MASTER

Simulation and optimization of a polysilicon emitter BICMOS process

Ahlrichs, F.W.

Award date: 1997

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

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

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
Simulation and optimization of a polysilicon emitter BICMOS process

F.W. Ahlrichs

TTE nr.: 544

Concerns: Graduation report
Faculty of Electrical Engineering
Eindhoven University of Technology
Department: Telecommunication Technology and Electromagnetics (TTE)

Place: Philips Semiconductors Nijmegen
Period of work: April - October 1997

Coaches: Ir. Som Nath, Philips Semiconductors Nijmegen
Ir. Anco Heringa, Philips Research Laboratories Eindhoven
Prof. Dr. F.M. Klaassen, TUE

Graduation coach: Prof Dr. L.M.F. Kaufmann, TUE

© Philips Electronics N.V., 1997
All rights are reserved. Reproduction in whole or in part is prohibited without the written consent of the copyright owner.
Abstract

Analysis of process fluctuations is critical in developing manufacturable technologies with minimal variations in electrical characteristics and in the determination of worst-case process descriptions. Physical device and process modeling can be used to identify and reduce the variations in the electrical characteristics due to fluctuations in processing. To perform such analysis correctly for the first-order transistor parameters such as the base current, the collector current and the transistor gain (Hfe) inclusion of physical phenomena like bandgap narrowing, intrinsic carrier concentration, minority hole and electron mobilities, Auger and Shockley-Read-Hall recombination and the modeling of the polysilicon/single-crystalline silicon interface are critical. Inconsistent use of these parameters in the device simulations leads to incorrect design and interpretation of experiments.

This project describes the process and device simulations for a polysilicon emitter bipolar transistor used in a BICMOS process. In the modeling the threshold adjust boron implantation appears to have much effect on the electrical behavior of the bipolar part of the process. Finally acceptable simulation results were found taking all earlier mentioned effects into account. Furthermore, for statistical modeling purpose it is necessary to predict the sensitivity of an electrical parameter to variations in process conditions such as the base implant dose. Once the process and device simulation showed acceptable results a statistical experiment was done to check the variations in the critical process parameters such as the base implant dose and energy, the etch back of silicon underneath the polysilicon and the temperature of the RTA emitter drive in. From this experiment, sensitivity curves were found which predict the fluctuation of the electrical parameters to variations in the above 4 process parameters, which will be very helpful for process engineers in a production environment.
# Contents

I. Introduction 9

II. Polysilicon emitters 10

2.1 Introduction 10
2.2 Device fabrication 10
2.3 Device modeling 12
  2.3.1 Introduction 12
  2.3.2 Modeling of the polycrystalline silicon region 13
  2.3.3 Modeling of the polysilicon/silicon interface 14
  2.3.4 Modeling of the single-crystalline silicon region and determination of the base current of a polysilicon emitter bipolar transistor 17

III. Physical models in device simulations 18

3.1 Introduction 18
3.2 Bandgap narrowing 18
3.3 Carrier mobility 19
  3.3.1 Introduction 19
  3.3.2 Mobility due to lattice scattering 19
  3.3.3 Mobility due to majority impurity scattering 20
  3.3.4 Mobility due to minority impurity scattering 21
  3.3.5 Mobility due to carrier-carrier scattering 22
  3.3.6 Total carrier mobility 23
3.4 Recombination 23
  3.4.1 Auger and Shockley-Read-Hall recombination 23
  3.4.2 Surface recombination 25

IV. Process and device simulation 27

4.1 Introduction 27
4.2 Process simulation 27
  4.2.1 Introduction 27
  4.2.2 SIMS and simulated profiles 28
    4.2.2.1 Introduction 28
    4.2.2.2 Simulated profiles 30
  4.2.3 Sheet resistances 34
  4.2.4 The role of the interfacial oxide break-up model in SUPREM-4 36
4.3 Device simulation 37
  4.3.1 Introduction 37
  4.3.2 Electrical characteristics 37
  4.3.3 Junction capacitances 42
4.4 Geometry dependence of the simulations 43
V. Statistical experiment

5.1 Introduction 45
5.2 Design of experiments 45
5.3 Analysis of the experiment results 47
  5.3.1 Introduction 47
  5.3.2 Base current 48
    5.3.2.1 The base current model based on a Vbe between 0.4 and 1.0 V 48
    5.3.2.2 Conclusions 50
    5.3.2.3 The base current model based on a Vbe between 0.6 and 0.7 V 52
    5.3.2.4 Conclusions 53
  5.3.3 The collector current 55
    5.3.3.1 The collector current model based on a Vbe between 0.4 and 1.0 V 55
    5.3.3.2 Conclusions 56
  5.3.4 The Hfe 58
    5.3.4.1 The Hfe model based on a Vbe between 0.4 and 1.0 V 58
    5.3.4.2 Conclusions 59

VI. General conclusions and recommendations for future work 61

References 62

Acknowledgment 64
I. Introduction

The major objective of this work is to understand what the influence is of variations in processing conditions on the electrical parameters of a polysilicon emitter bipolar transistor. To do this in a right way the help of process and device simulation programs is needed, such as SUPREM-4 and MEDICI. To get good results from the simulation programs a good understanding of the different physical mechanisms that play a part in the behavior of a polysilicon emitter bipolar transistor, is needed. Once the simulation results are matched to measurement results and the physical models are used in a right way, a statistical experiment can be set up to study the influence of variations in processing conditions on the electrical behavior of the device.

Not much work has been done in the past on setting up such an experiment for bipolar transistors, therefore this project was started. Good results and especially the outcome of the statistical experiment will produce sensitivity curves which are very helpful for process engineers. With the help of those sensitivity curves one is able to understand the influence of variation of a specific process parameter on the electrical behavior of the device. When used in a correct way these sensitivity curves will save a lot of money and time in a manufacturing environment.

In this report first of all the theory behind a polysilicon emitter bipolar transistor is described. Especially the interface between the polysilicon and the silicon layer plays a crucial role in the behavior of a polysilicon emitter transistor. The next step is to carefully determine which physical mechanisms play a significant role in modeling such a device. With a good physical basis, process and device simulations can be carried out.

The process and device simulations are well described in this project. All kind of effects on the electrical behavior of the transistor are well defined. As a result the simulations appeared to be correct and were matched to the measurement results.

Finally with the help of good process and device simulation input decks a statistical experiment was set up using NORMAN/DEBORA. This experiment was set up to study the effect of variations in four important process parameters on the electrical parameters of the transistor like the base current, collector current and gain (Hfe). The results of this experiment provided very useful sensitivity curves. The data for these curves was analysed using the STATISTICA analysis package.
II. Polysilicon emitters

2.1 Introduction

Polysilicon, to be complete polycrystalline silicon, is a material that consists of small, randomly oriented grains of single-crystalline silicon separated by disordered regions known as grain boundaries. Since the advent of polysilicon emitters in the 1970’s they are now commonly used in IC manufacturing processes. The major reason for this is their suitability for producing shallow emitter/base junctions necessary because lateral as well vertical dimensions will still require scaling for future technologies. By scaling the device, the peripheral component of the emitter/base capacitance is kept at a reasonable level. Another important aspect of polysilicon emitters is the increase of current gain by a factor 3-30 compared to diffused emitters, depending upon the fabrication conditions of the polysilicon emitter process [1].

Up to now there is some controversy over the exact mechanisms that control the current gain in polysilicon emitter bipolar transistors. Mechanisms that play a significant role in the modeling of these transistors, will be discussed in the next sections.

2.2 Device fabrication

The technology used for the process is a 0.6 μm CMOS technology combined with polysilicon emitter bipolar technology. This section only describes the polysilicon emitter bipolar part of the process. The bipolar is a vertical double base NPN transistor with a polysilicon emitter with a minimum emitter size of 0.9 * 0.9 (μm²). Initial substrate is p-type <100> oriented 10 Ω cm boron doped silicon. First an arsenic implant of 3.5e15 at.cm⁻² at 70 KeV is done to create a buried-N layer. After this a 1.8 μm thick epi-layer is grown with a phosphorus concentration of 7e15 at.cm⁻³. Next an SP-implant takes place BF2 of 4e15 at.cm⁻² at 40 KeV. To create the base of the NPN transistor a SPB implantation with a BF2 dose of 2.4e13 at.cm⁻² at 40 Kev is done. Upon this base region the processing of the polysilicon emitter takes place.

Processing of the polysilicon emitter in BICMOS technology is comprised of emitter window definition, chemical cleaning of the silicon surface prior to the deposition and doping of the polysilicon layer and rapid thermal annealing (RTA) to drive the emitter and to break up the native oxide at the poly/single-crystal silicon interface.

An etch in hydrofluoric acid (HF) is done to remove all oxide. After this a 150 nm polysilicon layer is formed at 620 °C from SiH₄ decomposition in a LPCVD reactor. Arsenic 5e15 at.cm⁻², 30 KeV is implanted into the polysilicon to dope the emitter of the bipolar transistor (figure 2.1). With the deposition of a polysilicon layer comes the growth of a thin continuous interfacial oxide layer approximately 7 Å thick. This interfacial oxide layer at the polysilicon/silicon interface plays a very significant role in the behavior of polysilicon emitter bipolar transistors [2].

An RTA-drive is performed at a temperature of 1075 °C for 12 seconds to activate the arsenic in the polysilicon and to break up the interfacial oxide layer. Arsenic diffuses
into the epi-silicon and forms an n’ layer in the base of the transistor. This gives an emitter/base junction reaching 60 nm deep into the epi-silicon starting from the polysilicon/epi-silicon interface. This thin emitter layer in the epi-silicon is essential to obtain a useful characteristics determined by the well known p-n junction current formulas [6].

Due to the rapid thermal annealing step the interfacial oxide layer breaks up. The rate at which the interfacial oxide layer breaks up depends critically on the initial thickness of the interfacial layer, the emitter doping concentration and the temperature. Because of these three factors, especially the high temperature (1075 °C) used for this process it is assumed that the interfacial oxide layer will break up and form a discontinuous layer varying in thickness from 0 to 8 Å, at even higher temperatures the interfacial oxide layer will break up completely and the oxide will form itself into spherical balls at the interface (figure 2.2) [1]. Nowadays it is possible with the help of TEM measurements to get insight into the interface structure.

Fig. 2.1: Cross section of a double base NPN polysilicon emitter bipolar transistor

Fig. 2.2: The process of the interfacial oxide break-up due to thermal annealing.
2.3 Device modeling

2.3.1 Introduction

In recent years the physics of polysilicon emitter bipolar transistors have been extensively studied. In spite of this widespread study, there is up to now still considerable controversy about the physical mechanisms that control the current gain of such a device. From literature it is known that the minority carrier injection into a polysilicon emitter is controlled by different complex processes: hole transport and recombination in the monocrystalline region, hole transport across the poly/singlecrystalline silicon interface and hole transport and recombination in the polysilicon [3]. By far the most difficult process to characterize is the hole transport across the poly/singlecrystalline silicon interface. This is due to the fact that the structural design of the interface critically depends on process conditions. As a consequence, the hole transport across the interface will be described by different physical mechanisms depending on how the structure of the interface looks. These mechanisms can be summarized as follows [4]:

1) Oxide tunneling model

In devices with an interfacial oxide layer, holes injected from the base into the emitter are forced to tunnel through this oxide layer, thereby blocking minority-carrier flow into the polysilicon and hence producing a reduction in base current. Also the majority-carrier electrons in the polysilicon have to tunnel through this oxide towards the singlecrystalline silicon region. This limits the maximum current carrying capability of the device, since it gives rise to an increase in emitter resistance.

2) Grain boundary mobility model

This model explains the improved gain in polysilicon emitters by the reduced mobility at the grain boundaries in the polysilicon as well as at the pseudo-grain boundary (disordered interfacial region) at the poly/singlecrystalline silicon interface. The base current is suppressed when the density of interface trapping states at the grain boundaries is small and is increased when the density of interface trapping states is high at the grain boundaries. The amount of interface traps determines the value of the recombination that will take place at the interface and hence influences the base current.

3) Segregation model

This model explains the improved gain due to dopant segregation at the polysilicon/silicon interface. Due to this pile-up of dopant a potential barrier occurs at the interface. The presence of this potential barrier will give a reduction in base current and hence a higher gain.

All three mechanisms can be considered as blocking mechanisms, because they show an improved gain by suppressing the base current. Due to these three different physical mechanisms, which can occur at the polysilicon/silicon interface, the modeling of a
polysilicon emitter transistor is very complex. An effective way to model such a device is by using the so-called effective recombination velocity (ERV) method.

The ERV-method describes each process which has influence on the hole current by means of a single parameter: the effective recombination velocity. The advantage of this method in comparing it with other methods is that it allows separate modeling of the polycrystalline silicon region, the polysilicon/silicon interface and the singlecrystalline silicon region [1],[3]. Therefore it is possible to combine detailed models for the polysilicon region and singlecrystalline silicon region with the models described earlier for the polysilicon/silicon interface, thus avoiding any unnecessary simplifications.

