexibility is often achieved when npn and pnp transistors are combined in the circuit design. Many present day bipolar production processes use the conventional *lateral* pnp (see fig. 1) as the standard pnp transistor structure.

In spite of their widespread use, however, literature on lateral pnp compact modeling for computer-aided circuit design purposes is not very abundant[1]. In practice, present day lateral pnp compact models tend to be adaptations of their vertical npn counterparts (c.f. [2]), neglecting the complex 2-D nature of the lateral pnp (see fig. 1). This approach leads to semi-empirical formulations which lack the superiority of a truly physics-based compact model.

In this paper, we present a new compact model, called MODELLA, for the lateral pnp (oxide or junction isolated) which has been derived from device physics. In MODELLA, the basic analytical equations for currents and stored minority charges are derived from an analysis of a 1-D lateral pnp and a 1-D parasitic vertical pnp. This approach was shown to be valid by comparing analytical formulations with 2-D numerical device simulation results, in the case where ohmic voltage drops are excluded. Under these simulation conditions it was found, for example, that the main 2-D hole current density between emitter and collector could be accurately described, at all injection levels, by means of the hole current density of the 1-D lateral pnp device. The scaling factor from the 1-D case to the 2-D case was found to be a function solely of device geometry.

The 2-D aspects of the lateral pnp are then incorporated by modifying and extending these basic equations to include the effects of series resistances, most notably current crowding.

Model derivation

This section describes how the most important basic modeling equations in MODELLA are derived from 1-D lateral pnp device physics and how they are adapted to incorporate 2-D effects. Considering the forward active case, the following physical approach is used to derive the basic equations for the collector current and stored minority base charge.

- High injection effects in the epitaxial layer base region are modeled by relating the collector current, I_c, and the stored minority charge, Q_b, to the bias dependent minority carrier concentration, $p'(x)$, in the base region of a 1-D lateral pnp. Here, $p'(x)$ can be assumed to have a linear distribution in the base at *all injection levels* (see for example [3]). The effective hole diffusion coefficient is therefore defined in the center of the base region. A consequence of relating the currents and charges *independently* to the injected minority carrier concentration (an approach first used in the MEXTRAM [4] compact model) is that the base transit time, $\tau_B (= \frac{dQ_b}{dI_c}|_{dV_{ce}=0})$, is seen to decrease by a factor of two at high injection levels thereby correctly modeling the Webster effect[5]. This contrasts with the extended G-P model[2] (E-GP model) where the use of the charge control principle yields a bias independent base transit time.
- The Early effect[6] is modeled by including the effect of the collector-base depletion width, X_{CB}, on the base width, X_B, in the expression for the collector current :

$$Ic = \frac{Ic'}{1 - X_{CB}/X_B} \tag{1}$$

Here, Ic' is the collector current without Early effect and X_{CB} is obtained by assuming asymmetric abrupt junctions. In this way, the collector-base voltage dependence of the Early voltage is also modeled.

Combining these approaches we obtain a new expression suitable for accurately modeling both the high injection and the forward Early effects in a homogeneously doped base transistor.

$$Ic = \frac{4I_f \,/\, (3 + \sqrt{1 + 16 \frac{I_f}{I_k}})}{1 - \sqrt{1 - \frac{V_{cb}}{V_d}} \,/\, (1 + \frac{V_{eaf0}}{2V_d})} \tag{2}$$

where I_f = the ideal reference current
$= \mathbf{I_{f0}} \exp \frac{V_{eb}}{V_t}$
and V_d = the diffusion voltage

© N.V. Philips' Gloeilampenfabrieken

There are three model parameters in this expression: the saturation current $\mathbf{I_{f0}}$, the high injection knee current $\mathbf{I_k}$ and the forward Early voltage at zero collector-base bias $\mathbf{V_{eaf0}}$. By including the effect of the emitter-base depletion width on the base width this equation is extended further, using one extra parameter $\mathbf{V_{ear0}}$, to model the reverse Early effect.

The modeling of 2-D effects such as current crowding is based on the influence of the series resistances shown in fig. 1. This figure shows how the hole current flow lines which make up the collector current can be roughly divided into two distinct components: a purely lateral flow which originates at the emitter sidewall and a flow along curved trajectories which originates from the bottom of the emitter. At low current levels the sidewall component dominates. At high current levels the voltage drop across the lateral emitter resistance, Re_{LAT}, leads to a reduced sidewall junction voltage thereby reducing the contribution from this component. The second component then dominates and current crowding in the region under the emitter contact is observed. Numerical device simulations on 2-D structures have revealed that the larger effective base width associated with these trajectories has the following consequences for device behavior at high current levels.

