High Frequency Characterization and Modeling of SiGe Heterojunction Bipolar Transistors

B. Gunnar Malm

Doctoral Thesis, 2002

KTH, Royal Institute of Technology
Department of Microelectronics and Information Technology
Laboratory of Solid State Devices
Stockholm, Sweden
High Frequency Characterization and Modeling of SiGe Heterojunction Bipolar Transistors

A dissertation submitted to Kungliga Tekniska Högskolan, Stockholm, Sweden, in partial fulfillment of the requirements for the degree of Teknisk Doktor.

© 2002 B. Gunnar Malm
KTH, Royal Institute of Technology
Department of Microelectronics and Information Technology
Electrum 299
SE-164 40 KISTA
SWEDEN

ISRN KTH/EKT/FR-2002/2-SE
ISSN 1650-8599
TRITA – EKT
Forskningsrapport 2002:2

Printed in 250 copies by Kista Snabbtryck AB, Kista 2002.
**Abstract**

High-speed, low voltage Silicon-Germanium (SiGe) heterojunction bipolar transistors (HBTs) have been designed, fabricated, electrically characterized and modeled. The SiGe HBTs are suitable for use in radio frequency (RF) integrated circuit (IC) applications and were fabricated using non-selective epitaxial growth. The design of the extrinsic base region has been investigated in detail. Transient enhanced diffusion of boron, caused by the extrinsic base implantation, was found to degrade DC and high-frequency electrical characteristics. It was also found that the low-frequency noise was affected by the base design. Furthermore, a hydrogen anneal was found to reduce the low-frequency noise in polysilicon emitter bipolar transistors. The high-frequency noise of SiGe HBTs was investigated experimentally as well as by device simulation. Noise parameter extraction methods based on direct admittance or Y-parameter measurement have been investigated in detail. Good agreement was found with conventional noise figure measurement. SiGe HBTs were fabricated to investigate the optimization of ion-implanted collector doping profiles. A novel concept was suggested where a low-energy (5–10 keV) antimony (Sb) implantation was combined with a standard selectively implanted collector (SIC) using phosphorous. Segregation of Sb was found to occur during the subsequent growth of the epitaxial SiGe base layer. The resulting broadening of the implanted Sb-profile degraded the DC-electrical characteristics of the device. The devices with an Sb-implantation exhibited a cut-off frequency of more than 60 GHz. A mixed-mode circuit and device simulation methodology was developed to investigate the RF harmonic distortion of SiGe HBTs. The influence on harmonic distortion of the Ge-profile as well as the collector doping profile was quantified. High-injection heterojunction barrier effects due to the presence of a valence band offset at the base-collector junction were found to significantly affect the harmonic distortion. Devices with a Ge-profile retrograded towards the collector exhibited significantly reduced harmonic distortion. Increasing the Ge-concentration at the base-emitter junction led to reduced harmonic distortion for low current operation. A non-uniform collector doping profile was shown to suppress the harmonic distortion and increase the breakdown voltage.

**Keywords**: Silicon-Germanium (SiGe), heterojunction bipolar transistor (HBT), low-frequency noise, high-frequency noise, harmonic distortion, linearity, device simulation, collector profile, epitaxial base integration, radio frequency (RF), radio frequency integrated circuit (RFIC).
Acknowledgments

First of all I would like to thank my supervisor professor Mikael Östling for giving me the opportunity to pursue the Ph.D. degree and for always supporting my work, reminding me to think positively and for suggesting many interesting research topics. I am also very grateful to all members of the bipolar group, especially Martin Linder, Martin Sandén and Johan Pejnefors. I am thankful to the co-authors of the appended papers for their ideas and comments. I would also like to acknowledge the collaboration with Ted Johansson, Torkel Arnborg and others at Ericsson Microelectronics AB. This work was financially supported by the Swedish Foundation for Strategic Research (SSF).

I dedicate this thesis to Ylva.

The fear of the Lord is the beginning of wisdom, and knowledge of the Holy One is understanding.

Proverbs 9:10

Att frukta HERREN är början till vishet, att känna den Helige är förstånd.

Ords. 9:10
Appended papers

I. Ge-profile Design for Improved Linearity of SiGe Double Heterojunction Bipolar Transistors
B. Gunnar Malm and Mikael Östling

II. Mixed Mode Circuit and Device Simulation of RF Harmonic Distortion for High-Speed SiGe HBTs
B. Gunnar Malm and Mikael Östling
Accepted for publication in Solid-State Electronics, 2002.

III. Lateral Base Design Rules for Optimized Low-Frequency Noise of Differentially Grown SiGe Heterojunction Bipolar Transistors
Martin Sandén, B. Gunnar Malm, Jan V. Grahn and Mikael Östling

IV. Implanted Collector Profile Optimization in a SiGe HBT Process
B. Gunnar Malm, Ted Johansson, Torkel Arnborg, Hans Norström, Jan V. Grahn and Mikael Östling

V. Influence of Transient Enhanced Diffusion of the Intrinsic Base Dopant Profile on SiGe HBT DC and HF Characteristics
B. Gunnar Malm, Jan V. Grahn and Mikael Östling

VI. Comparison of two High-Frequency Noise Characterization Methods for SiGe HBTs
B. Gunnar Malm, Jan V. Grahn, Mikael Östling, Jörgen Stenarson

VII. Decreased Low Frequency Noise by Hydrogen Passivation of Polysilicon Emitter Bipolar Transistors
Martin Sandén, B. Gunnar Malm, Jan V. Grahn, Mikael Östling
Related papers not included in the thesis

VIII. Influence of Ge-profile Design and Saturation Effects on SiGe HBT Linearity
B. Gunnar Malm and Mikael Östling
In proceedings of 31\textsuperscript{th} ESSDERC, pp. 451 - 454, 2001.

IX. Impact on Low-Frequency Noise Properties from Lateral Design of Differentially Grown SiGe HBTs
Martin Sandén, B. Gunnar Malm, Jan V. Grahn, Mikael Östling
In proceedings of 30\textsuperscript{th} ESSDERC, pp. 564 - 567, 2000.

In proceedings of 29\textsuperscript{th} ESSDERC, pp. 741 - 744, 1999.

XI. A low-Complexity 62-GHz $f_T$ SiGe HBT Process Using Differential Epitaxy and \textit{in situ} Phosphorous-Doped Poly-Si Emitter at Very Low Thermal Budget

XII. The Profile Control of n-Type Doping in Low and High Temperature Si Epitaxy for High Frequency Bipolar Transistors
H. H. Radamson, B. Mohadjer, B. G. Malm, J. V. Grahn, M. Östling and G. Landgren
Summary of Appended Papers

**Paper I.** This paper investigates the RF harmonic distortion for high-speed low-voltage SiGe HBTs. A novel simulation methodology, which includes the influence of the circuit environment is introduced and verified by experimental results. The influence of the Ge-profile and the collector doping profile on the harmonic distortion is quantified. The author suggested the study and performed all simulations and measurements and wrote the manuscript.

**Paper II.** This paper describes a mixed-mode circuit and device simulation methodology of harmonic distortion in SiGe HBTs. The simultaneous optimization of RF-distortion and noise properties is investigated. The different trade-offs for collector profile optimization for high breakdown voltage, improved high current operation and reduced bias and current dependence of the collector-base capacitance are discussed. This work is a continuation of paper I. The author performed all simulations and wrote the manuscript.

**Paper III.** The low-frequency noise characteristics of SiGe HBTs fabricated at KTH are investigated. The influence of various changes in the design of the extrinsic base region is discussed with respect to the low-frequency noise properties. The author performed electrical characterization of the devices and contributed to the noise measurement. The author was also involved in the planning of the study as well as performing the analysis and writing the manuscript.

**Paper IV.** This paper discusses optimization of an advanced SiGe HBT structure. Process and device simulation were used to find suitable epitaxial SiGe base and implanted collector profiles. Devices with different implanted collector profiles were fabricated and compared to the predicted results from the device simulations. A novel collector fabrication concept was proposed, where low-energy implantation of antimony (Sb) was combined with a standard high-energy implantation of phosphorous (P). The author performed all simulations as well as the major part of the electrical characterization and wrote the manuscript.

**Paper V.** The paper presents an in-depth electrical characterization of SiGe HBTs fabricated in a process developed at KTH. The main results concern the influence of transient enhanced diffusion on DC and HF characteristics. Device simulations, based on the measured SIMS-profiles from fabricated devices were calibrated to the measured DC and HF results. The author performed all electrical characterization and device simulations. The author has also contributed to the process development and device fabrication.