2.3.2 Modeling of the polycrystalline silicon region

The polycrystalline silicon layer consists of singlecrystalline grains separated by disordered grain boundaries each characterized by an effective recombination velocity (ERV) see figure (2.3). The polysilicon layer thickness is \( W_{\text{ep}} \), bounded at the left side in figure (2.3) by the metal contact and on the other side by the interfacial oxide layer \( \Delta \). In between the polysilicon layer, considered from a one-dimensional point of view, the structure consists of singlecrystalline grains with length/thickness \( d_g \) and grain boundary length/thickness \( \delta \). The entire grain length \( d \) is therefore equal to \( d_g + \delta \). Recombination at the metal contact is characterized by the recombination velocity \( S_M \) eqn. (2.1) [3].

\[
S_M = \sqrt{\frac{kT}{2m_{bps}^*}}
\]

\( k \) = Boltzmann’s constant
\( T \) = Temperature in Kelvin
\( m_{bps}^* \) = Hole effective mass in polysilicon
The next step is to calculate the effective recombination velocity $S_{g,i}$ (see figure 2.3) relative to the i-th grain which can be calculated using eqn. (2.2).

$$S_{g,i} = \frac{v_d + S_{gb,i} \tanh(\frac{d}{L_{pg}})}{v_d + S_{gb,i} \tanh(\frac{d}{L_{pg}})}$$

$v_d = \text{Diffusion velocity} = \frac{D_{pg}}{L_{pg}}$
$D_{pg} = \text{Hole diffusion coefficient in the grain}$
$L_{pg} = \text{Hole diffusion length in the grain}$

The ERV $S_{gb,i}$ relative to the i-th grain boundary is calculated using eqn. (2.3).

$$S_{gb,i} = S_{gb} + \frac{1}{\frac{1}{T_{gb}} + \frac{1}{S_{gb,i-1} + S_{gb}}}$$

$$T_{gb} = \frac{kT \mu_{pgb}}{q\delta}$$

$S_{gb} = \frac{1}{2} v_{th} c_p N_{st}$ = Recombination velocity at the interface between the grains and grain boundaries
$\mu_{pgb} = \text{Hole mobility in the grain boundary}$
$q = \text{Absolute value of the electronic charge}$
$N_{st} = \text{Density of trapping states at the grain boundary}$
$c_p = \text{capture cross section of holes}$
$v_{th} = \text{Thermal velocity}$

Starting from the metal contact on the left side of figure (2.3) calculation of $S_{g,1}$ can be done by substituting $S_M$ eqn. (2.1) into eqn. (2.2). Then by using eqn. (2.3) the calculations of all other effective recombination velocities can be determined. By doing this finally the value of $S_{poly} = S_{g,N}$ (where $N$ is the number of grains and grain boundaries) can be found. The way described in this section to calculate $S_{poly}$ for a multi grain model is based on the theory for a single grain model. Therefore this model is a start for modeling a polysilicon layer with more than one grain, but the above described model needs to be updated in the future. The value of $S_{poly}$ accounts for the hole injection into the polysilicon, and can therefore be used in the interface models discussed in the next section [3].

2.3.3 Modeling of the polysilicon/silicon interface

Due to the process of depositing polysilicon on the emitter window of the substrate a thin interfacial oxide layer at the polysilicon/silicon interface is introduced. This layer consists of silicon dioxide. Due to the fact that this material has a wider bandgap than silicon, this interfacial oxide layer forms a potential barrier for both electrons and holes.

Figure (2.4) shows the energy band diagram of a polysilicon emitter bipolar transistor with a thin insulator between the polysilicon and the silicon regions [5]. A rectangular potential barrier has been assumed for the interfacial oxide layer and $X_c$ and $X_h$ are the
potential barriers to electrons and holes respectively. Carriers pass through this interfacial oxide layer mainly by tunneling. When the potential barrier is less than a few kT then thermionic emission has to be included in the tunneling mechanism. An oxide layer is expected to give a potential barrier much higher than a few kT, so thermionic emission can be neglected. The polysilicon/silicon interface in this section is described by an ERV $S_p$ which accounts for hole transport across the interface $T_{block}$, hole injection into the polysilicon $S_{poly}$ and surface recombination at the interface $S_I$ eqn. (2.4).

\[ S_p = S_I + \frac{1}{\frac{1}{T_{block}} + \frac{1}{S_I + S_{poly}}} \]  

Equation (2.4) describes the effect of the polysilicon layer and the interface together by one parameter: the effective recombination velocity $S_p$. The parameter $S_{poly}$ equal to $S_{gN}$ is described in the previous section by equations (2.1), (2.2) and (2.3). This parameter defines the hole injection into the polysilicon layer and can therefore be seen in figure (2.3) at the interfacial-oxide/polysilicon interface. Parameter $S_p$ is found in figure (2.3) at the interfacial-oxide/singlecrystalline-silicon interface. For very thin interfacial oxide layers $S_p$ is approximately equal to $S_{poly}$, which indicates that the hole transport properties of the polysilicon layer dominate the device behaviour. Alternatively, for very thick interfacial oxide layers $S_p$ is approximately equal to $S_I$, which indicates that the surface recombination at the interface will dominate device behavior [1].

The term $T_{block}$ from eqn. (2.4) describes the blocking mechanism of the polysilicon/silicon interface. This blocking mechanism depends on the model used. As stated earlier in section 2.3.1 there are three different models each describing a
different physical mechanism with each having a different value for the term $T_{\text{block}}$. The expressions of $T_{\text{block}}$ for the different models are summarized in table 2.1 [4].

TABLE 2.1: Analytical expressions used for $T_{\text{block}}$ for the three different interface mechanisms.

<table>
<thead>
<tr>
<th>Blocking model:</th>
<th>Expression used for $T_{\text{block}}$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxide tunneling model</td>
<td>$\exp\left(-b_h\right) \sqrt{\frac{2\pi \Delta}{\hbar}} \frac{2\pi \Delta}{\hbar} \frac{2m_{hi}}{X_h}$, $h = \text{Planck's constant, } m_{hi} = \text{the effective mass of holes in the oxide, } m = \text{the effective mass of holes in silicon, } N_{E} = \text{electrically active doping density in the emitter, } N_{int} = \text{electrically active doping density of segregated doping at the interface.}$</td>
</tr>
<tr>
<td>Grain boundary mobility model</td>
<td>$\frac{kT \mu_{pgb}}{q \delta}$</td>
</tr>
<tr>
<td>Segregation model</td>
<td>$\frac{kT}{2\pi m_{hs} N_{int}} N_{E}$</td>
</tr>
</tbody>
</table>

The expression used for $T_{\text{block}}$ in the grain boundary mobility model is valid when a uniform doping concentration is assumed for the doping within the interfacial layer. However, due to dopant pile up, the doping distribution in the interfacial oxide layer is not uniform. The expression for $T_{\text{block}}$ will more generally be described by eqn. (2.5) [3].

$$T_{\text{block}} = \frac{1}{p_{0}(x_{p}) \int_{x_{p}}^{x_{s}} \frac{q}{kT \mu_{pgb} p_{0}(x)} dx}$$

$p_{0}(x) = \text{The hole concentration at thermal equilibrium}$

$x_{p}, x_{s} = \text{Boundary values of the interface (see figure 2.3)}$

In our device there is an interfacial oxide layer so the expression for $T_{\text{block}}$ to use is then the expression for the oxide tunneling model.
2.3.4 Modeling of the single-crystalline silicon region and determination of the base current of a polysilicon emitter bipolar transistor

Due to physical effects caused by high doping such as band gap narrowing, Auger and Shockley-Read-Hall recombination determination of the current injected into the single-crystalline silicon emitter region of a bipolar transistor is complex. These effects will be discussed in the next chapter in combination with their effect on the device simulations. In order to relate the effective recombination velocity \( S_p \) eqn. (2.4) to the base current, it is necessary to include recombination in the single-crystalline silicon emitter part of the transistor. The base current can now be accurately modeled, for the case of a shallow emitter profile, by the use of the approximate expression in eqn. (2.6a) [1].

\[
I_B = \frac{qA n_{i0}^2}{G_{eff}(W_E) + \frac{N_{Def}(W_E)}{S_p}} \times \\
\left[ 1 + \int_0^{w_p} \left( \frac{G_{eff}(W_E) - G_{eff}(x)}{\tau_{pe}(x) N_{Def}(x)} \right) dx + \frac{N_{Def}(W_E) w_p}{S_p} \int_0^{w_p} \frac{dx}{\tau_{pe}(x) N_{Def}(x)} \right] \exp \left( \frac{qV_{be}}{kT} \right)
\]

\( A = \) Surface area of the emitter/base junction
\( n_{i0} = \) Intrinsic equilibrium carrier density in the absence of band gap narrowing
\( W_E = \) Neutral emitter region thickness

\[
N_{Def}(x) = \frac{n_{i0}^2}{n_a^2(x)} N_D(x) = \text{Effective doping concentration in the emitter} \quad (2.6b)
\]

\[
n_a^2(x) = n_{i0}^2 \exp \frac{\Delta E_g}{kT} = \text{Effective intrinsic equilibrium carrier density} \quad (2.6c)
\]

\( \Delta E_g = \) The amount of band gap narrowing
\( \tau_{pe} = \) Hole lifetime in the emitter

\[
G_{eff}(x) = \int_0^{w_p} N_{Def}(x) dx = \text{Effective Gummel number} \quad (2.6d)
\]

\( V_{be} = \) Emitter / base voltage

The expression for the base current eqn. (2.6) now includes transport and tunneling mechanisms and recombination at the polysilicon/silicon interface due to the effective recombination velocity \( S_p \) and recombination in the single-crystalline silicon emitter part of the transistor by the term in square brackets.


III. Physical models in device simulations

3.1 Introduction

The results of device simulations critically depend on the physical models used in the device simulator. It is therefore very important to consider which model to use before starting any device simulation. In the next sections high doping effects and their use within device simulations are discussed.

3.2 Bandgap narrowing

It is observed for silicon that for high doping concentrations larger than $1 \times 10^{17} \text{ cm}^{-3}$ the measured current is higher than expected from the measured doping, minority carrier lifetime and minority carrier mobility. This effect could be explained by actual narrowing of the energy bandgap, a change in the density of states in the valence band, non ionization of some of the impurity atoms or, what seems most reasonable is a combination of all three. The way it manifests itself physically is by showing an increase in the effective intrinsic equilibrium carrier concentration $n_{ie}$ eqn. (3.1) [13].

$$n_{ie}^2(x) = n_{i0}^2 \exp \left( \frac{q\Delta E_g}{kT} \right)$$  (3.1)

$\Delta E_g$ = The amount of band gap narrowing

$n_{i0}$ = Intrinsic equilibrium carrier density in the absence of band gap narrowing

The amount of bandgap narrowing depends only on the doping concentration and not on the temperature and is described by eqn. (3.2).

$$\Delta E_g(N) = V_1 \left\{ \ln \left( \frac{N}{N_2} \right) + \sqrt{\ln \left( \frac{N}{N_2} \right)^2 + C_2} \right\}$$  (3.2)

$N$ = Doping concentration

In 1976 Slotboom and de Graaff described for the first time an empirical expression (3.2), which was derived from measurements of collector currents of npn-bipolar transistors, for p-type silicon [7]. Nine years later in 1985 del Alamo et al. came up with new values for the parameters $V_1$, $N_2$ and $C_2$ for describing the amount of bandgap narrowing in n-type silicon [8]. This expression yielded values about 32 mV lower than those described by Slotboom and de Graaff in 1976. The work done by Slotboom and de Graaff was verified by Swirhun et al. in 1986 [9].

The minority carrier mobility plays a crucial role in the interpretation of most of the experiments on calculating the amount of bandgap narrowing. Klaassen et al. came up with a new mobility model that was used to recalculate previous experiments on the bandgap narrowing. With the help of this model in 1992 Klaassen et al. published data for the amount of bandgap narrowing in p-type silicon as well as n-type silicon [10].
Both are described by the same set of parameters see table (3.1). The mentioned mobility model is described in the next section [10],[11],[12].

TABLE 3.1: Model parameters to calculate the amount of bandgap narrowing in eqn. (3.2).

<table>
<thead>
<tr>
<th></th>
<th>Slotboom and de Graaf (p-type)</th>
<th>del Alamo et al. (n-type)</th>
<th>Klaassen et al. (n-type, p-type)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1$ (mV)</td>
<td>9.0</td>
<td>9.35</td>
<td>6.92</td>
</tr>
<tr>
<td>$N_2$ (cm$^3$)</td>
<td>1e17</td>
<td>7e17</td>
<td>1.3e17</td>
</tr>
<tr>
<td>$C_2$</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The values for the parameters $V_1$, $N_2$ and $C_2$ given by Klaassen et al., have been used in the simulations described in this project.

3.3 Carrier mobility

3.3.1 Introduction

The mobility model described in this section takes into account the main contributions to mobility reduction such as: lattice scattering, donor and acceptor scattering and electron-hole scattering. Beside these effects screening of impurities by charge carriers and the temperature dependence of both majority and minority carrier mobility are included. The model described in this section gives the carrier mobility as a function of the donor, acceptor, electron and hole concentrations and of the temperature. This makes the model especially suited for device simulation purposes.

3.3.2 Mobility due to lattice scattering

At very low impurity concentrations the only scattering mechanism is lattice scattering. The electron mobility due to lattice scattering $\mu_{e,L}$ as well as the hole mobility due to lattice scattering $\mu_{h,L}$ are both given by eqn. (3.3).

$$\mu_{i,L} = \mu_{\text{max}} \left( \frac{300}{T} \right)^{\theta_i} \quad (3.3)$$

The subscript $i$ stands for $e$ (electron) or $h$ (hole) while the parameters $\mu_{\text{max}}$ and $\theta_i$ can be found in table (3.2).
3.3.3 Mobility due to majority impurity scattering

The mobility due to majority impurity scattering: the electron mobility due to donor scattering $\mu_{e,D}$ and the hole mobility due to acceptor scattering $\mu_{h,A}$, is given by eqn. (3.4a).

$$\mu_{i,I}(N_I) = \mu_{i,N}(\frac{N_{ref1}}{N_I})^{\alpha_1} + \mu_{i,c} \quad (3.4a)$$

with

$$\mu_{i,N} = \frac{\mu_{max}^2}{\mu_{max} - \mu_{min}} \left( \frac{T}{300} \right)^{3\alpha_1 - 1.5} \quad (3.4b)$$

and

$$\mu_{i,c} = \frac{\mu_{min}\mu_{max}}{\mu_{max} - \mu_{min}} \sqrt{\frac{300}{T}} \quad (3.4c)$$

The subscript $(i,I)$ in eqn. (3.4) stands for $(e,D)$ or $(h,A)$ while $N_D$ and $N_A$ are the donor and acceptor concentrations. The parameters $\mu_{max}$, $\mu_{min}$, $N_{ref1}$ and $\alpha_1$ can be found in table (3.2).