- the current gain, h_{fe}, decreases.
- the base transit time, τ_B, increases thereby reducing the cut-off frequency, f_t.
- the Early effect is reduced in agreement with the experimentally observed dependence of the Early voltage on the emitter-base voltage, V_{eb} (see fig. 6).

In MODELLA this current crowding effect is modeled using double emitter diodes and current sources to represent the sidewall (I_{FLAT}) and bottom (I_{FVER}) components of I_c (see fig. 2). I_{FLAT} and I_{FVER} are described by means of eqn.(2) using the *internal* junction voltages V_{e1b} and V_{e2b1} respectively. In addition, different saturation current and Early voltage parameters are used in the respective expressions. At low current levels I_{FLAT} dominates due to its larger saturation current, whereas at high current levels I_{FVER} takes over due to the voltage drop across Re_{LAT}. The Early voltage parameter associated with I_{FVER} will be larger than that used with I_{FLAT} thus modeling the dependence of the Early voltage on V_{eb}.

In order to model the increase in τ_B due to current crowding, the minority charge storage in the epitaxial base is also split into two components. These components represent charge storage under the emitter (Q_{FVER}) and charge storage between the emitter and collector (Q_{FLAT}). The V_{cb} dependency of the minority charge storage is modeled by adapting the sidewall component, Q_{FLAT}, to include the effect of the collector-base depletion width on the base width. Charge storage under the emitter is not affected by V_{cb}.

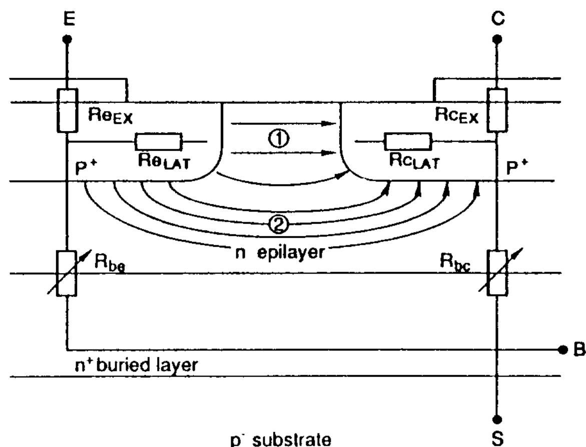


Figure 1: *Schematic cross-section of a lateral pnp transistor, showing the physical origin of the series resistances used in the equivalent circuit of MODELLA. The base resistances Rbe and Rbc both consist of constant and conductivity modulated parts. Also, the 2-D hole current flow in the epitaxial base region is shown to consist of two components: a purely lateral flow originating from the emitter sidewall and a flow along curved trajectories originating from under the emitter.*

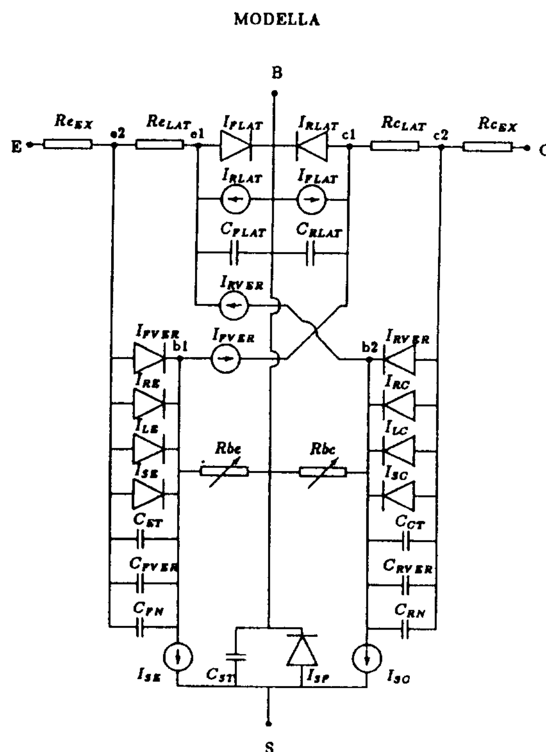


Figure 2: *Equivalent circuit diagram of MODELLA. Note that the emitter resistance is split into Re_{EX} and Re_{LAT} which, in combination with the double emitter diodes and current sources I_{FLAT} and I_{FVER}, models current crowding.*