**Paper VI.** This paper investigates a novel high-frequency noise characterization method, based on S-parameter measurement, using a network analyzer. This method was compared to results, obtained from conventional noise figure measurement. The author performed the S-parameter measurements and took part in the noise figure measurement. The author suggested the study, performed the analysis, the necessary calculations and wrote the manuscript.
Paper VII. The low-frequency noise in poly-silicon emitter Si-bipolar transistors is investigated. The hydrogen anneal (FGA) at the end of the processing is shown to reduce the low-frequency noise. The author contributed to the measurement and the analysis of the results as well as writing the manuscript. The author also took part in establishing the low-frequency noise set-up and measurement routines.
1 Introduction

This thesis is concerned with design, characterization and optimization of high-speed, low-voltage silicon-germanium (SiGe) heterojunction bipolar transistors (HBTs). The SiGe semiconductor devices are intended for use in electronic circuits for radio frequency (RF) applications as well as in high-speed optical networks. One example of these electronic circuits is the transceiver/receiver part of a cellular phone. An RF circuit is in turn composed by several smaller building blocks, such as low-noise amplifiers, power amplifiers, mixers and oscillators. In all of these the SiGe HBTs are needed to provide the gain or amplification of the RF-signal.

The dominating semiconductor technology today is silicon (Si) complementary-metal-oxide-semiconductor (CMOS) field-effect transistors. Electronic circuits fabricated in Si CMOS technology are used in a wide variety of consumer applications, ranging from computers to pacemakers. It would also be desirable to use CMOS for RF-applications such as cellular phones and wireless networks. For RF-applications transistors with high cut-off frequency, high power gain and low high-frequency noise are needed. SiGe HBTs with a narrow and highly doped epitaxial base are well suited to meet all of these requirements. Traditionally, so-called III-V compound HBTs have been used in many RF circuits. However, SiGe HBT based circuits offer a simpler and more cost-effective solution. A SiGe HBT process module can be integrated in a standard Si CMOS process flow, which reduces the fabrication cost and enables high yield. Furthermore SiGe HBTs and Si CMOS transistors can be used in the same circuit applications for added functionality. Another desirable property of SiGe HBTs is the low DC-power consumption, which can be used to extend battery lifetimes.

The SiGe technology development started about 13 years ago, by the successful demonstration of SiGe HBTs [1, 2]. Today commercial circuit applications are available from several vendors worldwide. Devices with cut-off frequency and maximum frequency of oscillation close to 200 GHz have recently been demonstrated [3-5]. These ultra-high-speed devices are suitable for use in e.g. 10 and 40 Gbit optical networks. This application area is not further discussed in this thesis. The rapid improvement in SiGe HBT performance is mainly due to careful tailoring of Ge and doping profiles, and also due to the continuing downscaling of device dimensions. Today state-of-the-art SiGe HBTs are fabricated using 0.18 μm deep UV lithography. Material growth development is also an important factor for the success of SiGe technology and the most recent example is the incorporation of carbon in the SiGe epitaxial layers, which further improves performance.

The focus of this thesis is on optimization of SiGe HBTs for RF-applications. Devices have been fabricated in two different SiGe HBT processes and thoroughly characterized by electrical measurements. The optimization of the Ge-profile and the collector region for high speed, low noise and reduced RF-distortion will be discussed in detail. Profile optimization is a task well suited for device simulation. Calibrated physical device simulations have been extensively used throughout thesis work. In SiGe HBTs with a non-selectively grown epitaxial base the layout of the
extrinsic base region becomes a central issue. In order to obtain the desired performance, transient enhanced diffusion of the vertical base profile should be minimized or avoided. For the investigated devices the base region design also affects the low-frequency noise, which is important for many RF-applications. Finally, it is of great interest to investigate the SiGe HBT performance in realistic circuit applications. Therefore a mixed-mode circuit and device simulation methodology has been developed, to investigate RF harmonic distortion for devices with different Ge-profiles and collector doping profiles.

The outline of the thesis is summarized below. In Chapter 2 we give a brief background to SiGe HBT device physics. The influence of Si/SiGe heterojunctions and a non-uniform energy bandgap in the base is discussed. This is followed by a background to finite element device simulation. The physical models and parameters for strained SiGe, which have been implemented in the simulator, are described in some detail. Process integration issues for epitaxial base and collector region fabrication has been investigated in two different SiGe HBT processes. In Chapter 3 SiGe epitaxial base integration is discussed, together with some results from paper V, concerning extrinsic base region design. Chapter 4 discusses collector fabrication. A novel concept for formation of highly doped launcher layer, using low energy antimony implantation is described. Experimental and simulation results from paper IV are used to illustrate the important trade-off between cut-off frequency and breakdown voltage. Chapters 5, 6 and 7 summarize RF-optimization issues. Chapter 5 deals with high-frequency noise characterization and modeling. These results are based on papers II and VI, together with additional measurements and simulations. In Chapter 6 on RF-distortion, we first give an in-depth background to the subject and then continue by discussing the major results from papers I and II. Guidelines for the choice of Ge-profile for suppressed harmonic distortion are given. Finally the low-frequency noise properties of Si and SiGe bipolar devices are discussed in Chapter 7. These results are based on the more process-oriented papers III and VII. Some concluding remarks about the findings and future perspectives for SiGe based semiconductor technology are given in Chapter 8.
2 SiGe HBT device physics and simulation

In this chapter the basic principles of heterojunction bipolar transistor operation are introduced. The influence of the Si/SiGe heterojunctions and a non-uniform energy bandgap on electrical characteristics such as current gain, Early voltage, and cut-off frequency is discussed. For a further discussion about bipolar transistor operation the reader is referred to textbooks [6, 7]. Finite element physical device simulations are introduced and the physical models for mobility, effective density of states and energy bandgap in strained SiGe are discussed. Finally, we briefly consider calibration of simulated DC- and HF-characteristics to measurement.

2.1 Electrical characteristics

The SiGe HBT derives its fundamental advantage over conventional Si bipolar transistors from a principle called bandgap engineering. Bandgap engineering can be used to improve several key figure-of-merits for a bipolar transistor, e.g. current gain ($\beta$), Early voltage ($V_A$) and base transit time ($\tau_B$). In this thesis we will also discuss RF-properties, such as noise figure and harmonic distortion. In SiGe HBTs the intrinsic base region is formed by an epitaxially grown strained Si$_{1-x}$Ge$_x$ alloy. Si$_{1-x}$Ge$_x$ refers to the binary alloy of Si and Ge, where $x$ is the Ge composition, below we often write SiGe for short. The Ge composition is typically between 5 and 30 atomic percent and can be varied inside the base. As an example of a profile with varying Ge concentration we will consider the graded or trapezoidal profile, where the Ge concentration increases linearly from the emitter side to the collector side of the base. The SiGe layer is in situ doped with boron during epitaxial growth, which enables precise control of the base doping profile and base width.

Semiconductor materials such as Si and Ge exhibit an energy band structure, with conduction bands populated by electrons, valence bands populated by holes and a forbidden energy bandgap ($E_g$). The magnitude of $E_g$ is given by the difference between the conduction band energy ($E_C$) and the valence band energy ($E_V$). Compressively strained Si$_{1-x}$Ge$_x$ has a smaller $E_g$ than Si. The reduction in $E_g$ has been determined experimentally [8] and is given by:

$$\Delta E_g = E_{g, Si} - E_{g, SiGe} = 0.896x - 0.396x^2$$

Eq. 1

Most of the bandgap reduction ($\Delta E_g$) in strained SiGe occurs as an offset in the valence band. SiGe HBTs are of the double heterojunction type, with a narrow bandgap SiGe-base and wider bandgap in the emitter and collector Si-regions. Thus heterojunctions are formed close to the base-emitter and the base-collector pn-junctions. The heterojunctions will influence the electron and hole transport. This will directly affect the terminal currents of the bipolar transistor – the base current
(I_B) and the collector current (I_C). The current gain $\beta$ is defined as the ratio between $I_C$ and $I_B$:

$$\beta = \frac{I_C}{I_B} \approx \exp \left( \frac{\Delta E_g}{kT} \right)$$

Eq. 2

Where $k$ is Boltzmann’s constant and $T$ is the temperature. The increase in $\beta$ will be exponentially related to $\Delta E_g$. We first discuss the $\beta$ increase qualitatively. This is followed by a more formal derivation. We begin by comparing $\beta$ for a Si BJT and a SiGe HBT, respectively. In this example a constant Ge-concentration of approximately 12 % was used in the neutral base region. We consider an npn-type bipolar transistor under forward bias, i.e., when a positive voltage $V_{BE}$ is applied between the base and emitter terminals. The valence and conduction bands energies $E_V$ and $E_C$ for this situation are schematically shown in Fig 1. Electrons are injected into the base from the emitter (left). They are transported through the base to the collector side (right) mainly by diffusion. It is clear that the electrons encounter a potential barrier between emitter and base. This barrier is reduced by the presence of Ge by an amount $\Delta E_C$. The collector current $I_C$ is exponentially related to the barrier height and hence will be very sensitive to the Ge-concentration at the base-emitter heterojunction.