**TABLE 3.2: Model parameters to calculate the carrier mobility.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Electrons (As)</th>
<th>Holes (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{max}$ (cm$^2$ V$^{-1}$ s$^{-1}$)</td>
<td>1417.0</td>
<td>470.5</td>
</tr>
<tr>
<td>$\mu_{min}$ (cm$^2$ V$^{-1}$ s$^{-1}$)</td>
<td>52.2</td>
<td>44.9</td>
</tr>
<tr>
<td>$N_{ref1}$ (cm$^3$)</td>
<td>9.68e16</td>
<td>2.23e17</td>
</tr>
<tr>
<td>$N_{ref2}$ (cm$^3$)</td>
<td>4e20</td>
<td>7.2e20</td>
</tr>
<tr>
<td>$\alpha_1$ (cm$^{-3}$)</td>
<td>0.68</td>
<td>0.719</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.21</td>
<td>0.40</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>2.285</td>
<td>2.247</td>
</tr>
</tbody>
</table>

At high carrier concentrations carriers tend to screen impurities from other carriers. This physical process is called screening. The theory which starts from a screened Coulomb potential requires a collision cross-section for impurity scattering that depends only on the carrier concentration $c$. When taking the effect of screening into account the mobility due to majority impurity scattering is then given by eqn. (3.5).
\[ \mu_{i,N}(N_i) = \mu_{i,N}(\frac{N_{\text{ref},i}}{N_i}) + \mu_{c,e} \frac{c}{N_i} \]  

(3.5)

From eqn. (3.5), because the second term in eqn. (3.5) is dominant at high impurity concentrations, it can be seen that due to screening at high impurity concentrations the mobility due to majority impurity scattering increases with the carrier concentration.

Ultra-high impurity concentrations have an effect on the carrier mobility. Above an impurity concentration of 1e20 cm\(^{-3}\) the carriers are no longer scattered by impurities possessing one electronic charge and a concentration \(N_i\), but by impurities with \(Z\) electronic charges and a cluster concentration \(N_i' = N_i/Z\). These ultra-high concentration effects can now be taken into account by replacing \(N_i\) in eqn. (3.5) by \(Z \times N_{i,s}\) where \(N_{i,s}\) accounts for the ionized impurity concentrations. Solving for \(Z\) at each impurity concentration \(N_{i,s}\) yields the clustering function \(Z(N_{i,s})\) given by eqn. (3.6a).

\[ Z_i(N_{i,s}) = 1 + \frac{1}{c_i + (\frac{N_{\text{ref},i}}{N_{i,s}})^2} \]  

(3.6a)

\[ N_i' = Z_i N_{i,s} \]  

(3.6b)

The values for the parameters \(N_{\text{ref},i}\) and \(c_i\) can be found in table (3.2).

3.3.4 Mobility due to minority impurity scattering

The mobility due to the minority impurity scattering, the electron mobility due to acceptor scattering \(\mu_{e,A}\) and the hole mobility due to donor scattering \(\mu_{h,D}\) are given by eqns. (3.7a/b).

\[ \mu_{e,A} = \frac{\mu_{e,D}}{G(P_e)} \]  

(3.7a)

\[ \mu_{h,D} = \frac{\mu_{h,A}}{G(P_h)} \]  

(3.7b)

The subscript e and h of P indicate that the effective mass of electrons or holes has to be used in the equation for \(G(P)\). The effective mass for electrons \(m_e\) used in the calculations and in the simulations is equal to \(m_0\) while the effective mass for holes \(m_h\) is equal to 1.258 \(m_0\). The parameter \(m_0\) is called the free electron rest mass. The ratio \(G(P)\) is defined as the collision cross-section for repulsive screened Coulomb potentials divided by the attractive screened Coulomb potentials and is given by eqn. (3.8).

\[ G(P_e) = \frac{\sigma_{s,rep}}{\sigma_{s,attr}} = 1 - \left[ s_1 \left( \frac{m_0 T}{m_e 300} \right)^{\nu_e} P_i \right]^{\nu_e} + \left[ \left( \frac{m_e 300}{m_0 T} \right)^{\nu_e} P_i \right]^{\nu_e} \]  

(3.8)
The values for the constants $s_1$ - $s_7$ are given in table (3.3). The parameter $P_i$, that takes the physical effect of screening into account, is given by eqn. (3.9).

$$P_i = \left[ \frac{2.459}{3.97e13(N_i^*)^{-1}} + \frac{3.828}{1.36e20m_i \text{ cm}_0} \right]^{-1} \frac{T}{300}$$

(3.9)

**TABLE 3.3 :** Numerical values for the constants $s_i$ and $r_i$.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$s_i$</th>
<th>$r_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.89233</td>
<td>0.7643</td>
</tr>
<tr>
<td>2</td>
<td>0.41372</td>
<td>2.2999</td>
</tr>
<tr>
<td>3</td>
<td>0.19778</td>
<td>6.5502</td>
</tr>
<tr>
<td>4</td>
<td>0.28227</td>
<td>2.3670</td>
</tr>
<tr>
<td>5</td>
<td>0.005978</td>
<td>-0.8552</td>
</tr>
<tr>
<td>6</td>
<td>1.80618</td>
<td>0.6478</td>
</tr>
<tr>
<td>7</td>
<td>0.72169</td>
<td></td>
</tr>
</tbody>
</table>

### 3.3.5 Mobility due to carrier-carrier scattering

The last scattering mechanism that contributes to the mobility reduction is carrier-carrier scattering. From all carrier-carrier scattering mechanisms the electron-hole scattering mechanism contributes most to the carrier mobility. Electron-electron and hole-hole scattering represent only a second-order effect on the mobility. The electron mobility due to hole scattering $\mu_{e,h}$ and the hole mobility due to electron scattering $\mu_{h,e}$ are given by eqns. (3.10a/b).

$$\mu_{e,h} = F(P_i)\mu_{e,D}$$

(3.10a)

$$\mu_{h,e} = F(P_i)\mu_{h,D}$$

(3.10b)

The mobility ratio $F(P_i)$ between stationary secondary scatterers with infinite mass and moving secondary scatterers with finite mass is given by eqn. (3.11).

$$F(P_i) = \frac{r_i P_i + r_2 + r_3 \frac{m_i}{m_j}}{P_i + r_4 + r_5 \frac{m_i}{m_j}}$$

(3.11)

In eqn. (3.11) $m_i$ is the effective carrier mass of the primary scatterer while $m_j$ is the effective mass of the secondary scatterer. This implies that if $i = e$ then $j = h$ and if $i = h$ then $j = e$. The values for the constants $r_1$ - $r_6$ are given in table (3.3).
3.3.6 Total carrier mobility

All scattering mechanisms described in the previous sub sections contribute to the total carrier mobility $\mu_i$. The total carrier mobility for electrons $\mu_{i_e}$ and for holes $\mu_{i_h}$ can now be described as a function of the donor, acceptor, electron and hole concentration by eqn. (3.12a).

$$\mu_i^{-1} = \mu_{i,L}^{-1} + \mu_{i,D}^{-1} + \mu_{i,A}^{-1} + \mu_{i,j}^{-1}$$  \hspace{1cm} (3.12a)

where

$$\mu_{i,L}^{-1} + \mu_{i,A}^{-1} + \mu_{i,j}^{-1} = \mu_{i,D}^{-1} + \alpha + \beta$$  \hspace{1cm} (3.12b)

With again the same meaning for the subscripts $i$ and $j$ as stated earlier. The last three terms in eqn. (3.12a) have been derived in this section for the situation where there is only one scattering partner, indeed there are three scattering partners each time for electrons as well as for holes. This gives the following expression for $\mu_{i,D+A+j}$ eqn. (3.13a):

$$\mu_{i,D+A+j} = \mu_{i,N} \frac{N_{i,sc} / N_{i,sc,eff}}{N_{i,sc} / N_{i,sc,eff}} \left( \frac{N_{i,sc} / N_{i,sc,eff}}{N_{i,sc} / N_{i,sc,eff}} \right)^{\gamma_1} + \mu_{i,s} \left( \frac{n + p}{N_{i,sc,eff}} \right)$$  \hspace{1cm} (3.13a)

where

$$N_{i,sc} = N_{i,D}^* + N_{i,A}^* + j$$  \hspace{1cm} (3.13b)

and

$$N_{s,sc,eff} = N_{i,D}^* + G(P_s)N_{i,A}^* + \frac{p}{F(P_s)}$$  \hspace{1cm} (3.13c)

and

$$N_{h,sc,eff} = N_{i,A}^* + G(P_h)N_{i,D}^* + \frac{n}{F(P_h)}$$  \hspace{1cm} (3.13d)

3.4 Recombination

3.4.1 Auger and Shockley-Read-Hall recombination

An important effect which occurs at high doping concentrations in silicon is Auger recombination. Due to Auger recombination a reduction in the carrier lifetime is obtained. There are two cases of Auger recombination. The first case is a recombination between an electron and hole, accompanied by the transfer of energy to another free hole. The second case is a recombination between an electron and hole that results in a transfer of energy to a free electron. The third free particle involved in this process will lose its energy to the lattice in the form of heat. The process involving two electrons and a hole occurs primarily in a heavily doped n-type material thus in our process this would occur only in the emitter part in the epi-silicon with
arsenic concentrations above 1e20 at.cm\(^{-3}\). The Auger recombination is given by eqn. (3.14).

\[
R_{\text{Auger}} = C_{\text{Au,n}}(pn^2 - nn_n^2) + C_{\text{Au,p}}(np^2 - pn_p^2)
\]  

(3.14)

The values of the Auger recombination coefficients \(C_{\text{Au,n}}\) and \(C_{\text{Au,p}}\) can be found in table (3.4) [12].

**TABLE 3.4: Parameter values for Auger and SRH recombination.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C_{\text{Au,n}}) (cm(^6)s(^{-1}))</td>
<td>2.78e-31</td>
</tr>
<tr>
<td>(C_{\text{Au,p}}) (cm(^6)s(^{-1}))</td>
<td>1.83e-31</td>
</tr>
<tr>
<td>(\tau_{0,n}) (s)</td>
<td>2.5e-5</td>
</tr>
<tr>
<td>(\tau_{0,p}) (s)</td>
<td>2.5e-5</td>
</tr>
<tr>
<td>(N_{\text{SRH,n}}) (cm(^{-3}))</td>
<td>3.401e16</td>
</tr>
<tr>
<td>(N_{\text{SRH,p}}) (cm(^{-3}))</td>
<td>1.33e17</td>
</tr>
</tbody>
</table>

The reduction of the electron \(\tau_n\) lifetime and hole \(\tau_p\) lifetime due to Auger recombination is given by eqns. (3.15a/b).

\[
\tau_n = \frac{1}{C_{\text{Au,n}}N^2}
\]  

(3.15a)

and

\[
\tau_p = \frac{1}{C_{\text{Au,p}}N^2}
\]  

(3.15b)

It should be noted that due to the strong dependence of Auger recombination on the majority carrier concentration, the layer at the silicon/polysilicon interface with very high doping concentrations (the emitter layer in the epi-silicon) may form a so-called dead layer. In this dead layer all injected excess carriers will recombine. This effect gives rise to a higher base current and thus a decrease in the gain.

It is also necessary to include Shockley-Read-Hall (SRH) recombination in the device simulation. The Shockley-Read-Hall theory of recombination assumes that a single recombination center, or trap, exists at an energy within the forbidden bandgap. There are four different processes defining SRH-recombination. The first is the capture of an electron from the conduction band by an initially neutral empty trap. The second process is the inverse of process 1: The emission of an electron from a trap level into the conduction band of the material. The third one is capture of a hole from the valence band by a trap containing an electron. The fourth is again the reverse of the third process: The emission of a hole from a neutral trap into the valence band. The SRH-recombination \(R_{\text{SRH}}\) is given by eqn. (3.16a).
The values of the parameters $\tau_{p,0}, \tau_{n,0}, N_{SRH,p}$ and $N_{SRH,n}$ can be found in table (3.4) [12]. The value of $q\Delta E$ is the difference in energy between the trap level and the intrinsic Fermi level.

### 3.4.2 Surface recombination

The use of polysilicon for emitter contacts allows to make shallower junctions as well as self aligned and hence smaller structures. The emitter efficiency is improved due to a reduction in minority carrier hole current which is injected from the base into the emitter for a npn transistor. A good way to describe this effect is with the help of the effective recombination velocity $S_p$ at the surface of the monocrystalline region. The minority carrier current $J_p$ crossing the polysilicon/silicon interface is related to $S_p$ given by eqn. (3.17) [14],[15].

$$J_p = qpS_p$$  \hspace{1cm} (3.17)

$p = $ Excess minority hole concentration at the monocrystalline silicon surface

All important aspects of the polysilicon and interfacial oxide layer related to the device characteristics are taken into account by the recombination velocity $S_p$ as discussed in section 2.3.3. When rewriting eqn. (2.4) $S_p$ becomes:

$$S_p = S_I + T_{block} \frac{S_I + S_{poly}}{T_{block} + S_I + S_{poly}}$$  \hspace{1cm} (3.18)

As a result of the transport mechanisms through the interfacial oxide, $S_p$ strongly depends on interfacial oxide thickness and the oxide barrier height $X_p$ for holes. The ERV $S_p$ also strongly depends on the polysilicon layer doping. Referring to Bakker et al. the dependence of the ERV $S_p$ on the interfacial oxide thickness with the polysilicon doping level and the barrier height for holes as parameters can be seen in figure (3.1) [14],[15].

In figure (3.1) $S_p$ has been calculated with the help of eqn. (3.18). It is noticed that $S_p$ will asymptotically tend to $S_I$ when the interfacial oxide layer thickness $\Delta$ increases.
due to the fact that the blocking factor $T_{\text{block}}$ decreases with increasing $\Delta$. The device used in this project has a discontinuous interfacial oxide layer with thicknesses between 0 - 8 Å. Therefore the only interesting part of the figure is the part where the interfacial oxide thickness is less than 8 Å.