Discussion and Results

The complete compact model includes ideal and non-ideal base currents, substrate currents with high injection effects, depletion charges, bias dependent base resistances and a substrate-base diode. The location of these elements in the equivalent circuit reflects the location of the underlying physical mechanisms in actual devices. For example, the ideal base current

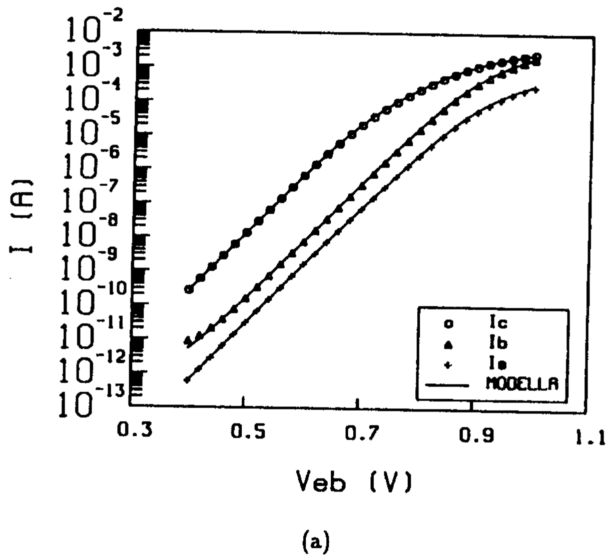
© N.V. Philips' Gloeilampenfabrieken

I_{RE} mainly originates from electrons injected into the emitter which recombine at the emitter contact. Because these minority carriers are injected from the epilayer base region under the emitter contact, the diode I_{RE} is located between nodes $e2$ and $b1$ in fig. 2. This means that the fall off in Ib at high current levels is unaffected by voltage drop across the lateral emitter resistance Re_{LAT}.

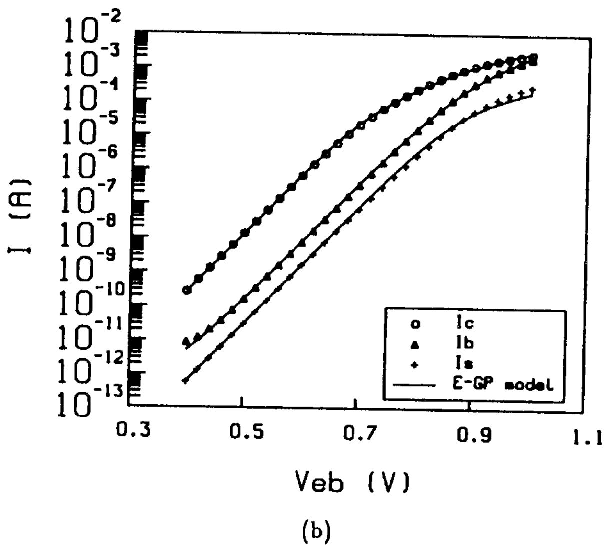
MODELLA has 42 parameters with 6 internal nodes which is more complex than the E-GP model (32 parameters with 3 internal nodes). However, due to its symmetry and the explicit nature of its equations which do not require any additional iterative solution procedures, MODELLA has shown excellent convergence properties when implemented in a circuit simulator[7]. In terms of CPU time, MODELLA is only a factor of 1.5 slower than the E-GP model and the number of iterations required for convergence is roughly equal. When compared to the semi-empirical E-GP approach, MODELLA's physical formulation possesses the following advantages:

- model parameters have a physical significance leading to easier, more reliable parameter determination.
- geometrical scaling rules can be applied with confidence.
- the influence of process variations on device behavior can be more accurately forecasted.
- the physical correlation between model parameters facilitates realistic statistical modeling.

A comparison between MODELLA and the E-GP model using a junction isolated lateral pnp is shown in figs. 3 to 6. In this example, the mask base width is $4\mu m$ and the collector completely surrounds the square $8 \times 8\mu m^2$ emitter. These

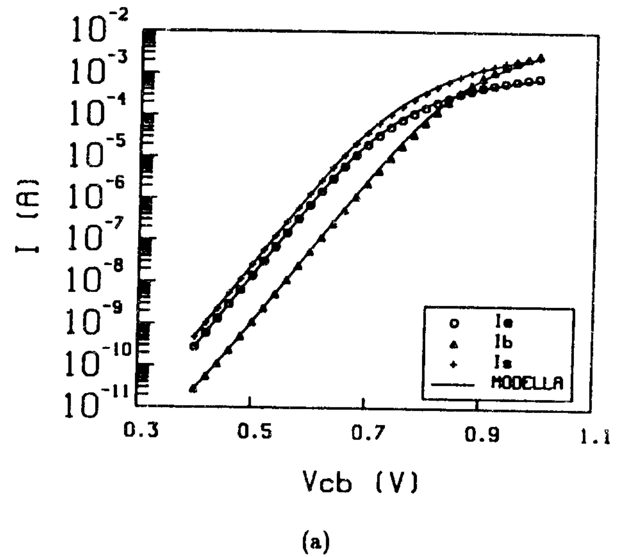


(a)