![Energy band diagram for 0.6 V forward bias for a Si BJT (dotted lines) and a SiGe HBT (solid lines).](image)

The base current $I_B$ is mainly composed of holes, which are injected into the emitter from the base. The base current magnitude is controlled by the hole diffusion and the hole recombination in the polysilicon emitter. For a fixed $V_{BE}$, $I_B$ remains unchanged by the introduction of Ge in the base, and therefore $\beta$ is increased due to the higher $I_C$. In the following section we give a formal derivation of the collector
current density, \( J_n = I_c / A_E \), where \( A_E \) is the emitter area. The collector current density \( J_n \) can be obtained from the generalized Moll-Ross relation, derived by Kroemer [9] as:

\[
J_n = -\frac{q \exp\left(\frac{qV_{BE}}{kT}\right)}{\int_0^{W_E} \frac{p(x)}{n_i^2(x)D_{mb}(x)} dx}
\]

Eq. 3

\[
G_B = \frac{\int_0^{W_B} \frac{p(x)}{n_i^2(x)D_{mb}(x)} dx}{\int_0^{W_B} \frac{p(x)}{n_i^2(x)D_{mb}(x)} dx}
\]

Eq. 4

Where \( p(x) \) is the position dependent hole concentration, \( D_{mb}(x) \) is the minority carrier diffusivity and \( n_i(x) \) is the intrinsic carrier concentration. The integral in the denominator (Eq. 4) is known as the base Gummel number (\( G_B \)). The integration limits, \( x = 0 \) and \( x = W_B \), correspond to the edges of the neutral base at the emitter and collector side respectively.

The intrinsic carrier concentration for strained SiGe is given by:

\[
n_{SiGe}^2 = \gamma n_{i0}^2(T) \exp\left(\frac{\Delta E_g(T)}{kT}\right)
\]

Eq. 5

Where \( n_{i0} \) is the intrinsic carrier concentration for Si and a factor \( \gamma \) is introduced to account for the reduction in the effective density of states in compressively strained SiGe [10, 11].

\[
\gamma = \frac{(N_C N_V)_{SiGe}}{(N_C N_V)_{Si}}
\]

Eq. 6

\( N_C \) and \( N_V \) are the density of states in the conduction and valence bands respectively. The factor \( \gamma \) varies between approximately 1 and 0.2, for typical values of the Ge-concentration (0 – 30 %). This means that the exponential factor in Eq. 5, depending on \( \Delta E_g \), will dominate and thus the intrinsic carrier concentration increases with the Ge-concentration. We find that for identical base doping profiles a SiGe HBT will have a smaller \( G_B \) and thus higher \( I_C \) than a Si BJT. This conclusion applies for a fixed \( V_{BE} \) and agrees with the more qualitative discussion above. From Eq. 3 we also find that increasing the base doping reduces \( I_C \). Taking advantage of the higher gain due to the Ge bandgap engineering allows a higher base doping level to be used. This results in lower base sheet resistance. A low base resistance (\( R_B \)) is important to achieve high power gain and low high-frequency noise.
So far we have assumed that we have a constant Ge-concentration inside the neutral base. We now discuss the influence of a bandgap grading, since it will improve $V_A$ and reduce $\tau_B$. The Early voltage $V_A$ is an important figure-of-merit in analog circuit applications and is defined as [12]:

$$\int \frac{p(x)}{n_i^2(x) D_{nb}(x)} \, dx$$

Eq. 7

Where $C_{BC}$ is the base-collector depletion capacitance. To illustrate the improvement in Early voltage we now consider the case of a uniform base doping and a linearly graded Ge-concentration:

$$p(x) = N_B$$

Eq. 8

$$n_{i,\text{SiGe}}^2 = n_{i0}^2 \exp \left( \frac{\Delta E_{g,\text{SiGe}}}{kT} \frac{x}{W_B} \right), \quad 0 \leq x \leq W_B$$

Eq. 9

Where $\Delta E_{g,\text{SiGe}}$ represents the total bandgap reduction at the collector side of the neutral base corresponding to $x = W_B$. Inserting Eq. 9 in Eq. 7 and performing the integration over the neutral base width leads to:

$$V_A = \frac{q N_B W_B}{C_{BC}} \frac{kT}{\Delta E_{g,\text{SiGe}}} \left[ \exp \left( \frac{\Delta E_{g,\text{SiGe}}}{kT} \right) - 1 \right]$$

Eq. 10

It is found that $V_A$ depends exponentially on the bandgap difference in the neutral base, which leads to a considerable enhancement, e.g. approximately a factor of 10 for 8 % Ge-grading.

Finally, we consider the influence on the cut-off frequency ($f_T$) or forward transit time ($\tau_F$). The total transit time in a bipolar transistor is given by the sum of several contributions:

$$\tau_F = \tau_B + \tau_E + \tau_{EB} + \tau_{BC} + \frac{kT}{qI_C} \left( C_{BE} + C_{BC} \right)$$

Eq. 11

Where $\tau_B$ is the base transit time, $\tau_E$ is the emitter transit time, $\tau_{EB}$ is the base-emitter depletion region transit time and $\tau_{BC}$ is the base-collector depletion region transit time. The final term in Eq. 11 accounts for the charging times of the base-emitter depletion capacitance $C_{BE}$ and the base-collector depletion capacitance $C_{BC}$. The cut-off frequency $f_T$ is defined as the frequency where the ac current gain equals unity:
The forward transit time $\tau_F$ is inversely related to $f_T$:

$$f_T = \frac{1}{2\pi \tau_F} \quad \text{Eq. 13}$$

Figure 2 illustrates the transit time distribution as a function of $I_C$ in a typical 50 GHz SiGe HBT structure. The transit times have been obtained from 1D numerical simulations in SCORPIO [13] using measured secondary ion mass spectrometry (SIMS) doping and Ge-profiles as input data. For $I_C$ around 2 mA, corresponding to peak $f_T$, $\tau_B$ is the dominating contribution and accounts for about 50% of the total transit time. We also find that $\tau_{BC}$ and $\tau_E$ have a significant contribution. In Fig. 2 the transit time in the undepleted part of the lowly doped collector (τC) is also shown.

![Transit time distribution vs. collector current in a 50 GHz SiGe HBT structure, obtained from numerical simulations.](image)

The base transit time $\tau_B$ represents the time the injected minority carriers (electrons) need to travel through the quasi-neutral base. The transport of minority carriers occurs mainly by diffusion. If velocity saturation at the base-collector junction is neglected the diffusion limited transit time, for a uniform base doping, is given by:

$$\tau_B = \frac{W_B^2}{2D_{nB}} \quad \text{Eq. 14}$$
In the case of a non-uniform Ge-concentration (or base doping profile) a conduction band gradient will be present, giving rise to an additional driving force. This is sometimes referred to as a quasi-electrical drift field. Kroemer [9] derived the following expression for $\tau_B$:

$$\tau_B = \int_0^z \frac{w_x^2 n_i^2(z)}{p(z)} \left[ \int \frac{p(y)dy}{D_{eb}(y)n_i^2(y)} \right] dz$$

Eq. 15

For simplicity, assume a constant base doping and a linearly graded Ge-profile, according to Eq. 8 and Eq 9. The base transit time can be written as:

$$\tau_B = \frac{W^2}{D_{eb} \Delta E_{g, SiGe}} \left[ 1 - \frac{kT}{\Delta E_{g, SiGe}} \left[ 1 - \exp \left( -\frac{\Delta E_{g, SiGe}}{kT} \right) \right] \right]$$

Eq. 16

If we compare a graded (6 →12 %) and a constant (12 %) Ge-profile we find that the reduction in $\tau_B$, due to the drift field, leads to about 5 % increase in $f_T$, as shown in Fig 3.