![Diagram](image_url)

**Fig. 3.1: Effective recombination velocity as a function of interfacial oxide thickness.**

In figure (3.1) it can be seen that for the device used in this project, with a polysilicon doping of about $3 \times 10^{20}$ at.cm$^{-3}$ as taken from SIMS measurements, the value of the ERV $S_p$ is between the $1 \times 10^5$ and $2 \times 10^5$ cm$^2$s$^{-1}$. An exact value of $S_p$ cannot be given, owing to the fact that $S_p$ depends on the physical properties of the polysilicon layer and the polysilicon/silicon interface, which are known to vary significantly with doping and other processing steps. The value used for $S_p$ in device simulations must therefore be used as a fitting parameter.

The value of $S_p$ only influences the base current of a device and not the collector current, because it has influence on the recombination rate of excess minority carrier holes in the emitter of a npn transistor and therefore it determines the minority carrier hole concentration. The minority carrier hole concentration determines the base current of a npn transistor. Decreasing $S_p$ decreases the base current and hence increases the gain because the collector current is not affected. Especially for shallow junctions the improvement in gain will be larger when $S_p$ is decreased, due to the fact that the recombination in the thin monocrystalline silicon emitter region then becomes insignificant.
IV. Process and device simulation

4.1 Introduction

With the help of process and device simulation programs available today it is possible to accurately predict the device characteristics and behavior. For an accurate prediction of device behavior one needs measurement results and a good insight into the physical models used to characterize a device. Without measurement data an accurate prediction of device behavior can not be made, due to the fact that the simulation programs are not exactly calibrated for each different technology step.

Important measurement results used to calibrate the simulation towards the real process of the polysilicon emitter bipolar transistor described in this project are: SIMS profiles, sheet resistances and junction capacitances. Also available were the gummel plots and Hfe plots for realized polysilicon emitter bipolar transistors with different dimensions. These plots were used to calibrate and check the device simulation.

The process and device simulations are described in the following sections. In the first section the process simulation is described while in the section that follows the device simulation is described. In the above sections all simulation results discussed refer to a polysilicon emitter bipolar transistor having the smallest dimensions from which measurement results were available. The transistor used is a vertical double base NPN transistor with an emitter window of 0.9 * 0.9 (μm²). In the last section the results will be checked for a polysilicon emitter bipolar transistor with different dimensions.

4.2 Process simulation

4.2.1 Introduction

The process simulator used in this project is the two dimensional process simulator TSUPREM-4, Technology Modeling Associates, Inc., Palo Alto, CA USA (version 6.3) [19]. TSUPREM-4 is a TCAD (Technology Computer Aided Design) computer program for simulating the processing steps used in the manufacturing of silicon integrated circuits and discrete devices. All different types of processing steps can be modelled. For example: Ion implantation, diffusion, silicon and polysilicon oxidation, epitaxial growth, and low temperature deposition and etching of various materials. TSUPREM-4 simulates the incorporation and redistribution of impurities in a two-dimensional device cross section perpendicular to the surface of the silicon wafer. The output information given by the program includes the boundaries of the various layers, the distribution of impurities in each layer, and stress produced by oxidation, thermal cycling or film deposition.
4.2.2 SIMS and simulated profiles

4.2.2.1 Introduction

Secondary Ion Mass Spectroscopy (SIMS) provides a direct method for measuring the impurity profile of an implanted or epitaxial layer. The technique directly measures the amount of impurity in silicon as the sample surface is slowly sputtered. The sputtering removes about 40 nm of the surface layer. SIMS measures electrically active as well inactive impurity levels. The big advantage of the SIMS technique in comparing it with other techniques like for example Auger Emission Spectroscopy (AES), Electron Spectroscopy for Chemical Analysis (ESCA) and Rutherford Backscattering Spectroscopy (RBS) is that SIMS is capable of detecting all elements and it can identify elements present in very low concentration levels. It is therefore the most widely used surface analysis technique with the ability to measure doping level concentrations in electronic materials. SIMS is an excellent tool for generating concentration profiles of dopants in silicon at levels down to the $1 \times 10^{15}$ at. cm$^{-3}$ range.

To apply correct impurity profiles in the process simulation, these need to be matched with the profiles measured by SIMS. Therefore it is necessary to look at the process technology steps that have an influence on the impurity concentrations. These steps include:

- **Silicon and polysilicon oxidation**

When growing silicon oxide, impurities that are implanted through this SiO$_2$ layer are affected. Due to this SiO$_2$ layer the implanted impurities are blocked and the concentration of the impurities underneath the SiO$_2$ layer will be less compared to a situation without a SiO$_2$ layer. The thickness of the SiO$_2$ plays a significant role, the thicker this layer, the lower the impurity concentration underneath this layer. Boron atoms will go through the SiO$_2$ layer while arsenic atoms are blocked because of their bigger size. Therefore it is very important to match the thicknesses of the different oxide layers with reality. To do this you have to change in the process simulator the time in the diffusion step used to grow the oxide.

Also the models used to grow an oxide play an important part. In the simulations the *vertical* model is used to grow all oxides except for the oxidation of LOCOS (LOCal Oxidation of Silicon). To grow the 530 nm thick LOCOS the *viscous* model has to be used. The *vertical* model is the simplest of the numerical oxidation models in TSUPREM-4. In this model the oxide growth is only in the -y direction and therefore the growth rate depends only on the orientation of the silicon substrate. The *viscous* model simulates in two dimensions viscous flow of the oxide during oxidation which allows accurate values of stress to be computed. Also the variation of crystal orientation along the interface is taken into account when calculating the oxidation rate.
Implantation models, dose, energy, wafer rotation and tilt

In order to benefit from the ability to control the number of impurities implanted into a substrate, it is necessary to know where the implanted atoms are located after implantation. This information is necessary for process engineers for selecting appropriate doses and energies when designing a fabrication process sequence for new or modified integrated circuit devices. To make accurate predictions of implanted impurity profiles a theoretical model is needed, which is based on the energy interaction mechanisms between the impurity ions and the substrate. Within the process simulation program SUPREM-4 there are three different theoretical models to use: Gaussian, Pearson and Montecar type.

The Gaussian model specifies that a simple Gaussian distribution can be used to represent the implanted impurity profile. The simple approach of calculating an impurity profile with the help of a Gaussian distribution function does not account for several effects that occur when implants are made into single-crystalline silicon. Therefore other analytical models to calculate impurity profiles have been developed. Nevertheless, the Gaussian distribution is still commonly used to provide quick (less simulation time) estimates of doping distributions into amorphous and single-crystalline materials.

The Pearson model specifies that a Pearson distribution is used to represent the implanted impurity profile. This model is more complicated than the Gaussian model and therefore it gives a better calculation of the implanted profile. The mathematical expressions of this model fall beyond the scope of this project.

The Montecar model specifies that a numerical Monte Carlo analysis is used to simulate the implantation. This model includes the effect of reflected ions and produces damage information in the form of vacancy and interstitial profiles. In Monte Carlo analysis, ion implantation is simulated by following the history of an energetic ion through collisions with target atoms using the binary collision assumption. The calculation of each trajectory begins with a given energy, position and direction. In the simulations 1000 ion trajectories are calculated and the depth at which each ion stops is determined (it is possible to increase this number, but that was not necessary). The Montecar model is very complicated and time consuming.

When to use which implantation model appears to have influence on the doping profile, but especially on the electrical characteristics of the device. More about this in the next section when discussing the simulation results of the impurity profiles.

It is clear that changing the implantation dose and energy most affects the impurity doping profiles. It is therefore necessary to retain the values of these two parameters as close as possible to reality. Yet in practice it is allowed to change these parameters by 20-30 percent due to physical effects that occur in reality, which are not taken into account in the simulations.

In the simulations rotation of the wafer is not included. The tilt angle was in the simulations and in reality set to 7°. The effect of the 7° tilt angles at which wafers are held to minimize channeling, is to charge the scan rate at the wafer surface so that a
linear gradient in the implant dose across the wafer is created. Taking tilt in the simulations into account, seems to have influence on the doping profiles but especially on the base sheet resistance (see section 4.2.3).

**Diffusion time and temperature**

Once an implantation step is done, it will be followed for most of the time by a diffusion step to outdiffuse the impurities in the silicon layer to create a more deeper/less steeper profile. To match the simulated profile with the SIMS profile it is sometimes necessary to change the time of the diffusion step. A change in temperature is normally avoided because this will also affect the concentrations to differ from the already implanted impurities in the silicon.

4.2.2.2 Simulated profiles

Two profiles were measured with the SIMS technique:

1) The buried-N profile, just after the epi growth and after complete wafer processing.
2) The emitter profile, through the middle of the emitter.

The first profile checked in the simulations was the buried-N profile just after the epi-silicon layer growth. In the simulations the same values as in the real process for the dose and energy were used and a Pearson distribution was chosen to describe the profile. The dashed line in figure (4.1) is the simulated buried-N profile, while the solid line is the SIMS-profile.

**Fig. 4.1: The simulated and SIMS buried-N profile.**
In figure (4.1) there is also a plot of the buried-N SIMS profile and simulated profile (dashed line) after complete wafer processing. Due to outdiffusing of the arsenic into the epi-silicon layer the profile is wider and a little lower than the buried-N profile just after the epi layer growth. The outdiffusion of the arsenic is caused by high temperature diffusion steps, especially the RTA-drive at the end of the process. As can be seen in the figure the simulated profiles are very well matched to the SIMS profiles.

The most important profile in the process of making a polysilicon emitter bipolar transistor is the profile of the emitter. Figure (4.2) shows this profile. The dashed line is the simulated emitter profile while the solid line is the profile measured with the SIMS technique. It can be seen that the doping concentration in the polysilicon layer is almost uniform. This simplifies the transport equations and it makes possible the analysis in section 2.3.

In the polysilicon layer arsenic is implanted with a dose of $5 \times 10^{15}$ at. cm$^{-2}$ at 30 KeV. Due to the RTA-drive at 1075 °C for 12 seconds arsenic diffuses down the grain boundaries in the polysilicon layer and after a while it diffuses into the bulk of the grains and into the monocrystalline silicon. The large arsenic concentration peak at the polysilicon/silicon interface is caused by segregated arsenic at the pseudo-grain boundary between the polysilicon and the silicon. The peak value in the simulated profile $6 \times 10^{20}$ at. cm$^{-3}$ is a little higher than the peak value from the SIMS measurements $4 \times 10^{20}$ at. cm$^{-3}$.

The arsenic dose measured by the SIMS measurement technique is $3.1 \times 10^{15}$ at. cm$^{-2}$, with this dose the arsenic profile is plotted in figure (4.2). The arsenic dose implanted in the real process and in the simulations is $5 \times 10^{15}$ at. cm$^{-2}$. The simulated arsenic profile in figure (4.2) (dashed line) gives therefore a little higher peak concentration than the arsenic profile measured by SIMS. When doing a simulation with an arsenic dose of $3.1 \times 10^{15}$ at. cm$^{-2}$ the peak concentration of the arsenic at the polysilicon/silicon interface will decrease to the same value as the peak concentration of the SIMS profile.

The above described difference in peak concentration can as described above simply be matched by lowering the arsenic implant dose. This has not been done because lowering the arsenic implant dose gives an increase in the base Gummel number $G_B$ resulting in a decrease in the collector current. Also the emitter Gummel number $G_E$ will decrease when the arsenic implantation dose is lowered which results in a slightly higher base current (see table 4.1). Thus lowering the arsenic implant dose will decrease the transistor gain $H_{fe}$.

**TABLE 4.1: Calculated emitter and base Gummel numbers**

<table>
<thead>
<tr>
<th></th>
<th>Standard/real</th>
<th>Arsenic dose 3.1e15 at.cm$^2$</th>
<th>RTA temp. 1100 °C</th>
<th>RTA drive 20 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_E$ (at.cm$^{-2}$)</td>
<td>2.37e14</td>
<td>1.40e14</td>
<td>4.06e14</td>
<td>3.94e14</td>
</tr>
<tr>
<td>$G_B$ (at.cm$^{-2}$)</td>
<td>5.86e12</td>
<td>6.88e12</td>
<td>4.18e12</td>
<td>4.28e12</td>
</tr>
</tbody>
</table>

31
The RTA emitter drive affects the diffusion of the arsenic into the monocristalline silicon emitter part. Therefore the RTA emitter drive determines the emitter/base junction depth, the emitter $G_E$ and base $G_B$ Gummel number (see eqns. 4.1a/b) and the breakup of the interfacial oxide layer, which allows the arsenic to diffuse into the silicon.

$$G_E \approx N_D x_E$$

(4.1a)

$$G_B \approx N_A x_B$$

(4.1b)

$x_E =$ Emitter/base junction depth in the epi-silicon

$x_B =$ Base/collector junction depth in the epi-silicon

In eqns. (4.1a/b) uniform doping concentrations in the emitter and base are assumed, but in the real process there are of course emitter and base non uniformities (see also figure 4.2) and the emitter and base gummel number will therefore be given by eqn. (2.6d). The base Gummel number determines the collector current and is therefore a crucial factor for the device simulation. The junction depths can be found in table (4.2). The calculated base and emitter Gummel numbers for an increased RTA temperature can be found in table (4.1).

![Graph](image)

**Fig. 4.2:** SIMS and simulated emitter-base profile.

**TABLE 4.2:** Simulated and SIMS junction depths

<table>
<thead>
<tr>
<th></th>
<th>SIMS measurements junction depths</th>
<th>Simulated junction depths</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_E , (\mu m)$</td>
<td>0.058</td>
<td>0.044</td>
</tr>
<tr>
<td>$x_B , (\mu m)$</td>
<td>0.33</td>
<td>0.35</td>
</tr>
</tbody>
</table>
The base/collector junction depth \( x_B \) taken from the SIMS measurement technique cannot be predicted exactly from figure (4.2). In this figure the part where the base/collector junction depth lies is difficult to accurately determine. However, the simulated junction depth (exactly calculated) of 0.35 \( \mu \)m is not that much different from the value obtained from the SIMS profile from figure (4.2). The value for the simulated emitter/base junction depth \( x_E \) differs no more than 20 percent.