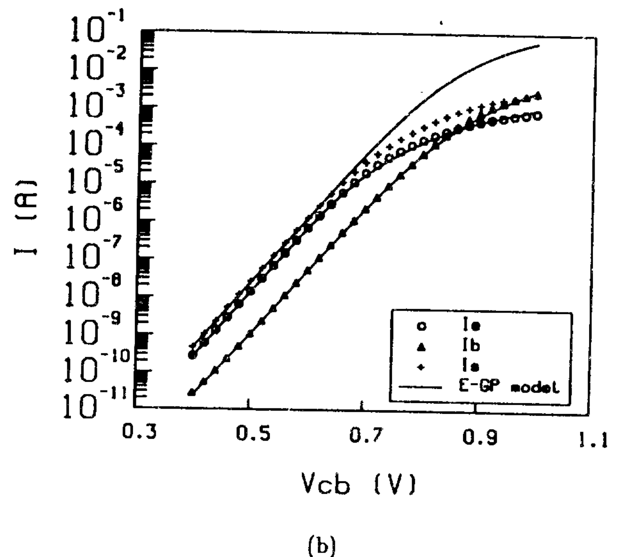


(b)

Figure 3: *Forward active Gummel plots comparing (a) MODELLA and (b) the E-GP model with measured characteristics.*



(a)



(b)

Figure 4: *Reverse active Gummel plots comparing (a) MODELLA and (b) the E-GP model with measured characteristics.*

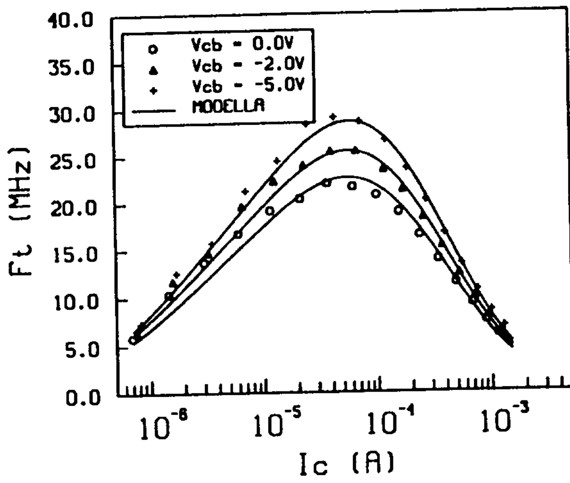
© N.V. Philips' Gloeilampenfabrieken

For Internal Use Only

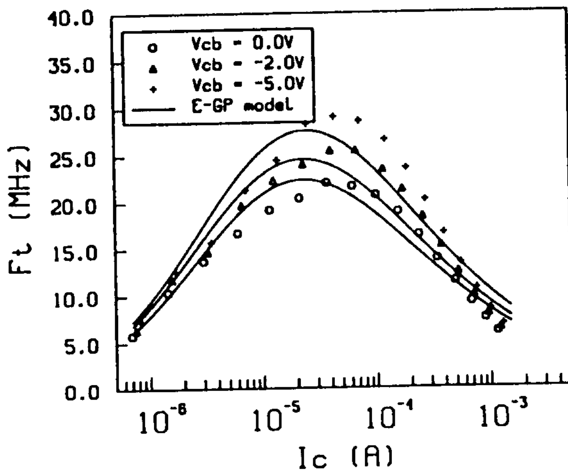
comparisons show the improved performance of MODELLA especially with respect to:

- reverse active behavior,
- the bias dependence of the cut-off frequency f_t,
- the bias dependence of the Early effect.

With the E-GP model the Early voltage, V_{eaf}, has a negligible dependence on Veb and is seen to be completely independent of Vcb. MODELLA, on the other hand, shows excellent agreement with measured data due to the physical modeling of both base width modulation by Vcb and current crowding under the emitter. This bias dependence of V_{eaf} is often a critical aspect in lateral pnp compact modeling e.g. in circuit applications involving lateral pnp current sources.