![Fig. 3 Simulated cut-off frequency for a constant and a graded Ge-profile.](image)

For a higher maximum Ge-content, e.g. 20 % the Ge-gradient increases. This leads to a larger reduction in base transit time due to the drift field. However a large Ge-gradient in the neutral base might lead to adverse effects such as a bias dependent $\beta$ [14], which is undesirable for many circuit applications.
Finally, we briefly consider the influence of the Ge-profile on $\tau_E$. The emitter transit time $\tau_E$ depends on the minority charge storage, i.e., injected holes, in the emitter region. The hole injection is reduced by the valence band offset between the base and the emitter, which creates a potential barrier. For the very narrow base widths, which are possible to achieve with epitaxy, $\tau_E$ and $\tau_B$ become comparable in magnitude. This means that $\tau_B$ is not limiting the total transit time. It becomes important to reduce $\tau_E$, which is inversely proportional to $\beta$ and to the Ge-concentration. Also, since quite high base doping levels are typically used the Ge-concentration at the base-emitter junction needs to be raised to maintain sufficient $\beta$. This leads to a trade-off for the design of very high speed SiGe HBTs [15].

### 2.2 Device simulation

This section discusses physical device simulation. Here, a finite element mesh is used to represent the semiconductor device. The mesh structure is composed of several regions with differing electrical properties; Semiconductors - Si and Si$_{1-x}$Ge$_x$ alloys, insulators - mainly represented by SiO$_2$ and contacts - formed by Al or silicides. For the different regions in the mesh structure, physical models and parameters for carrier transport, e.g. mobility, intrinsic carrier concentration and lifetimes are specified. To find the carrier concentrations and potential in each of the mesh node points a system of coupled partial differential equations is solved numerically by an iterative method. The system of differential equations is derived from an approximate solution of the so-called Boltzmann transport equation. In the drift-diffusion (DD) approximation three equations are used. Namely the continuity equations for electrons and holes and the Poisson equation for the potential. A more detailed approximation level includes the local carrier temperatures for electrons and holes, the so-called energy-balance (EB) or hydrodynamic formalism. This leads to a total number of five coupled differential equations. Normally voltage boundary conditions are used. In this manner the terminal currents of the device can be obtained for both DC and AC conditions. The AC solution is found by performing a small-signal linearization of the problem, starting from the DC steady state solution. The output of an AC-simulation is the conductance and capacitance between each pair of terminals, e.g. base and emitter. The conductance and capacitance values are used to calculate the admittance, $Y$-parameters of the device (two-port). The $Y$-parameters can be converted to impedance, $Z$-parameters or scattering, $S$-parameters. The $S$-parameters are needed to find important figure-of-merits such as $f_T$ and power gain.

Two-dimensional (2D) device simulations, based on measured SIMS profiles, are sufficient to obtain good agreement between measurement and simulations. On the other hand one-dimensional (1D) simulations are particularly useful for profile optimization, due to the much shorter simulation times. However, 1D simulations tend to overestimate $f_T$ and $\beta$ due to the lack of 2D effects such as the contribution from the peripheral part of the base-emitter capacitance and non-ideal base current components resulting from the high extrinsic base doping. Three-dimensional (3D) device simulations are not widely used, but are needed for some special applications. For non-isothermal problems, where the lattice heat flow equation is solved, 3D simulation should be used in order to get realistic boundary conditions.
For advanced sub-micron devices the EB formalism must be used to account for non-local effects, such as impact ionization and velocity overshoot at the base-collector junction. Using the EB-formalism, with 5 coupled equations, results in at least twice as long simulation times, compared the DD-formalism. The number of mesh points, that can be used in the structure is mainly limited by the simulation time. For SiGe HBTs a very dense mesh with a minimum spacing on the order of 5 Å is needed in regions with large doping or Ge-gradients. Fortunately, automatic mesh generation tools are available to design useful structures, improve convergence and reduce simulation time.

2.3 Physical models for strained SiGe

To obtain reliable simulation results accurate models are needed for different physical properties of the materials in the device structure such as mobility, intrinsic carrier concentration and carrier lifetimes. These properties are functions of the temperature, the impurity doping level and the alloy composition for the Si$_{1-x}$Ge$_x$ case. The electrical characteristics are very sensitive to the Ge-profile and therefore we concentrate on the physical models and parameters for SiGe alloys. In the simulations we used Ge-concentration dependent models for the effective density of states in valence and conduction bands and the bandgap for strained SiGe [16]. Models for the low-field mobility and velocity saturation in SiGe [17, 18] were also used. The physical models were implemented in the commercial finite element device simulator SILVACO/ATLAS [19] using the c-code interface. In the following section we first briefly discuss the mobility models. This is followed by a detailed discussion about the intrinsic carrier concentration in strained SiGe.

2.3.1 Low-field mobility and velocity saturation

The low-field mobility in SiGe is influenced both by the compressive lattice strain and the alloy scattering. The resulting effect, mobility enhancement or degradation, depends on the carrier type, i.e., electrons or holes as well as the transport direction – parallel or perpendicular to the growth direction of the strained SiGe layer. For bipolar devices different mobility models for minority and majority respectively are needed. The most widely used model for SiGe HBT simulation [13, 20] is the Klaasen unified mobility model, which was originally developed for Si-devices. In this study a modified version of the Caughey-Thomas mobility model [21] was used to account for the difference in minority and majority mobility. In SiGe HBTs mobility enhancement is observed for hole transport perpendicular to the growth direction. This will mainly influence the base resistance. However the mobility increase is relatively small for typical Ge-concentrations, used in SiGe HBTs (< 20 %). In SiGe HBTs the electron transport occurs parallel to the growth direction of the strained SiGe base layer. This leads to somewhat reduced electron low-field mobility, compared to a Si-device. Furthermore the electron saturation velocity decreases with increasing Ge-concentration.

2.3.2 Intrinsic carrier concentration in strained SiGe

In the following section we discuss two important effects that affect the intrinsic carrier concentration $n_i^{SiGe}$ in strained SiGe. The intrinsic carrier concentration is
needed to calculate the collector current (Eq. 3). In SiGe $n^\text{SiGe}$ is influenced by the smaller effective energy bandgap as well as a strain-induced reduction in effective density of states (DOS). The intrinsic carrier concentration is given by the following relation:

$$n^\text{SiGe}_i = \sqrt{N^\text{SiGe}_C N^\text{SiGe}_V} \cdot \exp\left(-\frac{E_{g,\text{eff}}}{2kT}\right)$$

Eq. 17

In Eq. 17, the effective bandgap $E_{g,\text{eff}}$ is used, which accounts both for the bandgap reduction in strained SiGe and apparent bandgap narrowing due to heavy doping effects:

$$E_{g,\text{eff}} = E_{g,\text{Si}} - \Delta E_{g,\text{SiGe}} - \Delta E_{g,\text{dop}}$$

Eq. 18

The following interpolation formula for $\Delta E_{g,\text{SiGe}}$ was implemented in the simulator [16]:

$$\Delta E_{g,\text{SiGe}} = 0.96x - 0.2x^2$$

Eq. 19

This relation is slightly different from Eq. 1. The reason for this is that a consistent set of physical models should be used in the simulations. The effective bandgap value in Eq. 19 should be used together with the concentration dependent values for $N^\text{SiGe}_C(x)$ and $N^\text{SiGe}_V(x)$, which will be given below. The apparent bandgap narrowing $\Delta E_{g,\text{dop}}$ has been found to be slightly smaller in SiGe, compared to Si [11, 20]. However, we used the Klaasen bandgap narrowing expression for Si [22], which agrees relatively well with the SiGe model found in Ref. [11].