The emitter/base junction depth \( x_E \) critically depends on the RTA-drive. Increasing the RTA temperature or duration results in an increase in the arsenic out-diffusion from the polysilicon emitter and thus in a reduced base Gummel number resulting in an increase in the collector current. Increasing the RTA duration from 12 seconds to 20 seconds has the same effect on the emitter/base junction depth \( x_B \) as an increase of RTA temperature by 25 °C. The emitter/base junction depth \( x_E \) increases 12 nm towards 0.056 \( \mu \)m when changing the RTA duration or temperature by these values. This has not been done in the simulations, because a reduced base Gummel number increases the base sheet resistance. The base sheet resistance is a very important parameter for this process and has to be matched with reality (see section 4.2.3).

The boron profile in figure (4.2) determines the base profile of the polysilicon emitter bipolar transistor. In the SIMS profile only the isotope \( B^{11} \) (dose 2.0e13 at.cm\(^{-2}\)) is measured while in the simulated profile the total boron dose was considered. The \( B^{10} \) isotope concentration in the process is 20 percent of the \( B^{11} \) concentration (BF\(_2\)-implant includes both isotopes) which accounts for an increase of the boron SIMS profile in figure (4.2). Taking this increase of the boron SIMS profile (compared to the simulated profile) into account the profile will still be a little higher than the SIMS profile. This is due to the fact that the dose of the BF\(_2\) (base)-implantation, to form the base of the npn transistor, in the simulations was increased from 2.4e13 towards 2.8e13 at.cm\(^{-2}\) to obtain a better value for the base sheet resistance (section 4.2.3).

A very crucial processing step is the first boron implantation in the process. The first boron implantation in the process is one with a dose 1.7e12 at.cm\(^{-2}\) and energy 25 KeV. This implantation is done to tune the \( V_t \) (threshold voltage) for the CMOS part of the process. This implantation, initially done to only influence the threshold voltage of the MOS transistors used in the process, also affects the bipolar part of the process. The influence of this implantation on the device electrical characteristics is dependant on the implantation model chosen. This will be discussed further in section 4.3.2.

A good example to show the effect of the different implantation models on the implanted profiles can be found in figure (4.3). In this figure the profile of the arsenic implanted in the polysilicon emitter region is plotted prior to the RTA drive. The only difference in the three profiles plotted is the use of a different implantation model. The implantation models used are: Gaussian, Pearson and Montecarlo respectively.

A comparison of these simulated profiles with the profile of the real process can not be made, because there was no available SIMS profile of this implant. In fact in figure (4.3) it can be seen that the simulated arsenic profile with the use of the Montecarlo model is not a physically expected profile. The Monte Carlo calculated profile is too steep and shows no arsenic concentration at the polysilicon surface. Therefore, in the simulations, the profile used has been calculated by a Pearson distribution function.
Most of the time the *Pearson* model is chosen for the different implantation steps throughout the whole process, except for implantations with energies above 200 KeV. For these implantations no Pearson tables are available in the process simulator SUPREM-4 and therefore a Monte Carlo calculation has been chosen.

![Graph](image)

**Fig. 4.3**: The arsenic profile in the polysilicon emitter before annealing for three different implantation models.

### 4.2.3 Sheet resistances

One of the most important parameters to be measured during fabrication of wafers is the pinched/intrinsic base sheet resistance $R_{be}$. This is the sheet resistance of the quasi-neutral base under the emitter diffusion region and is directly related to the Gummel integral by eqn. (4.2).

$$R_{be} \, s q = \frac{1}{W_b} q \mu_p \int_0^{W_b} N_{eff}(x) dx$$

(4.2)

$W_b$ = The neutral base width

Another important parameter is the extrinsic base sheet resistance $R_b$, given by the same eqn. as the eqn. for the intrinsic base sheet resistance (eqn. 4.2). The calculation of both the extrinsic and intrinsic base sheet resistances is a two-dimensional problem.

Both sheet resistances are very much dependent on the different processing steps. Therefore it is necessary to match the simulated sheet resistances with the measured sheet resistances to be sure that the simulations are correct. The simulated as well as the measured intrinsic and extrinsic base sheet resistances are found in table (4.3).
TABLE 4.3: Calculated and simulated square base sheet resistances.

<table>
<thead>
<tr>
<th></th>
<th>Measured values</th>
<th>Simulated values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_b$ (kΩ/sq)</td>
<td>2.75</td>
<td>2.86</td>
</tr>
<tr>
<td>$R_{be}$ (kΩ/sq)</td>
<td>9.0</td>
<td>8.67</td>
</tr>
</tbody>
</table>

The most critical step in the process for the extrinsic as well as for the intrinsic base sheet resistance is the base-implantation. In the real process this implantation is a BF$_2$ implant of $2.4\times10^{13}$ at.cm$^{-2}$ at 40 KeV, while in the simulations the implantation dose was increased towards $2.8\times10^{13}$ at.cm$^{-2}$ to obtain better values for the base sheet resistances. The values for the extrinsic and intrinsic base sheet resistance given in table (4.3) are those for the simulation with a implantation dose of $2.8\times10^{13}$ at.cm$^{-2}$. In fact the value for the extrinsic and intrinsic base sheet resistance for an implantation dose of $2.4\times10^{13}$ at.cm$^{-2}$ is 3.04 kΩ/sq and 9.47 kΩ/sq respectively.

With a dose of $2.4\times10^{13}$ at.cm$^{-2}$ the value for the intrinsic and extrinsic base sheet resistance is somewhat higher than the values with an implantation dose of $2.8\times10^{13}$ at.cm$^{-2}$. The main reason for using a higher implantation dose in the simulations is that the extrinsic as well as the intrinsic base sheet resistance showed better values compared to reality. Also the Hfe plot gives much better values in the low injection region when using the higher dose of $2.8\times10^{13}$ at.cm$^{-2}$. More about this in section 4.3.2 where the device simulations are discussed.

When changing the RTA emitter drive time from 12 to 20 seconds the pinched base sheet resistance increases from 8.67 kΩ/sq towards 12.3 kΩ/sq. This difference can be explained with eqn. (4.2). When the RTA diffusion time is increased towards 20 seconds the base Gummel number decreases 27 percent (see table 4.1). This gives an increase in the pinched base sheet resistance by 37 percent towards 11.87 kΩ/sq due to eqn (4.2). This value is close to 12.3 kΩ/sq, and the difference can be explained by a slightly reduced hole mobility.

The situation is different when the same RTA drive time is increased from 12 seconds towards 20 seconds for a simulation in which the dose of the first boron implantation ($V_t$-implantation) is increased from $1.7\times10^{12}$ at.cm$^{-2}$ towards $2.1\times10^{12}$ at.cm$^{-2}$. The pinched base sheet resistance for an implantation dose of $2.1\times10^{12}$ at.cm$^{-2}$ and 12 seconds RTA emitter drive time is 8.17 kΩ/sq. When changing the RTA emitter drive time towards 20 seconds the pinched base sheet resistance is increased towards 37.3 kΩ/sq. Due to the RTA drive time increase the base Gummel number decreases from $5.88\times10^{12}$ at.cm$^{-2}$ towards $2.17\times10^{12}$ at.cm$^{-2}$. This results with the help of eqn. (4.2) in a pinched base sheet resistance of 22.14 kΩ/sq. This value is more than 40 percent lower than the simulated value. This is unlikely to be due to a reduced hole mobility. This needs to be investigated further.
4.2.4 The role of the interface oxide break-up model in SUPREM-4

In SUPREM-4 version 6.3 modeling of polycrystalline material has been incorporated for the first time. The behavior of dopants in polycrystalline material is strongly influenced by the boundaries between crystalline grains. The rate of segregation at the grain boundaries depends on the rate of grain growth, while the number of diffusion paths along the boundaries depends on the grain size.

A thin interfacial oxide layer is typically present between a deposited polysilicon layer and an underlying single-crystalline silicon layer. This interfacial oxide layer presents a barrier to epitaxial realignment of the poly layer. Due to the high temperature RTA emitter drive the interfacial oxide layer breaks up and a discontinuous oxide layer is formed for the device described in this project.

In SUPREM-4 a model is described in which the parameters for the model are specified in terms of a characteristic break-up time for the thinnest (5Å) interfacial oxide layer (see eqn. 4.3a).

\[
\frac{dR_{\text{void}}}{dt} = \frac{1}{t_{bu}} \left( \frac{5\text{Angstrom}}{t_{ox}} \right)^3 \frac{1}{\sqrt{\pi N_{EA}}} \tag{4.3a}
\]

\[R_{\text{void}} = \text{The radius of the voids in the interfacial oxide layer}\]

\[t_{bu} = TBU.0 \exp \left( \frac{-TBU.E}{kT} \right) \tag{4.3b}\]

\[t_{ox} = \text{interfacial oxide thickness}\]

\[N_{EA} = \text{areal density of the voids}\]

The parameter \(t_{ox}\) can be changed in the process simulator SUPREM-4, also the parameters \(TBU.0\) (7e-19), (a constant), and \(TBU.E\) (-5.0), (the activation energy of the oxide break-up process), can be changed.

The above model looks interesting but only values for \(t_{ox}\) varying between 0 - 8 Å had some influence on the electrical characteristics of the device. The base current was slightly reduced for a greater value of \(t_{ox}\). This is less significant because the value of the effective surface recombination velocity of holes in the emitter, which can be used to influence the base current in the device simulator MEDICI, appeared to have a much larger effect on the base current. The above described interfacial oxide model in SUPREM-4 needs to be updated in the future, so that it will be possible to even model a discontinuous oxide layer perfectly. Presently this model shows a better base current, but the model is not so good that it is possible to see what the influence will be of different settings of the parameters on the device electrical characteristics.
4.3 Device simulation

4.3.1 Introduction

The device simulator used in this project is the two-dimensional device simulator MEDICI, Technology Modeling Associates, Inc., Sunnyvale, CA USA (version 2.1.2) [16]. MEDICI is a powerful device simulation program that simulates the behavior of MOSFETS, bipolar transistors and other semiconductor devices. It makes use of the two-dimensional distribution of potential and carrier concentrations in a device. The output of the program is the electrical characteristics of a device for arbitrary bias condition.

A number of physical models have been incorporated into the program for accurate simulation of different devices. There are models for recombination, impact ionization, photogeneration, band-gap narrowing, band to band tunneling, mobility, and lifetime of the carriers. MEDICI also incorporates both Boltzmann and Fermi-Dirac statistics, including incomplete ionization of impurities.

4.3.2 Electrical characteristics

MEDICI calculates all kinds of different electrical characteristics for all kinds of devices. For a polysilicon emitter bipolar transistor the most important characteristics looked at in this project are:

- The base sheet resistances
- The Gummel plot
- The Hfe plot

To make physical calculations it is very important to include the right physical models within MEDICI. In chapter 3 these models are discussed. The following physical effects are included in the device simulations: Bandgap narrowing, Auger and Shockley-Read-Hall recombination, finite surface recombination, Philips unified mobility model to calculate the carrier mobility and tunneling.

In figure (4.4) the Gummel plot of the device is given, while in figure (4.5) the Hfe plot of the device is given. In both figures the solid line is the measured plot while the dashed line (Hfe) or the line with the markers (Gummel) is the simulated plot.

To check the Gummel plot more accurate absolute values are taken from the measurements as well as from the simulations for three different operating regions (low injection region \( V_{be} \approx 0.4V \), linear region \( V_{be} \approx 0.7V \) and the high injection region \( V_{be} \approx 1.0V \)) and those values are given in table (4.4).

The simulated collector current over the complete range of base emitter voltages is well matched to the measured values. This is not surprising since the collector current depends on the base Gummel number and therefore on the pinched base sheet resistance. The value of the pinched base sheet resistance in the simulations as discussed in section 4.2.3 is almost equal to the measured value.
By far the most important parameter for this simulation, the parameter which has the most influence on both the collector current and on the base current, is the temperature $T$. The exact temperature at which the measurements took place is not exactly known due to different room temperatures at different times/periods. Therefore this value is taken as 295 K. To obtain good values for the simulations the temperature was set to 293 K which is the lowest reasonable value for room temperature. The collector and base current are both increasing with increasing temperature and decreasing with reduced temperature.

![Gummel plot](image)

**Fig. 4.4: Measured and simulated Gummel plot**

**TABLE 4.4: Absolute values for the collector and base current.**

<table>
<thead>
<tr>
<th>Vbe (V)</th>
<th>$V_{bc} = -1V$</th>
<th>Simulated values; $T = 293$ K $V_{bc} = -1V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>2.29e-13</td>
<td>4.51e-14</td>
</tr>
<tr>
<td>0.70</td>
<td>2.66e-9</td>
<td>3.12e-9</td>
</tr>
<tr>
<td>1.0</td>
<td>2.10e-5</td>
<td>5.99e-5</td>
</tr>
</tbody>
</table>

For example when the temperature in the simulations is increased to the MEDICI default setting of 300 K the collector current for a base emitter voltage of 0.7 V increases by about 70 percent, while the base current increases by 55 percent. This gives rise to an $Hfe$ ($I_c/I_b$) increase of only 8 percent.

The values for the simulated base current see table (4.4) do not well match the measured values. Only for the linear region (the normal operating region of a transistor) the base current is well matched. For the low injection region the simulated base current is about a factor 5 lower than the measured value, while for the high
injection region the simulated base current is almost a factor 3 higher than the measured value.

Naturally the value of the Effective Recombination Velocity \( S_p \) (see section 3.4.2) has much influence on the base current. In the simulations a finite surface recombination velocity is taken into account. In this process only the surface recombination velocity for holes \( v_{surfP} \) in the emitter has influence on the base current. When decreasing this factor the base current decreases and when increasing the surface recombination velocity for holes the base current also increases. Therefore this is a parameter that can be used to affect the gain \( (Hfe) \) of a transistor in the simulations.