(a)



(b)

Figure 5: *Comparison between (a) MODELLA and (b) E-GP model predictions of the measured cut-off frequency, f_t, as a function of I_c at different V_{cb} values.*

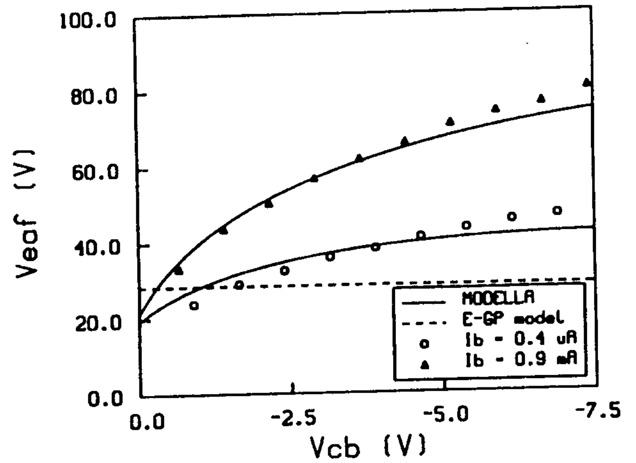


Figure 6: *Comparison between MODELLA and the E-GP model predictions of the measured forward Early voltage, V_{eaf} as a function of V_{cb}. The Early voltage is defined as $V_{eaf} = I_c|\frac{dV_{eb}}{dI_c}|_{I_b=const.} - V_{cb}$. The two I_b values correspond to $V_{eb} = 0.7V$ and $V_{eb} = 0.94V$, respectively.*

Conclusions

A new physical approach to lateral pnp compact modeling, which incorporates high injection effects, current crowding effects and a bias dependent output conductance, has been presented. MODELLA's superior performance has been demonstrated by comparison with the extended Gummel-Poon model using measured device characteristics. Furthermore, MODELLA facilitates a better intuitive understanding of device behavior due to its close link with device physics and to the physical significance of its parameters.

References

[1] I. Kidron: Int. J. Electr., Vol.31, 421 (1971).

[2] H.C. de Graaff, F.M. Klaassen: Compact transistor modelling for circuit design, Springer-Verlag (1990).

[3] Lindmayer and Wrigley: Fundamentals of Semiconductor Devices, p. 243, Van Nostrand, 1968.

[4] H.C. de Graaff, W.J. Kloosterman: IEEE Trans. Electr. Dev. ED-32, 2415 (1985).

[5] W.M. Webster: Proc. IRE, 42, 914 (1954).

[6] J.M. Early: Proc. IRE, 40, 1401 (1952).

[7] PANACEA: Internal Philips circuit simulator.

© N.V. Philips' Gloeilampenfabrieken

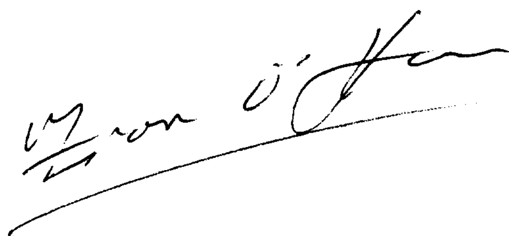
Acknowledgements

The work described in this thesis has been submitted to Trinity College Dublin, as partial fulfillment of an M.Sc. degree by research. In this context, I would firstly like to thank Dr. Brian Foley, my project supervisor at Trinity College Dublin, for making this M.Sc. project possible. His active support and encouragement throughout the project are greatly appreciated. Dr. Henk de Graaff kindly accepted the role of project supervisor at Philips and his clear technical insight and constructive advice are also appreciated.

A warm thanks to my former and current group leaders at Philips Research Laboratories, Dr. Paul Hart and Drs. Wil Josquin, for their support throughout the project and to Drs. W. Gelling for agreeing to the Trinity College involvement.

Many thanks to Jan van den Biesen for all the help and advice in the initial stages of the project.

I would also like to express my gratitude to many members of the group for their help and advice, especially Fred Hurkx, Willy Kloosterman and Martin Versleijen. Thanks also to Merten Koolen, Hans Geelen and Hans Timmerman for providing all the measurements appearing in this thesis and, finally, to Mari Knuvers and Ad van Run for keeping me sane through it all.



Fran O'Hara
Eindhoven, November 1990,
Philips Research Laboratories,
Nederlandse Philips Bedrijven B.V.

© N.V. Philips' Gloeilampenfabrieken