We also need interpolation formulas for the effective DOS for varying Ge concentration x. The DOS is calculated from the carrier density-of-states effective masses in the conduction and valence band respectively. The carrier masses can be found from Monte-Carlo simulation using analytical or full-band models of strained SiGe. For the DOS in the conduction band the following formula was used:

$$N^\text{SiGe}_C(x) = 2 \left(\frac{2\pi m_0 m_{dc,\text{SiGe}}(x) kT}{\hbar^2}\right)^{3/2} \frac{2 \exp\left(-0.63x\right)}{kT} + 4$$

Eq. 20

Where $m_0$ is the rest mass of a free electron. The first factor accounts for the reduction in the electron effective mass and the second factor accounts for the strain splitting of the conduction bands. The density-of-states effective mass is given by:

$$m^\text{SiGe}_{dc}(x) = \left(m^\text{SiGe}_{n,\perp} m^\text{SiGe}_{n,k} m^\text{SiGe}_{n,\parallel}\right)^{1/3}$$

Eq. 21
This relation uses the average of the effective masses in the directions perpendicular (⊥) and parallel (∥) to the growth direction of the strained SiGe layer. The effective electron masses are given by [23]:

\[
m_{e_n,⊥}^{\text{SiGe}} = 0.1905 \frac{1.1698}{E_G^{\text{SiGe}}(T)} - 0.004x \tag{22}
\]

\[
m_{e_n,∥}^{\text{SiGe}} = 0.918 - 0.065x \tag{23}
\]

The effective DOS in the valence band is given by:

\[
N_{v}^{\text{SiGe}}(x) = 2 \left( \frac{2\pi m_v kT}{\hbar^2} \right)^{3/2} \left( \frac{m_{d_v}^{\text{SiGe}}(x)m_{d_v}(T)}{m_{d_v}^{\text{SiGe}}(0)} \right)^{3/2} \tag{24}
\]

The derivation of an interpolation formula for the DOS in the valence band is more complex, since the effective hole masses in several different valence bands must be used. The valence band will be split in sub-bands for heavy holes (hh), light holes (lh), and a split-off (so) band. The density-of-states effective hole mass is given by:

\[
m_{d_v}^{\text{SiGe}}(x) = \left[ m_{hh}^{3/2} + m_{lh}^{3/2} \exp\left(-\frac{\delta_{so}}{kT}\right) + m_{so}^{3/2} \exp\left(-\frac{\delta_{so}}{kT}\right) \right]^{2/3} \tag{25}
\]

Where

\[
\delta_{so} = E_{so} - E_{hh} = 0.044 + 0.530x \tag{26}
\]

and

\[
\delta_{ev} = E_{ev} - E_{hh} = 0.28x - 0.536x^2 + 0.71x^3 \tag{27}
\]

The effective masses in the different valence bands are given by [16]:

\[
m_{hh} = m_0 \left( 0.689 - 2.251x + 4.954x^2 - 5.580x^3 \right) \tag{28}
\]

\[
m_{lh} = m_0 \left( 0.269 - 0.195x - 0.317x^2 - 0.292x^3 \right) \tag{29}
\]
\[ m_{so} = m_o \left( 0.126 + 0.051x + 0.876x^2 - 2.589x^3 \right) \]  

**Eq. 30**

For the temperature dependence of the density-of-states hole mass \( m_{dv}(T) \) Si-values were used [24]:

The effective DOS for the valence band and conduction band in strained SiGe are lower than for Si. In Fig. 4, the effective DOS calculated from Eq. 20 and Eq. 24 are shown, normalized to the Si-values.

![Effective DOS](image)

**Fig. 4** Valence and conduction band effective density of states for Si\(_{1-x}\)Ge\(_x\) normalized to the values for Si.

### 2.4 Device simulation calibration

Typically, simulations can be calibrated to measurement within 10 – 15 percent accuracy, provided that the Ge- and doping profiles are known see e.g. Ref. [13]. The doping profiles can be obtained from SIMS performed with high depth resolution. The concentration can be obtained with high accuracy. It is, however, difficult to resolve large concentration gradients. The SIMS measurements give the physical concentration of dopant atoms. In many cases the electrically active concentration will be significantly lower. In the simulations it is reasonable to assume that all dopants are activated in the epitaxially grown base as well as in the relatively lowly doped (~\(10^{17}\) cm\(^{-3}\)) collector. In the polysilicon emitter, physical doping concentrations in the order of \(5 \times 10^{20}\) cm\(^{-3}\) are used. Depending on the choice of emitter fabrication procedure the activation level will vary. For an *in situ* doped emitter, activation levels of 30 – 40 % can be achieved [25]. The activation level in ion-implanted emitters is typically lower. If experimental SIMS-profiles are not available process simulations of ion-implantation can be used. Some parameters
such as the active doping concentration levels in emitter and base polysilicon could be extracted from Hall-measurement. The largest uncertainty in the simulation input data is related to the Ge-profile. The rising and falling edges of the Ge-profile are typically very abrupt, making SIMS analysis difficult. The average Ge-concentration or total amount of Ge-atoms can also be obtained using other physical analysis tools such as X-ray diffraction (XRD) and Rutherford backscattering (RBS).

For the calibration of simulated DC-characteristics to measurements we first discuss the collector current $I_C$. The effective bandgap reduction will lead to increased $I_C$ for a fixed $V_{BE}$ and the reduction of the effective DOS leads to a reduction in $I_C$. For useful doping and Ge-concentrations in the base the former effect (collector current increase) will prevail over the latter one. A small uncertainty is introduced by the difficulty to determine the effective emitter area for very small devices. In Fig. 5 we compare measured and simulated Gummel characteristics, $I_B$ and $I_C$ vs. $V_{BE}$, for a SiGe HBT with $5 \times 1 \, \mu m^2$ emitter area. The agreement between measured and simulated $I_C$ is good for the whole $V_{BE}$ bias range. Calibration of the base current $I_B$ is more complicated since non-ideal effects such as Shockley-Read-Hall (SRH) recombination might have a significant influence for low $V_{BE}$. For the investigated devices the non-ideal $I_B$-component was relatively large and showed a pronounced bias dependence. The non-ideal $I_B$-component originates from the high dose extrinsic base implantation, which was placed relatively close to the emitter window edge. This causes a strong band-bending in the depletion region at the p$^+$-n$^+$ base-emitter junction. This modifies the capture cross-sections and hence the recombination rate and was modeled by introducing a field-dependent term in the SRH recombination model [26]. Good agreement was observed for the ideal part of $I_B$. The ideal $I_B$-component can be fine-tuned by changing the surface recombination velocity at the emitter contact or by adjusting the Auger coefficients in the highly doped polysilicon emitter.

For calibration of the high-frequency characteristics we have mainly considered $f_T$, since other important parameters such as maximum frequency of oscillation ($f_{MAX}$) are more affected by the extrinsic device parasitics $R_B$ and $C_{BC}$. Both $R_B$ and $C_{BC}$ are slightly underestimated, by using a simplified simulated structure. The choice of mobility model in the polycrystalline extrinsic base was also found to have a large influence on the simulated $f_{MAX}$. On the other hand reasonably good agreement was obtained between the simulated and measured $f_T$-values. In Fig. 6 we observe a larger slope for the simulated $f_T$ at low $I_C$ due to reduced 2D effects, which also leads to higher peak $f_T$. The influence of 2D effects was clearly observed by comparing measured $f_T$-values for devices with different emitter area-to-perimeter ratios. A small difference is also observed in the high current $f_T$ roll-off, which is very sensitive to the collector doping level as well as the Ge-profile at the base-collector junction.
Fig. 5 Comparison of measured and simulated Gummel-characteristics for a $5 \times 1 \ \mu\text{m}^2$ SiGe HBT, $V_{CB} = 0 \ \text{V}$.

Fig. 6 Comparison of measured and simulated $f_T$ for a $5 \times 1 \ \mu\text{m}^2$ SiGe HBT, $V_{CB} = 1.0 \ \text{V}$. 
3 SiGe epitaxial base integration

In this chapter we discuss process integration issues for SiGe HBTs with a non-selectively grown epitaxial base. In the first section we briefly describe the process flow including the device isolation, base epitaxy and emitter formation steps. In section 2 we discuss extrinsic base design and transient enhanced diffusion (TED) effects caused by the extrinsic base implantation. Optimization of the base design is further discussed in paper V.