From figure (3.1) in section 3.4.2 it can be seen that for the process described in this project the recombination velocity has a value between 1e5 and 2e5 cms\(^{-1}\). The value used in the device simulation is 1.4e5 cms\(^{-1}\), a value in good agreement with other authors [14],[15]. The level of the gain has been matched by changing the value for the effective surface recombination velocity. This has been done for the linear region, because this is the normal operating region of the transistor.

![Fig. 4.5: The simulated and measured Hfe plot](image)

In figure (4.5) the measured and the simulated Hfe plot are shown. The Hfe \( (\beta = \frac{I_c}{I_b}) \) plot can be divided in three separate regions: The low injection region, the normal operating region of the transistor and the high injection region.

In the low injection region, at low forward bias of the base-emitter junction, the current gain \( \beta \) will be reduced. This is due to the fact that at low forward bias, there is an important additional component of base current due to recombination of excess carriers inside the emitter-base space charge layer added to the normal base current. Therefore the base current increases at low forward bias of the base-emitter junction and the gain automatically decreases because the collector current is not affected.
Especially in the simulated Hfe plot there is a slight fall off of the gain in the normal region of operation. The reason for this fall off is due to the so called reverse Early effect. At high base-emitter voltages the emitter-base depletion layer decreases which increases the base Gummel number due to an increased effective base width. An increased base Gummel number accounts for a less dramatic increase in the collector current at high base-emitter voltages. Therefore the gain is slightly reduced for higher base-emitter voltages.

This slight fall off of the gain in this ideal region of operation can be reduced by performing a simulation without taking the effect of band-gap narrowing into account. When doing a simulation without bandgap narrowing and by stating infinite surface recombination in the emitter there is almost no fall off of the gain at higher base-emitter voltages.

The effect of bandgap narrowing increases the effective intrinsic equilibrium carrier density eqn. (2.6c), it decreases the effective doping concentration in the emitter eqn. (2.6b) and therefore the base current is increased eqn. (2.6a). In the simulations the base current is increased by a factor 40 when taking bandgap narrowing into account. The collector current is only increased less than by a factor 2 with bandgap narrowing. Therefore the gain is decreased more than a factor of 20 with bandgap narrowing taken into account.

The decreased gain due to the effect of bandgap narrowing can be compensated by taking finite surface recombination in the emitter into account. Finite surface recombination of the minority carrier holes in the emitter decreases the base current and has no influence on the collector current. Therefore the gain is increased when using finite surface recombination of holes in the emitter instead of infinite surface recombination of holes at the emitter contact.

The reverse Early effect has more influence when the base doping is higher. Bandgap narrowing is a high doping effect and has influence for doping concentrations larger than 1e17 at.cm⁻³. For the high doping concentrations in the thin epi-silicon emitter part of the structure these two effects have much influence. The slight fall off of the gain in the ideal operating region due to the Early effect can therefore best seen in the Hfe plot where the effect of band-gap narrowing is neglected. In the other plot, the plot with bandgap narrowing, the effect of bandgap narrowing and the Early effect both account for the reduced gain for higher base-emitter voltages. A simulation without taking the effect of bandgap narrowing into account is indeed not physically allowed. The effect of bandgap narrowing is nowadays very well understood and has to be taken into account in the simulations.

The strong fall off of the gain in the high injection region is due to the Kirk effect. The actual mechanism of the Kirk effect consists in a vertical widening of the effective neutral base region due to a decrease in the collector-base depletion layer. The decrease in the collector-base depletion region is much stronger than the decrease in the base-emitter depletion region due to the inverse Early effect. Therefore the base Gummel number is much more increased for the Kirk effect than for the inverse early
effect. The collector current decreases and hence the gain decreases by a factor of 2 or 3 at high level injection.

The curve of the simulated $H_{fe}$ plot figure (4.5) shows fall off of the gain in the low injection region, fall off of the gain in the normal operating region of the transistor and a strong fall off of the gain in the high injection region. Comparing the simulated $H_{fe}$ plot with measured values differences can be seen.

The fall off of the gain in the high injection region due to the Kirk effect for the simulated curve is similar to the measured curve except that the fall off starts for the same collector current but only with a lower $H_{fe}$. This is due to the fact that the fall off of the gain in the normal operating region of the transistor is much more for the simulated plot than for the measured plot. This effect, as stated before, can be overcome by doing simulations without bandgap narrowing and taking an infinite surface recombination into account.

Furthermore the fall off of the gain in the low injection region for the simulated curve starts at a collector current much lower than the onset of fall off for the measured curve. This problem can possibly be solved by taking an higher effective recombination velocity in the simulations into account at the surface of the emitter base depletion region. Simulations with changing the effective recombination velocity, only at this position in the structure, have not been done. This is indeed recommended for future work.

In the simulations the fall off of the gain in the high injection region was always very well present for all kind of changes in the process technology steps. The problem in fact is the fall off of the gain in the low injection region. The falloff of the gain in this region, due to increased base current, is much dependable on the different steps in the process simulator SUPREM-4, like implant doses, diffusion times and models chosen.

From the simulation results it appears that the processing step with the biggest influence on the fall off of the gain in the low injection region is the first boron implantation in the process (dose $1.7e12 \, \text{at.cm}^{-2}$, energy 25 KeV). This boron implantation is done to tune the threshold voltage for the CMOS part of the process. It is very interesting that the simulations show this implantation to have much influence on the bipolar part of the process.

The implantation model chosen to simulate this boron implantation accounted for the difference in fall off of the gain in the low injection region, but it had no influence on the normal operating region of the transistor or on the high injection region. The only way to get a correct $H_{fe}$ plot was by choosing the Monte Carlo model for this implant.

A simulation was done with two equal SUPREM-4 input decks, the only difference was the implantation model for the first boron implant. The results can be seen in figure (4.6). The $H_{fe}$ plot with the Monte Carlo model for this boron implantation is already shown in figure (4.5), this is re-plotted to show the differences with the other models. When using the Pearson model in MEDICI, the fall off of the gain in the low injection region starts for a collector current a factor 100 lower compared to the Monte Carlo model. This is not realistic anymore compared to the measured profile.
Also important to get a correct Hfe plot, when using the Montecar model for the threshold voltage adjust boron implant, as well as the BF₂ implant to create the base of the transistor. In the real process this implantation is done at 2.4e13 at.cm⁻² at 40 KeV. The Hfe plot when using these values for the base implant with the Montecar model for the first boron implant can be seen in figure (4.6). This plot also shows a fall off in the low injection region for much lower collector current compared to the dose of 2.8e13 at.cm⁻². Furthermore the base sheet resistances showed better values for this increased dose, therefore this dose was choosen for the simulations.

The implantation dose of the first boron implantation (V₁ - adjust) seemed to have influence on the Hfe characteristics. Increasing this dose showed a fall off of the gain in the low injection region for even higher collector currents and showed a slighter fall off of the gain in the ideal region. This dose was not increased otherwise the pinched base sheet resistance became much too low.

![Hfe plot](image)

Fig. 4.6: Simulated Hfe plots for different process parameters

A closer look at the first boron implantation is clearly required. In the real process this implantation is done for the CMOS part of the process, but the simulations showed much influence on the bipolar part of the process. Therefore experiments have to be done on real silicon to see what influence in reality this process step has on the behavior of the polysilicon emitter bipolar transistor. With this information the bipolar part of the BIMOS-3 process can be further optimized.

### 4.3.3 Junction capacitances

The base doping profile can be confirmed by simulating and measuring the base-emitter $C_{be}$ and for a part the base-collector $C_{bc}$ junction depletion layer capacitances, which are determined by the net active doping concentration in the base. The simulated as well as the measured base-emitter and base-collector junction depletion layer capacitances can be found in table (4.5) for a 0.9 * 0.9 μm² emitter window.
TABLE 4.5: Base-emitter and base-collector junction depletion layer capacitances.

<table>
<thead>
<tr>
<th></th>
<th>Measured values</th>
<th>Simulated values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V_{be} = 0V$ ; $V_{bc} = 0V$</td>
<td>$V_{be} = 0V$ ; $V_{bc} = 0V$</td>
</tr>
<tr>
<td>$C_{be}$ (fF($\mu m$)$^2$)</td>
<td>2.24</td>
<td>2.95</td>
</tr>
<tr>
<td>$C_{bc}$ (fF($\mu m$)$^2$)</td>
<td>2.77</td>
<td>1.90</td>
</tr>
</tbody>
</table>

The measured values of the capacitances are not correct due to inaccuracy of the measurements of small single emitter transistors. It is clear that the base-collector junction capacitance at zero bias is lower than the base-emitter junction capacitance which is shown by the simulation results.

4.4 Geometry dependence of the simulations

In this chapter all results were discussed based on one double base transistor with an emitter window size of 0.9 * 0.9 $\mu m^2$. The input deck of both the process simulator as well as the device simulator were calibrated for this transistor (see section 4.2 and 4.3).

TABLE 4.6: Mask sizes in the x-direction for two transistors with different dimensions.

<table>
<thead>
<tr>
<th>Masks:</th>
<th>0.6 * 2.4 emitter window Values in $\mu m$</th>
<th>0.9 * 0.9 emitter window Values in $\mu m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buried-N (BN)</td>
<td>-4.2 ↔ 5.1</td>
<td>-3.75 ↔ 6.15</td>
</tr>
<tr>
<td>Oxide definition (OD)</td>
<td>-6 ↔ -4.95 ; -4.35 ↔ -3.15 ; 3.15 ↔ 4.35 ; 5.85 ↔ 8.4</td>
<td>-6 ↔ -3.15 ; 3.15 ↔ 4.95 ; 6.45 ↔ 8.4</td>
</tr>
<tr>
<td>N-well (NW)</td>
<td>-5.25 ↔ -4.5 ; 6.15</td>
<td>-4.65 ↔ -4.05 ; 4.65 ↔ 7.05</td>
</tr>
<tr>
<td>Shallow N$^+$ (SN)</td>
<td>-5.55 ↔ -3.75 ; 3.75 ↔ 6.45</td>
<td>4.35 ↔ 7.05</td>
</tr>
<tr>
<td>Shallow P$^+$ (SP)</td>
<td>-3.75 ↔ -1.05 ; 1.05 ↔ 3.75</td>
<td>-3.75 ↔ -0.9 ; 0.9 ↔ 3.75</td>
</tr>
<tr>
<td>Contact to silicon (CS)</td>
<td>-0.3 ↔ 0.3</td>
<td>-0.45 ↔ 0.45</td>
</tr>
<tr>
<td>Polysilicon emitter (PSE)</td>
<td>-0.9 ↔ 0.9</td>
<td>-0.75 ↔ 0.75</td>
</tr>
</tbody>
</table>

In this section the results will be checked on a transistor with different dimensions: a double base polysilicon emitter transistor with an emitter window size of 0.6 * 2.4 $\mu m^2$.

The exact mask sizes for both transistors are given in table (4.6). The value through the middle of the emitter is for both transistors set to 0. MEDICI is a 2-D and not a 3-D device simulator therefore it takes a thickness of 1 $\mu m$ into account as the third dimension. Therefore the mask sizes in table (4.6) are only given for the x-direction. To get good values for the currents for the transistor with an emitter window size of 0.6 * 2.4 $\mu m^2$ the currents have to be multiplied with a factor 2.4. This was also done in the last section but then of course with a factor 0.9.

All results given in the last two sections for the transistor with an emitter window of 0.9 * 0.9 $\mu m^2$ are checked for the transistor with an emitter window of 0.6 * 2.4 $\mu m^2$. 

43
All simulation results appeared to be the same as the measured results with some differences mentioned in the next paragraphs.

The Gummel plot and the hfe plot for this transistor with larger dimensions showed that the electrical characteristics were also well defined with the same input decks used in the simulators SUPREM-4 and MEDICI. Although MEDICI and SUPREM-4, two-dimensional device and process simulation programs are, dependent on the thickness can be neglected for this structure. This is also known for the real process where transistors with different dimensions were measured. Only for a transistor with a very large emitter window size, for about 10 μm * 10 μm, there was a difference in gain.

For such a big transistor, not readily used for manufacturing, the edge capacitance can be neglected. This gives a decrease in current gain (Hfe) going from 135 for the normal size emitter window transistors, as the ones discussed in this project, towards 100.

To get the correct level of gain (about 135) for the simulated 0.6 * 2.4 μm transistor, the effective surface recombination velocity for holes in the emitter had to be changed. The value was changed from 140.000 cms⁻¹ towards 220.000 cms⁻¹. This states for a dependence of the ERV on some mechanisms which contribute to the device emitter and base current which are dependable on the emitter dimensions, like edge capacitances which become important in structures with small emitter dimensions. This needs to be investigated more in the future.
V. Statistical experiment

5.1 Introduction

Especially for process engineers it is interesting and extremely beneficial to see what the influence of some process parameters is on the electrical behavior of a device.

In the process of making a polysilicon emitter bipolar transistor there are a few important process parameters which are known to affect the electrical behavior of a device. In the experiment, described in this chapter, the influence of 4 critical process parameters on the electrical behavior of the device have been investigated. These parameters are:

- The base implantation energy
- The base implantation dose
- The etch-back of silicon underneath the polysilicon
- The temperature of the RTA emitter drive in

It should be clear to the reader that the base implantation dose and energy are important process parameters with much influence on the currents within the device. There will always be a little etch back of epi-silicon when polysilicon is deposited. Measurements showed that this etch-back influences the device behavior and therefore this parameter is taken into account in the experiment to quantify it’s influence on the electrical behavior. As known from chapter 4 the RTA emitter drive temperature is critical for the device behavior.

Two important parameters the polysilicon implant dose and energy are not taken into account because it is known that they have much influence on the effective recombination velocity as can be seen in figure (3.1). The exact dependence of the ERV on the polysilicon implantation is still not known and therefore not taken into account in this experiment.

5.2 Design of experiments

First of all a design of experiment had to be chosen. The choice for a design depends on the information needed and especially for this experiment it was important to reduce the amount of experiments (simulations) as much as possible because of the long simulation time (\(\approx 10\) hours CPU time per simulation). To gather enough information from the simulations for a good estimation of the quadratic effects and interactions between two parameters, a Box-Behnken design was chosen [17].