3.1 Process flow

The base region in a SiGe HBT is formed by epitaxial growth of an in situ boron doped strained SiGe layer. The SiGe layer is grown either by reduced pressure chemical vapor deposition (RPCVD) or by ultra high vacuum chemical vapor deposition (UHV/CVD). The epitaxial growth can occur either selectively or non-selectively. Non-selective SiGe epitaxy is used in several industrial SiGe HBT processes today [3, 27, 28]. In comparison with selective epitaxy [29, 30], non-selective epitaxy offers more straightforward process integration, better control of the SiGe layer thickness uniformity and reduced local loading effects. In this thesis we have investigated SiGe HBTs fabricated using two different processes. In both cases non-selective growth of the epitaxial base was used. Following the base epitaxy the polysilicon emitter can be formed in two ways, either using a single-poly or a double-poly approach. In the single-poly approach the emitter-base structure is formed in quasi-self-aligned manner (QSA). In the double-poly approach the emitter window is self-aligned to an additional deposited polysilicon extrinsic base layer, hence the name double-poly. The choice of process is a trade-off between performance and manufacturability. The fully self-aligned process minimizes the parasitics, such as \( R_B \) and \( C_{BC} \), and gives very good RF-performance. To obtain similar performance, using the QSA approach, careful optimization of the base region layout and design rules is needed. This is discussed in more detail in Paper V. The main advantage with the QSA is that it is relatively easy to integrate into a conventional CMOS process flow, by adding four lithographic mask levels [31, 32].

In a standard process flow the growth of the epitaxial base follows directly after the device isolation step, using either local oxidation of silicon (LOCOS) or deep and shallow trench isolation. The device isolation defines the active area, where the intrinsic device is formed. Inside the active area epitaxial SiGe growth occurs, whereas poly-crystalline SiGe is simultaneously grown on the isolation oxide (LOCOS or oxide-filled trenches). For the selective case no SiGe growth occurs on exposed oxide or nitride surfaces. For device applications a composite Si/SiGe/Si stack is typically used. The stack consists of a Si-buffer layer, a strained, boron doped SiGe-layer with intrinsic (undoped) spacers and a Si-cap layer. The buffer layer is needed to improve the quality of the SiGe layer and the Si-cap is important to avoid strain relaxation. The thickness of the Si-cap layer also controls the positioning of the emitter-base junction. A schematic cross-section of the intrinsic
transistor region after SiGe base epitaxy and poly-Si emitter formation is shown in Fig. 7.

Fig. 7 Schematic illustration of single-poly SiGe HBT with quasi-self-aligned (QSA) emitter-base structure.

The poly-crystalline SiGe on the isolation oxide forms the extrinsic base electrode and is implanted with a high dose of boron to obtain low sheet resistance. This is important to obtain high power gain as well as low high-frequency noise. A small part of the epitaxial SiGe base region is also implanted with boron, shown as the p⁺-implantation pockets.

The region between the edge of the extrinsic base implantation and the intrinsic device underneath the emitter window is often referred to as the base-link region. To minimize the base link resistance, the extrinsic base implantation should be self-aligned to the poly-Si emitter using spacers [33]. Misalignment of the implantation with respect to the emitter window might result in a non-ideal base current and reduced \( f_T \). This discussion only applies to a single-poly device structure. For a double-poly structure, the extrinsic base region is formed by a deposited polysilicon layer, which is self-aligned to the emitter window using so-called inside spacers. A double poly-approach was used in paper IV.

3.2 Transient enhanced diffusion of boron

Ion-implantation in the epitaxial SiGe base region creates a large number of highly mobile point defects i.e. Si self-interstitials and defect clusters, which contribute to transient enhanced diffusion of boron [34]. On the other hand interstitials, created in the poly-crystalline SiGe region, are effectively captured at the grain boundaries and hence do not contribute significantly to TED. The TED effect causes a significant broadening of the intrinsic vertical base profile. This might be detrimental for device performance, since out-diffusion of boron past the Si/SiGe heterojunctions results in formation of conduction band barriers [35], reducing \( I_C \) and \( f_T \). To avoid out-
diffusion of boron past the base-collector heterojunction an undoped SiGe spacer layer is always used between the highly doped SiGe base and the collector.

In Paper V devices with different design of the extrinsic base region were investigated to quantify the influence of TED. The extrinsic base design also affects the low-frequency noise properties, which will be further discussed in Chapter 7. Out-diffusion of the boron profile past the base-collector heterojunction will affect both DC- and HF-characteristics of the devices. In this study several extrinsic base design parameters were changed to find an optimized design with high $f_T$, $V_A$ and ideal base current $I_B$. By changing the position of the extrinsic base implantation relative to the interface between the poly-crystalline and epitaxial part of the extrinsic base region devices with different amounts of implantation-created interstitials could be compared. The size of the LOCOS active area was varied, while the distance from the implantation edge to the emitter window was kept fixed. In this way the influence of extrinsic base encroachment could be minimized. This is important since placing the implantation too close to the emitter window edge is also known to degrade $f_T$ and base current ideality.

The devices, where the boron implantation was mainly inside the poly-crystalline part of extrinsic base, exhibited the highest $f_T$. However these devices suffered from a poor base current ideality, which was attributed to recombination at the interface between the poly-crystalline and epitaxial base regions. A significant reduction in $f_T$ was observed for devices, which had received a boron implantation in the epitaxial part of the extrinsic base. Both $V_A$ and $I_C$ were also reduced, compared to the situation above. On the other hand the base current ideality was clearly improved. Therefore, the influence of TED should be minimized, by choosing an appropriate intrinsic SiGe spacer thickness, which accommodates for some boron out-diffusion.

In this context we should also mention that TED effects can be significantly reduced by introducing a relatively high concentration of carbon (<$10^{20}$ cm$^{-3}$ 0.1 - 0.2 atomic percent) in the SiGe lattice [36, 37]. Boron diffuses mainly by an interstitial mechanism, i.e., coupled diffusion of boron and interstitials. Since carbon diffusion is also mediated by interstitials, boron and carbon diffusion will be competing reactions. Therefore the TED of boron, due to the super-saturation of interstitials will be reduced in a SiGe:C epitaxial layer [38]. As a result devices with narrower base widths and higher base doping levels can be fabricated, leading to a significant performance increase [4, 27].
4 Collector design

In a high-speed, low-voltage SiGe HBT an optimized collector doping profile is needed to achieve high cut-off frequency, sufficiently high breakdown voltage and to minimize parasitic capacitances. The RF-distortion properties will also be affected; this is further discussed in Chapter 6. In the first section of this chapter we discuss impact ionization and breakdown voltage. We also introduce a non-local description of impact ionization, which is needed for accurate simulation of low-voltage devices with high collector doping levels. Section 2 discusses the influence of collector design on high-frequency properties such as $f_T$. In the next section the commonly used selectively implanted pedestal collector (SIC) is discussed. In the final section of this chapter we introduce a launcher collector profile where a narrow and highly doped peak is formed close to the base-collector heterojunction. A novel concept to form an ultra shallow doping profile using antimony (Sb) low-energy implantation is investigated.

4.1 Breakdown voltage

For a reverse biased pn-junction avalanche breakdown limits the maximum voltage, which can be used for stable operation of the device. In the case of an npn-bipolar transistor the electrons at the base-collector junction might gain sufficient energy from the high electric field to cause an impact ionization event where an electron-hole pair is generated. The probability for an impact ionization event to occur is given by the ionization rate $\alpha$. The generated electrons and holes might in turn give rise to additional impact ionization events. The number of generated electron-hole pairs is given by the multiplication factor $M - 1$. The breakdown voltage is defined as the voltage where the multiplication factor $M$ goes to infinity. In bipolar devices the base-collector breakdown voltage is denoted $BV_{CBO}$ and the collector-emitter breakdown voltage is called $BV_{CEO}$. Because of a feedback of generated electron-hole pairs $BV_{CEO}$ is smaller than $BV_{CBO}$.

For a situation where the electrons in the base-collector (BC) depletion region can be assumed to be in local equilibrium with the lattice the impact ionization rate is an exponential function of the local electric field. However, if the electric field varies rapidly with position in the BC depletion region, the electrons will not be in equilibrium with the lattice. The so-called carrier temperatures or energies will not follow the electric field distribution, due to the finite energy relaxation times [39]. Hence, for devices with high base- and collector-doping levels a non-local impact ionization description should be used. This means that the impact ionization rates are calculated from the carrier temperatures, rather than from the local electric field. In Fig. 8 the electric field and impact ionization rate are shown for the SiGe HBT structure investigated in Paper IV. The maximum impact ionization rate occurs deep inside the depleted collector and essentially follows the electron local temperature profile (not shown here). On the other hand the electric field has a peak close to the base-collector SiGe/Si heterojunction. The impact ionization rate close to the
electric field peak will be severely overestimated by using a local-field description. This leads to a significant reduction in the calculated $BV_{CEO}$.