The experiment is a 4-factors 3-level Box-Behnken design. This means that for the 4 factors (process parameters) from section 5.1 three values are chosen: A high value (+1), a default value (0) and a low value (-1). The exact values chosen to run this experiment can be found in table (5.1).
TABLE 5.1: Settings for the process parameters.

<table>
<thead>
<tr>
<th>Factor standardized setting</th>
<th>low setting (-1)</th>
<th>default setting (0)</th>
<th>high setting (+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>energy (KeV)</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>dose (at.cm⁻²)</td>
<td>2.24e10¹³</td>
<td>2.80e10¹³</td>
<td>3.36e10¹³</td>
</tr>
<tr>
<td>etch back (nm)</td>
<td>0</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>temperature (°C)</td>
<td>1050</td>
<td>1075</td>
<td>1100</td>
</tr>
</tbody>
</table>

The standard 4-factor 3-level Box-Behnken design consists of 27 simulation runs. From these 27 runs three runs are performed in the center point. For a computer experiment this center point is the run with the default value for all 4 factors. This run has to take place only one time in a computer experiment, in contrast to empirical experiments. Therefore the computer experiment consists of 25 runs. Note that when estimating every point in the simulations \(3^4 = 81\) runs are necessary. A Box-Behnken experiment indeed gives a good approximation with only 25 runs (see table 5.2).

TABLE 5.2: Settings of the process parameters for the 25 runs used in the experiment.

<table>
<thead>
<tr>
<th>Run</th>
<th>Energy (KeV)</th>
<th>Dose (at.cm⁻²)</th>
<th>Etch back (nm)</th>
<th>Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>2.80e10¹³</td>
<td>0</td>
<td>1050</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>2.24e10¹³</td>
<td>10</td>
<td>1050</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>2.80e10¹³</td>
<td>10</td>
<td>1050</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>2.80e10¹³</td>
<td>10</td>
<td>1050</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>3.36e10¹³</td>
<td>10</td>
<td>1050</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>2.80e10¹³</td>
<td>20</td>
<td>1050</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>2.24e10¹³</td>
<td>0</td>
<td>1075</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>2.80e10¹³</td>
<td>0</td>
<td>1075</td>
</tr>
<tr>
<td>9</td>
<td>35</td>
<td>2.24e10¹³</td>
<td>10</td>
<td>1075</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>2.24e10¹³</td>
<td>10</td>
<td>1075</td>
</tr>
<tr>
<td>11</td>
<td>35</td>
<td>3.36e10¹³</td>
<td>10</td>
<td>1075</td>
</tr>
<tr>
<td>12</td>
<td>45</td>
<td>3.36e10¹³</td>
<td>10</td>
<td>1075</td>
</tr>
<tr>
<td>13</td>
<td>40</td>
<td>2.24e10¹³</td>
<td>20</td>
<td>1075</td>
</tr>
<tr>
<td>14</td>
<td>35</td>
<td>2.80e10¹³</td>
<td>20</td>
<td>1075</td>
</tr>
<tr>
<td>15</td>
<td>45</td>
<td>2.80e10¹³</td>
<td>20</td>
<td>1075</td>
</tr>
<tr>
<td>16</td>
<td>40</td>
<td>3.36e10¹³</td>
<td>20</td>
<td>1075</td>
</tr>
<tr>
<td>17</td>
<td>40</td>
<td>2.80e10¹³</td>
<td>0</td>
<td>1100</td>
</tr>
<tr>
<td>18</td>
<td>45</td>
<td>2.80e10¹³</td>
<td>0</td>
<td>1100</td>
</tr>
<tr>
<td>19</td>
<td>40</td>
<td>3.36e10¹³</td>
<td>0</td>
<td>1100</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>2.24e10¹³</td>
<td>10</td>
<td>1100</td>
</tr>
<tr>
<td>21</td>
<td>35</td>
<td>2.80e10¹³</td>
<td>10</td>
<td>1100</td>
</tr>
<tr>
<td>22</td>
<td>45</td>
<td>2.80e10¹³</td>
<td>10</td>
<td>1100</td>
</tr>
<tr>
<td>23</td>
<td>40</td>
<td>3.36e10¹³</td>
<td>10</td>
<td>1100</td>
</tr>
<tr>
<td>24</td>
<td>40</td>
<td>2.80e10¹³</td>
<td>20</td>
<td>1100</td>
</tr>
<tr>
<td>25</td>
<td>40</td>
<td>2.80e10¹³</td>
<td>10</td>
<td>1075</td>
</tr>
</tbody>
</table>
Run 25 from table (5.2) is the run with the default settings of the four process parameters. This run was optimized with the help of measurement results described in chapter 4.

To run this Box-Behnken experiment the TCAD program NORMAN/DEBORA (IMEC, Leuven, Belgium) was used [18]. Only the NORMAN part of the program was used for handling the experiment runs, the analysis of the data was done with an additional statistical program called STATISTICA instead of DEBORA.

Optimizing a process technology involves finding an optimum set of parameter settings such that a number of relevant results meet predefined targets. The methodology implemented in NORMAN for solving this problem is based on systematic application of the concepts of statistical DOE (Design Of Experiments) for planning a number of experiments for different settings of the input factors.

NORMAN does have a fair number of built-in DOE such as the composite designs (CCF, CCC, CCI), Plackett-Burman, Taguchi and Latin-hypercube designs, random, grid and diagonal designs, 3-level full factorial, Box-Behnken, etc. As stated earlier for the experiment described in this section a 4-factor 3-level Box-Behnken design was chosen. When using the 4-factor 3-level Box-Behnken design built in NORMAN a design is made consisting of 57 runs, this is clearly wrong based on the theory of Box-Behnken [17]. Therefore the built in Box-Behnken design was not used instead the 25 runs in table (5.2) were used.

The input file of NORMAN [RNR-G-97x-034 ; Philips, internal report] consists of three parts:

- Definition of the settings of the process parameters
- Definition of the simulation path
- Definition of the process and device simulation input decks

The first part describes the settings of the process parameters as can be seen in table (5.1). The second part describes which DOE has to be used and handles the input and output of the process and device simulator programs. The third part consists the device and process simulator input decks.

5.3 Analysis of the experiment results

5.3.1 Introduction

For a full description of the analysis see internal report [RNB-E88-97/288].

The responses looked at in the experiment are the base current I_b, the collector current I_c and Hfe. The relation between the currents and the base-emitter voltage is an exponential function known as the pn-junction formula. Therefore the logarithm of the responses are analysed.
At each Vbe setting (25 settings between 0.4 and 1.0 V) a model per response is built. All effects that significantly influence the response for at least one setting of Vbe are adopted in the models per Vbe setting. This results in 25 models for each response. Each effect (main linear, main square or interaction) has its own coefficient, depending on the Vbe setting taken into account. In the next sections the models for all three responses are discussed.

5.3.2 Base current

5.3.2.1 The base current model based on a Vbe between 0.4 and 1.0 V

Building a model for the log of the base current results in a model including the mean, the energy (A), the square of energy (A2), the dose (B), the etch back (C), the square of etch back (C2) and the temperature (D). The interactions between energy and dose (AB), energy and etch back (AC), energy and temperature (AD) and dose and etch back (BC) also differ significantly from zero. In table (5.3) the coefficients of these effects as function of VBE are shown.

<table>
<thead>
<tr>
<th>effect</th>
<th>coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-34.862-48.751<em>VBE+266.985</em>VBE(^2)-427.895<em>VBE(^3)+360.815</em>VBE(^4)-125.947*VBE(^5)</td>
</tr>
<tr>
<td>A</td>
<td>-4.561+29.167<em>VBE-70.055</em>VBE(^2)+78.958<em>VBE(^3)-41.002</em>VBE(^4)+7.491*VBE(^5)</td>
</tr>
<tr>
<td>A(^2)</td>
<td>11.486-71.798<em>VBE+180.687</em>VBE(^2)-228.914<em>VBE(^3)+145.535</em>VBE(^4)-37.007*VBE(^5)</td>
</tr>
<tr>
<td>B</td>
<td>-7.796+52.574<em>VBE-139.443</em>VBE(^2)+181.888<em>VBE(^3)-116.868</em>VBE(^4)+29.658*VBE(^5)</td>
</tr>
<tr>
<td>C</td>
<td>14.126-88.467<em>VBE+217.167</em>VBE(^2)-261.116<em>VBE(^3)+153.761</em>VBE(^4)-35.454*VBE(^5)</td>
</tr>
<tr>
<td>C(^2)</td>
<td>18.541-116.889<em>VBE+294.043</em>VBE(^2)-367.882<em>VBE(^3)+228.665</em>VBE(^4)-56.481*VBE(^5)</td>
</tr>
<tr>
<td>D</td>
<td>-1.561+10.585<em>VBE-30.438</em>VBE(^2)+46.4<em>VBE(^3)-36.656</em>VBE(^4)+11.74*VBE(^5)</td>
</tr>
<tr>
<td>AB</td>
<td>0.775-1.017<em>VBE-8.001</em>VBE(^2)+23.556<em>VBE(^3)-23.461</em>VBE(^4)+8.149*VBE(^5)</td>
</tr>
<tr>
<td>AC</td>
<td>-10.092+53.522<em>VBE-112.596</em>VBE(^2)+116.721<em>VBE(^3)-59.375</em>VBE(^4)+11.819*VBE(^5)</td>
</tr>
<tr>
<td>AD</td>
<td>-1.719+8.026<em>VBE-13.994</em>VBE(^2)+11.315<em>VBE(^3)-4.367</em>VBE(^4)+0.75*VBE(^5)</td>
</tr>
<tr>
<td>BC</td>
<td>-13.959+90.794<em>VBE-232.781</em>VBE(^2)+293.917<em>VBE(^3)-182.544</em>VBE(^4)+44.546*VBE(^5)</td>
</tr>
</tbody>
</table>

The coefficients in table (5.3) are based upon the standardized settings of the parameters. This means that the low values of the parameters have value -1, the default setting has value 0 and the high setting has value 1.

This model fits well, as also shown by the performance indicators $R^2$, which denotes the percentage of variation explained by the model, and $R^2_{adj}$, which adjusts $R^2$ for the number of estimated parameters. For this model we have $R^2 = 0.9996$ and $R^2_{adj} = 0.9996$. This is also shown in figure (5.1), where the observed and predicted ln($I_b$) are plotted against each other.

When the residuals are investigated, it is clear that the model fits better for higher Vbe values (figure 5.2). Clearly there still remains some systematic effect, not adopted in the model. This is probably due to a lack of fit of the models per Vbe setting and due to building the overall model for all Vbe settings.
In the normal operating region of the transistor the base-emitter voltage is 0.66 V. For this value of Vbe it is nice to see, when changing the 4 process parameters, what the influence is on the base current. Therefore the sensitivity codes are plotted in figure (5.3) for only those parameters who's interaction has influence on the base current (see table 5.3 AB, AC, AD and BC). Note that the sensitivity plots are based upon a model predicting Vbe in the range [0.4V, 1.0 V].
5.3.2.2 Conclusions

With a base implantation energy lower than the default value the base current increases with increasing base implantation dose. This is due to the fact that a higher base implantation dose increases the base current due to a higher recombination rate in the base and to a possible decreased emitter gummel number. With a base implantation energy above the default value a reduced base current is observed when the base implantation dose is increased. This effect can probably be described towards an increased emitter gummel number.

With increasing RTA drive temperature the base current increases. With increasing temperature the emitter gummel number increases which counts for a decreasing base current. Therefore the effect of increasing temperature will have more effect on the recombination rate in the base then on the increase of the emitter gummel number. Therefore the base current increases with increasing RTA drive temperature.

The sensitivity plot of energy and dose and energy and etch back show saddle points. For the factors dose and etch back the result is that the base current increases with decreasing dose. For constant dose, the base current is lowest at the default setting of the etch back (10 nm).
Fig. 5.3: Sensitivity curves for the base current modeled for Vbe between 0.4 V - 1.0 V
5.3.2.3 The base current model based on a $V_{be}$ between 0.6 and 0.7 V

Since the base current is not well modeled in the simulations for comparison with measured results over the whole range of base-emitter voltages a new model is made for a smaller range of $V_{be}$ between 0.6 V and 0.7 V. The base current was not well modeled in the simulations for the low and high injection regions see table (4.4). The coefficients for the new model can be seen in table (5.4). The coefficients as function of $V_{be}$ (between 0.6 V - 0.7 V) are approximated by a 2nd degree polynomial while the coefficients as function of $V_{be}$ (between 0.4 V - 1.0 V) are modeled by a 5th degree polynomial.

**TABLE 5.4: Coefficients for $\ln(I_b)$ for $V_{be}$ between 0.6 V and 0.7 V.**

<table>
<thead>
<tr>
<th>effect</th>
<th>coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>$-46.701 + 38.191 \cdot V_{BE} + 0.948 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>A</td>
<td>$0.325 - 0.767 \cdot V_{BE} + 0.485 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>B</td>
<td>$0.05 - 0.165 \cdot V_{BE} + 0.12 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>C</td>
<td>$0.049 - 0.237 \cdot V_{BE} + 0.255 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>C2</td>
<td>$0.369 - 0.843 \cdot V_{BE} + 0.52 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>D</td>
<td>$-0.115 + 0.402 \cdot V_{BE} - 0.303 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>AB</td>
<td>$0.308 - 1.018 \cdot V_{BE} + 0.747 \cdot V_{BE}^2$</td>
</tr>
<tr>
<td>AC</td>
<td>$-1.007 + 2.453 \cdot V_{BE} - 1.518 \cdot V_{BE}^2$</td>
</tr>
</tbody>
</table>

The fit of this model is shown in figure (5.4), where the predicted $\ln(I_b)$ is plotted versus the observed $\ln(I_b)$. It is shown that this fit is very well.