![Graph of electric field and impact ionization rate vs. depth](image1)

**Fig. 8** Simulated electric field and impact ionization rate vs. depth for the SiGe HBT structure in Paper IV. Dashed line indicates the position of the Ge-profile, maximum concentration 12%.

In Fig. 9 the breakdown voltage using a local-field description and non-local carrier temperature description are compared for the same SiGe HBT structure as above. A significant reduction in breakdown voltage is observed for the local-field description.

![Graph of collector current vs. collector-emitter voltage](image2)

**Fig. 9** Simulated breakdown voltage for SiGe HBT structure in Paper IV, using local and non-local impact ionization description.
4.2 Cut-off frequency and Kirk effect

The maximum available or peak $f_T$ is limited by the onset of the Kirk effect or base push-out. The onset of the Kirk effect occurs when the collector current density becomes equal to the doping level in the collector. This means that the effective doping at the base-collector junction is reduced and the electrical field vanishes. As a result the effective base becomes wider and $\tau_B$ will start to increase. Using a high collector doping level suppresses the onset of the Kirk effect. Consequently, higher peak $f_T$ can be obtained, due to the reduced base widening at high current densities. Increasing the collector doping level has two major disadvantages on the device performance. Firstly, $C_{BC}$ will be increased, leading to reduced power gain or $f_{MAX}$. Secondly, the maximum electric field at the base-collector junction will be higher, leading to more impact ionization and thus lower $BV_{CEO}$. The reduced $BV_{CEO}$ in a device with high collector doping level and high $f_T$ leads to a trade-off for the collector profile design.

The cut-off frequency $f_T$ is determined by a sum of transit times in different regions of the transistor. The base-collector depletion region transit time $\tau_{BC}$ can be reduced by using a higher collector doping. In SiGe HBTs, $\tau_{BC}$ is not limiting whereas the influence is larger in III-V HBTs [40]. By considering analytical expressions for $\tau_{BC}$ and $BV_{CEO}$ a theoretical figure-of-merit for bipolar transistors has been derived [41]. The so-called Johnson limit is defined as $f_T \times BV_{CEO}$ and has a value of 200 GHz-V. It is clear from recent experimental results that this simplified relation does not accurately predict the performance of state-of-the-art SiGe HBTs, where values as high as 340 GHz-V have been reported [3, 42].

4.3 Selectively implanted collector (SIC)

Using a SIC provides a high collector doping self-aligned to the emitter window. In this way $C_{BC}$ is minimized, since there is essentially no overlap between the highly doped extrinsic base and collector regions. Using medium to high-energy ion implantation (100 – 400 keV) arbitrary collector profiles can be formed. The peak of the implantation profile can be placed close to the base-collector junction or deeper inside the epitaxial collector, close to the buried layer. In current high-speed, low voltage processes the retrograded profile with a concentration increasing towards the buried layer is preferred. Such a design reduces the maximum electrical field at the base-collector junction compared to a uniformly doped collector, while maintaining high $f_T$ [43]. In SiGe HBTs a first implantation can be performed before epitaxial base formation and a second implantation can be performed self-aligned to the emitter window opening [44]. Leaving out or masking the second implantation step enables fabrication of devices with higher breakdown voltage on the same wafer [29].
4.4 Launcher profiles

Narrow doping peaks with high concentration, so-called launchers, have been demonstrated in both compound semiconductor [45] and SiGe HBT technologies [46]. The highly doped layer close to the base-collector junction suppresses the onset of the Kirk effect and compensates for boron out-diffusion from the base. Therefore the effective base width can be made shorter. In single heterojunction HBTs this allows operation at very high current densities, whereas the launcher concept is less efficient for double heterojunction HBTs with a valence band offset. A typical launcher thickness is about 50 - 100 nm. Since the layer is thin electrons traversing this region cannot gain enough energy from the electric field to cause an impact ionization event [47]. Therefore it is possible to increase the collector doping level locally without affecting $BV_{CEO}$. For correct design of the launcher layer thickness and doping a non-local description of impact ionization should be used. Introducing a highly doped layer will increase the $CBC$ slightly, however if the layer is totally depleted by the application of a reverse bias the change will be negligible. Important figure-of-merits such as $f_{MAX}$ will not be degraded.

In Paper IV a systematic study of launcher like collector doping profiles, obtained by ultra shallow ion-implantation, was performed. The minimum width of the implantation profile is limited by the implantation energy and thermal diffusion. Today low-energy implantations (2-5 keV) are routinely available and used for shallow junction formation. In this study antimony was chosen for the narrow launcher profile since the diffusivity of antimony is lower compared to arsenic and phosphorous. Therefore the implantation profile is relatively unchanged by diffusion during subsequent high temperature processing steps.

SiGe HBTs were fabricated in a fully self-aligned double-poly process and dedicated to study the influence of the collector doping profile on breakdown and high-frequency performance. The reference devices received only a conventional high energy phosphorous SIC implantation and the resulting doping profile was relatively uniform throughout the depth of the epitaxial collector. Typical doping and Ge-profiles (reproduced from Paper IV) are shown in Fig. 10.

Process simulations with varying energy and dose for the Sb-implantation indicated that narrow launcher like Sb-peaks could be formed. From the SIMS-profiles of the fabricated device it can be observed that some Sb-segregation has occurred during the growth of the epitaxial base. Similar segregation effects are observed during low temperature RPCVD n-type epitaxy, see Paper XII and Ref. [48]. Therefore, highly-doped non-uniform collector profiles are difficult to realize by CVD. The segregation effect needs to be minimized in order to avoid punch-through electrical characteristics due to a too high donor concentration in the base.
Fig. 10 Doping and Ge-profiles from a high-speed (60 GHz $f_T$) low voltage SiGe HBT from Paper IV. Antimony implantation performed prior to SiGe base epitaxy and phosphorous SIC implantation performed after base epitaxy.
5 High-frequency noise

The possibility to use a high base doping in a SiGe HBT, without degrading the current gain, has led to significantly improved high-frequency noise properties [49]. The high base doping reduces the base resistance, which results in lower thermal noise. SiGe HBTs also have higher low-current $f_T$ compared to Si BJTs, which reduces the influence of the collector shot noise [50]. For the design of SiGe HBT low-noise circuits accurate characterization and extraction of the noise parameters is important. However, high-frequency noise characterization is relatively complex as well as time-consuming. A dedicated set-up is needed, including a thermal noise source, noise figure meter and impedance tuners. We have investigated two admittance or Y-parameter based methods. The Y-parameters are obtained directly from S-parameter measurement using a vector network analyzer (VNA), which is routinely used for RF-characterization. Both investigated Y-parameter based methods show good qualitative agreement with the conventional approach. Furthermore, these methods are well suited for noise optimization, using a device simulator. In this chapter we first give an introduction to some basic high-frequency noise concepts and then investigate the Y-parameter based extraction methods in some detail. Optimization of the Ge-profile for reduced high-frequency noise is also investigated by device simulation. Finally, we briefly discuss noise parameter de-embedding.

5.1 Basic definitions

A commonly used figure of merit is the noise figure ($F$). The noise figure can be directly related to the noise temperature ($T_e$) using the following relation:

$$F = 1 + \frac{T_e}{290}$$  \hspace{1cm} \text{Eq. 31}

The reference temperature of the source is usually assumed to be 290 K. The noise figure is often given on a logarithmic (dB) scale:

$$NF = 10 \log_{10}(F)$$  \hspace{1cm} \text{Eq. 32}

The noise figure $F$ is a relative measure, defined as signal to noise ratio (SNR) at the input divided by the SNR at the output. This relation can be used at device or circuit level. In other words the total output noise power is normalized to the total input noise power, which is given by the thermal noise of the source (sometimes referred to as the generator noise). Under small signal conditions $F$ is a parabolic function of the source admittance ($Y_S$), and will be given by the following equation:
The minimum achievable noise figure is called $F_{\text{MIN}}$ ($NF_{\text{MIN}}$). $R_n$ is the noise resistance and $Y_{\text{OPT}}$ is the optimum source admittance, where:

$$Y_{\text{OPT}} = G_{\text{OPT}} + jB_{\text{OPT}}$$ \hspace{1cm} \text{Eq. 34}

In many situations it is convenient to use the optimum source reflection factor $\Gamma_{\text{OPT}}$, which is related to $Y_{\text{OPT}}$ by:

$$\Gamma_{\text{OPT}} = \frac{1-Y_{\text{OPT}}Z_0}{1+Y_{\text{OPT}}Z_0}, \quad Z_0 = 50\Omega$$ \hspace{1cm} \text{Eq. 35}