![Fig. 5.4: Observed versus predicted $\ln(I_b)$ for $V_{be}$ between 0.6 V - 0.7 V](image)

The sensitivity curves for $V_{be} = 0.66$ V for this model are given in figure (5.5). These plots are more reliable than the plots for the $V_{be}$ range between 0.4 V - 1.0 V because of the better modeled base current in this operating region of the transistor.
5.3.2.4 Conclusions

For a given implant energy level, the base current will increase when the dose decreases or the temperature increases. With increasing RTA drive temperature the base current increases. With increasing temperature the emitter gummel number increases which counts for a decreasing base current. Therefore the effect of increasing temperature will have more effect on the recombination rate in the base then on the increase of the emitter gummel number. Therefore the base current increases with increasing RTA drive temperature.

The sensitivity curves of energy and etch back show a saddle point.
Fig. 5.5: Sensitivity curves for the base current modeled for Vbe between 0.6 V - 0.7 V
5.3.3 The collector current

5.3.3.1 The collector current model based on a Vbe between 0.4 and 1.0 V

The natural logarithm of the collector current can be described by a model including the mean, the energy (A), the square of energy (A²), the dose (B), the etch back (C) and the temperature (D). The interactions between energy and dose (AB), energy and etch back (AC) and energy and temperature (AD) differ significantly from zero. In table (5.5) the coefficients of these effects as a function of Vbe are shown. Note that these coefficients are based upon standardized settings of the factors.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-82.564 +356.194<em>VBE-948.581</em>VBE²+1350.59<em>VBE³-903.819</em>VBE⁴+221.373*VBE⁵</td>
</tr>
<tr>
<td>A</td>
<td>-3.385+29.571<em>VBE-98.706</em>VBE²+160.176<em>VBE³-126.692</em>VBE⁴-39.023*VBE⁵</td>
</tr>
<tr>
<td>A²</td>
<td>5.329-45.096<em>VBE+143.109</em>VBE²-220.415<em>VBE³+164.622</em>VBE⁴-47.574*VBE⁵</td>
</tr>
<tr>
<td>B</td>
<td>5.953-50.527<em>VBE+163.801</em>VBE²-258.509<em>VBE³+198.551</em>VBE⁴-59.257*VBE⁵</td>
</tr>
<tr>
<td>C</td>
<td>-4.908+42.697<em>VBE-138.612</em>VBE²+218.522<em>VBE³-167.553</em>VBE⁴+49.882*VBE⁵</td>
</tr>
<tr>
<td>D</td>
<td>-4.644+40.496<em>VBE-131.992</em>VBE²+208.745<em>VBE³-160.558</em>VBE⁴+47.993*VBE⁵</td>
</tr>
<tr>
<td>AB</td>
<td>-4.791+41.649<em>VBE-135.411</em>VBE²+213.094<em>VBE³-163.037</em>VBE⁴+48.506*VBE⁵</td>
</tr>
<tr>
<td>AC</td>
<td>-3.154+25.974<em>VBE-81.962</em>VBE²+125.684<em>VBE³-93.564</em>VBE⁴+27.018*VBE⁵</td>
</tr>
<tr>
<td>AD</td>
<td>-0.7+3.826<em>VBE-7.479</em>VBE²+3.703<em>VBE³+3.626</em>VBE⁴+2.981*VBE⁵</td>
</tr>
</tbody>
</table>

For this model we have $R^2 = 0.9989$ and $R^2_{adj} = 0.9988$. Thus the model fits well, as can be seen in figure (5.6) where the observed and predicted In ($I_c$) are given against each other.

![Figure 5.6: Observed versus predicted In ($I_c$)](image)

In figure (5.7) the residuals (observed values minus the predicted values) are plotted versus the settings of Vbe. When the residuals are investigated, it is clear that the model fits better for values of Vbe close to 1.
Fig. 5.7: Residuals versus Vbe when predicting ln (Ic).

In figure (5.8) the sensitivity curves are given for Vbe = 0.66 V for the factors in the experiment.

5.3.3.2 Conclusions

The collector current decreases for an increased base implantation dose. This is due to the fact that the base gummel number is increased for a higher base implantation dose.

The collector current increases for an increased RTA drive temperature. This is due to the fact that the base gummel number is decreased for a higher RTA drive temperature.

The sensitivity plot for the RTA drive temperature against the base implantation energy is almost the same as the sensitivity plot for the silicon etch back under the polysilicon layer against again the base implantation energy. Changing the etch back therefore has the same effect as a change in the RTA drive temperature. Increasing the etch back will also give a lower base gummel number and therefore a higher collector current.

A model for the collector current over a range of Vbe between 0.6 V and 0.7 V was not necessary because the collector current is well modeled over the whole range of Vbe's as can be seen in table (4.4). To check this a model was made for this range of Vbe and indeed the sensitivity curves were the same as in figure (5.8).
With band gap narrowing

Fig. 5.8: Sensitivity curves for the collector current modeled for Vbe between 0.4 V - 1.0 V
5.3.4 The \( H_{fe} \)

5.3.4.1 The \( H_{fe} \) model based on a \( V_{be} \) between 0.4 and 1.0 V

For the log of \( H_{fe} = \ln \left( \frac{I_c}{I_b} \right) \) the model includes the mean, the energy (A), the square of energy (A²), the dose (B), the etch back (C), the temperature (D) and the square of temperature (D²). The interactions between energy and dose, energy and etch back and energy and temperature also differ significantly from zero. In table (5.6) the coefficients of these effects as function of \( V_{be} \) are shown. Note that these coefficients are based upon standardised settings of the factors.

**TABLE 5.6: Coefficients for \( \ln (H_{fe}) \) for a \( V_{be} \) between 0.4 V and 1.0 V**

<table>
<thead>
<tr>
<th>effect</th>
<th>coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>62.087 - 500.022 * ( V_{be} ) + 1464.87 * ( V_{be}^2 ) + 2101.087 * ( V_{be}^3 ) + 1470.33 * ( V_{be}^4 ) - 399.037 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>A</td>
<td>-7.994 + 58.543 * ( V_{be} ) - 167.116 * ( V_{be}^2 ) + 235.851 * ( V_{be}^3 ) - 165.055 * ( V_{be}^4 ) + 45.755 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>A²</td>
<td>-0.566 + 17.447 * ( V_{be} ) - 77.548 * ( V_{be}^2 ) + 139.824 * ( V_{be}^3 ) - 113.352 * ( V_{be}^4 ) + 34.212 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>B</td>
<td>-13.578 + 101.52 * ( V_{be} ) - 298.085 * ( V_{be}^2 ) + 432.686 * ( V_{be}^3 ) - 310.01 * ( V_{be}^4 ) + 87.468 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>C</td>
<td>18.908 - 129.364 * ( V_{be} ) + 348.907 * ( V_{be}^2 ) - 468.542 * ( V_{be}^3 ) + 313.193 * ( V_{be}^4 ) - 83.11 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>D</td>
<td>5.606 - 46.099 * ( V_{be} ) + 142.767 * ( V_{be}^2 ) - 214.277 * ( V_{be}^3 ) + 156.25 * ( V_{be}^4 ) - 44.216 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>D²</td>
<td>-4.579 + 35.246 * ( V_{be} ) - 103.357 * ( V_{be}^2 ) + 144.807 * ( V_{be}^3 ) - 97.496 * ( V_{be}^4 ) + 25.393 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>AB</td>
<td>5.566 - 42.663 * ( V_{be} ) + 127.403 * ( V_{be}^2 ) - 189.526 * ( V_{be}^3 ) + 139.566 * ( V_{be}^4 ) - 40.354 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>AC</td>
<td>-6.802 + 28.187 * ( V_{be} ) - 35.733 * ( V_{be}^2 ) + 1.129 * ( V_{be}^3 ) + 26.097 * ( V_{be}^4 ) - 12.874 * ( V_{be}^5 )</td>
</tr>
<tr>
<td>AD</td>
<td>1.581 - 12.289 * ( V_{be} ) + 35.205 * ( V_{be}^2 ) - 44.853 * ( V_{be}^3 ) + 24.762 * ( V_{be}^4 ) - 4.389 * ( V_{be}^5 )</td>
</tr>
</tbody>
</table>

For this model we have \( R^2 = 0.936 \) and \( R^2_{adj} = 0.931 \). Thus the model fits well, as can be seen in figure (5.9), where the observed and predicted \( \ln (H_{fe}) \) are plotted against each other. It is shown that the fit is less than for the base current and collector current.

![Observed versus predicted \( \ln (H_{fe}) \)](image.png)

**Fig. 5.9: Observed versus predicted \( \ln (H_{fe}) \)**

When the residuals are investigated, it shows that the model fits better for higher \( V_{be} \) values (figure 5.10). Clearly there is still some systematic effect left, not adopted in
the model. In figure (5.10) the residuals (observed values minus the predicted values) are plotted versus the settings of Vbe.

![Graph showing residuals versus Vbe](image)

**Fig. 5.10: Residuals versus Vbe when predicting ln (Hfe)**

In figure (5.11) the sensitivity curves are given for VBE=0.66 and the factors in the experiment.

### 5.3.4.2 Conclusions

The sensitivity plots show that for a given energy setting, the Hfe increases when the dose decreases, the etch back increases or the temperature increases. The reasons for this can be described to increasing or decreasing base/collector currents as described in sections 5.3.2.2, 5.3.2.4 and 5.3.3.2.
With band gap narrowing

Fig. 5.11: Sensitivity curves for the gain (Hfe) modeled for Vbe between 0.4 V - 1.0 V
VI. General conclusions and recommendations for future work

Well calibrated input decks for the process and device simulation of a polysilicon emitter bipolar transistor have been designed. There still is some problem in the accurate simulation of the base current for such a device. In the low injection region and in the high injection region the device simulator MEDICI can not model the base current in a correct way. To improve the modeling of the base current a polycrystalline model was introduced in the process simulator. An improved base current is obtained with the use of this model but exact modeling is still not possible. Therefore this model needs to be updated in the future and therefore a better understanding of the physical mechanisms that occur at the polysilicon/silicon interface is needed. By changing the value for the effective recombination velocity in the device simulator the base current can be changed to obtain a good match with reality. Due to the fact that the value of the effective recombination velocity varies significantly with changes in processing parameters and the exact dependence of these variations is not known one can use the effective recombination velocity as a match factor. The exact dependence of this effective recombination velocity on processing parameters needs to be investigated in the future by doing matrix experiments on real silicon. The simulations showed quite surprisingly that the threshold voltage adjust implantation had significant effect on the electrical behavior of the polysilicon emitter bipolar transistor. Tests have to be done on real silicon to see the exact influence of this processing step on the electrical characteristics of the bipolar transistor. The statistical experiment brought up some very useful sensitivity plots. With the help of these plots the influence of variations and interactions between the base implant dose and energy, the etch back of silicon underneath the polysilicon and the temperature of the RTA drive on the electrical parameters of the device can be understood.
References

[1] Ashburn, P.
DESIGN AND REALIZATION OF BIPOLAR TRANSISTORS. 1st ed.
Southampton 1987

THE ROLE OF THE INTERFACIAL LAYER IN POLYSILICON EMITTER BIPOLAR TRANSISTORS

ON THE MODELING OF POLYSILICON EMITTER BIPOLAR TRANSISTORS

POLYSILICON EMITTERS FOR BIPOLAR TRANSISTORS: A REVIEW AND RE EVALUATION OF THEORY AND EXPERIMENT

COMPARISON OF EXPERIMENTAL AND THEORETICAL RESULTS ON POLYSILICON EMITTER BIPOLAR TRANSISTORS

GUMMEL PLOT NONLINEARITIES IN POLYSILICON EMITTER TRANSISTORS-INCLUDING NEGATIVE DIFFERENTIAL RESISTANCE BEHAVIOR

MEASUREMENTS OF BANDGAP NARROWING IN SI BIPOLAR TRANSISTORS

MEASUREMENT OF STEADY-STATE MINORITY-CARRIER TRANSPORT PARAMETERS IN HEAVILY DOPED N-TYPE SILICON

[9] Swirhun, S.E. et al.
MEASUREMENT OF ELECTRON LIFETIME, ELECTRON MOBILITY AND BANDGAP NARROWING IN HEAVILY DOPED P-TYPE SILICON

A UNIFIED MOBILITY MODEL FOR DEVICE SIMULATION-I. MODEL EQUATIONS AND CONCENTRATION DEPENDENCE

A UNIFIED MOBILITY MODEL FOR DEVICE SIMULATION-II. TEMPERATURE DEPENDENCE OF CARRIER MOBILITY AND LIFETIME
PHYSICAL MODELS FOR DEVICE SIMULATION
Nat. Lab. report Nr. 6799/94 (1994)

STUDIES OF DIFFUSED BORON EMITTERS: SATURATION CURRENT, BANDGAP NARROWING, AND SURFACE RECOMBINATION VELOCITY

EFFECTIVE RECOMBINATION VELOCITY OF POLYSILICON CONTACTS FOR BIPOLAR TRANSISTORS
Electronic Letters (1984), Vol 20, NO 15, p 622-624

[15] Roulston, D.J.
BIPOLAR SEMICONDUCTOR DEVICES INTERNATIONAL EDITION 1st. ed.

[16] MEDICI MANUAL, version 2.1.2
Technology Modeling Associates (TMA), Inc., Sunnyvale, CA USA.

STATISTICAL DESCRIPTION AND ANALYSING OF EXPERIMENTS

[18] Booth, R. et al.
NORMAN/DEBORA manual
Imec, Leuven, Belgie 1995

[19] SUPREM-4 manual, version 6.3.0
Technology Modeling Associates (TMA), Inc., Palo Alto, CA USA
Acknowledgment

The author of this report thanks the following people:

Som Nath, for giving the opportunity to work on this project and for his coaching during the work.

Anco Heringa, for helping with UNIX problems and his overall coaching.

Marijke Swaving, for her help with setting up and analysing the statistical part.

Wim Peters and Mark Murphy, for their help on process technology discussions.

Prof. Dr. L.M.F. Kaufmann and Prof. Dr. F.M. Klaassen, for their support from the university.