Thus $F$, for a certain source admittance $Y_S$, is determined by a total of four noise parameters – $F_{\text{MIN}}$, $R_n$, $G_{\text{OPT}}$ and $B_{\text{OPT}}$. To be able to extract the noise parameters one needs to measure the noise figure corresponding to four different values of the source admittance. In practice, approximately 10 different source admittance values are used. This gives reasonable accuracy, for fitting of a parabolic noise figure surface to the measurement data. The source admittance tuning procedure is relatively time-consuming, thus limiting the number of source admittance values that can be used. In the ATN 5B set-up, used in this study, the source admittance is changed using electronic PIN-diode tuners. The tuner provides more than 300 different fixed source admittance states. However, if the input reflection factor of the device is large ($\Gamma_{\text{IN}} > 0.8$), fewer tuning states close to the optimum source admittance will be available. This will make the extraction of the noise parameters less accurate. The uncertainty in the noise parameters will be observed as a large scatter in $Y_{\text{OPT}}$ and $NF_{\text{MIN}}$ for different bias or frequency as well as a high $R_n$. The uncertainty in the noise parameter extraction will be exemplified by measurement in the following section and has also been discussed in paper VI.

Another important quantity, which is closely related to the noise parameters, is the associated gain ($G_A$). The definition of $G_A$ is the maximum available gain for the optimum source admittance. Inserting $\Gamma_{\text{OPT}}$ into the expression for maximum available gain gives:

$$G_A = \frac{\left(1 - \left|\Gamma_{\text{OPT}}\right|^2\right)\left|S_{21}\right|^2}{\left|1 - S_{11}\Gamma_{\text{OPT}}\right|^2\left(1 - \left|S_{22}\right|^2 + \frac{S_{12}S_{21}\left|\Gamma_{\text{OPT}}\right|^2}{1 - S_{11}\Gamma_{\text{OPT}}}\right)}$$ \hspace{1cm} \text{Eq. 36}

It should be mentioned here that the optimum noise matching does not necessarily correspond to the highest gain. This leads to a trade-off between low noise and high
gain, which can be avoided by introducing an inductive emitter degeneration in a common-emitter LNA stage [51].

### 5.2 Noise parameter extraction

In this section we discuss noise parameter extraction from either measured or simulated Y-parameters of the SiGe HBT. This methodology has a number of advantages, compared to conventional noise figure measurements. Firstly the measurements are easier and no dedicated set-up is needed. This is exemplified in Paper VI. Secondly using 2D numerical device simulations allows the Ge-profile to be optimized for low noise.

The total noise from a SiGe HBT can be represented by the current (and voltage) noise sources as shown in Fig. 11. This approach is used in the so-called thermodynamical noise model [52, 53].

**Fig. 11** Schematic representation of the noise sources in the thermodynamical noise model for a bipolar transistor.

The noise sources at the input (base) side are:

\[ S_i = 2qI_B \] (shot noise)  \hspace{1cm} Eq. 37

\[ S_i = 4kT \text{Re}\{Y_{11}\} \] (thermal noise) \hspace{1cm} Eq. 38

The noise sources at the output (collector) side are given by:

\[ S_i = 2qI_C \] (shot noise) \hspace{1cm} Eq. 39

\[ S_i = 4kT_C \text{Re}\{Y_{22}\} \] (thermal noise) \hspace{1cm} Eq. 40

Where \( T_C \) is the average carrier temperature in the collector region. Here we assume \( T_C = T \). In practice the collector thermal noise is only important for very high
operating frequencies, close to the \( f_T \) of the device. This is further exemplified in Paper VI.

In Fig. 12 a similar schematic representation is shown of the so-called SPICE noise model [51]. This model includes the parasitic resistances \( r_B \) and \( r_E \). The thermal noise sources corresponding to \( r_B \) and \( r_E \) are represented by voltage noise sources. It has been verified that the thermodynamical and SPICE models yield similar results for the noise figure [20]. However, due to the different configuration of the input thermal noise source a difference in \( R_n \) and \( Y_{OPT} \) will be observed.

The noise current values can be calculated from the \( Y \)-parameters, which are frequency, bias and temperature dependent. \( S \)-parameters are measured, using a conventional network analyzer and converted to \( Y \)-parameters. This is a relatively straightforward procedure and the time-consuming source matching is avoided since \( S \)-parameters are measured in a 50 \( \Omega \) impedance environment. The expression for \( F \) as a function of the \( Y \)-parameters at a certain bias (\( I_B \) and \( I_C \)) can be calculated using the thermodynamical approach [20, 52].

\[
F = 1 + \frac{2q|I_B| + 4kT \text{Re}\{Y_{11}\} + (2q|I_C| + 4kT \text{Re}\{Y_{22}\})} {4kTG_S} \left( \frac{\text{Re}\{Y_{11}\} + G_S}{Y_{21}} \right)^2
\]

\text{Eq. 41}

Where \( G_S \) is the real part of the source admittance \( Y_S \). It is seen that the noise figure \( F \) is given by the sum of the shot noise from the input circuit and thermal noise from the input circuit and the input referred thermal and shot noise from the output circuit. The optimum source admittance \( Y_{OPT} \), corresponding to \( F_{MIN} \), is given by [20]:

Fig. 12 Schematic representation of the noise sources in the SPICE noise model for a bipolar transistor.
Here it was assumed that the influence of collector thermal noise could be neglected. The noise resistance $R_n$ is found from:

$$ R_n = \frac{I_C}{2kT q} \left| Y_{21} \right|^2 $$

Eq. 44

The noise parameter expressions in the SPICE model are similar but also include the real part of the input impedance, which is given by the sum of base and emitter resistances $r_{BE} = r_b + r_e$. The extraction procedures for $r_{BE}$ use either S-parameters in the lower frequency range < 1 GHz [51] or the semi-circle method where $r_{BE}$ is found from the input impedance [20]. The minimum noise figure is given by:

$$ F_{min} = 1 + \frac{qI_C}{kT \left| Y_{21} \right|^2} \left( Re\{Y_{11}\} + A \right) $$

Eq. 45

$$ A = \sqrt{\left[ 1 + \frac{2kT \left| Y_{21} \right|^2 r_{BE}}{qI_C} \right] \left[ \left| Y_{11} \right|^2 + \frac{I_B \left| Y_{21} \right|^2}{I_C} \right] - \left( Im\{Y_{11}\} \right)^2} $$

Eq. 46

The optimum source admittance $Y_{OPT}$ and the noise resistance $R_n$ in the SPICE model are found from the following expressions:

$$ G_{OPT} = \sqrt{\frac{I_B \left| Y_{21} \right|^2 + I_C \left| Y_{11} \right|^2}{\frac{2kT}{q} \left| Y_{21} \right|^2 r_{BE} + I_C} - \left( \frac{I_C \left( Im\{Y_{11}\} \right)}{\frac{2kT}{q} \left| Y_{21} \right|^2 r_{BE} + I_C} \right)^2} $$

Eq. 47

$$ B_{OPT} = \frac{-Im\{Y_{11}\}}{\frac{2kT}{q} \left| Y_{21} \right|^2 r_{BE} / I_C + 1} $$

Eq. 48

$$ R_n = \frac{I_C}{\frac{2kT}{q} \left| Y_{21} \right|^2} + r_{BE} $$

Eq. 49
In the following section we illustrate the extraction of noise parameters for the SiGe HBT process investigated in Paper VI. We show results using the two different Y-parameter based extraction methods, described above. The extraction methods are compared to conventional noise figure measurement using a noise figure meter and an impedance tuner at the input. We have chosen a collector bias point close to the minimum noise bias at $I_C$ of approximately 0.4 mA. The transistor size (emitter area) was $5 \times 1.6 \ \mu m^2$. For optimum source impedance match transistors with larger emitter lengths are commonly used but such devices were not available. In Fig. 13 we compare $NF_{MIN}$ and $G_A$ for the three extraction methods. Measured values refer to the conventional method where the noise figure is obtained directly, whereas $NF_{MIN}$ for the thermodynamical and SPICE models respectively was calculated from the measured S-parameters of the SiGe HBT. Relatively good qualitative agreement is found between the methods. The Y-parameter based models give slightly lower value for $NF_{MIN}$. It is reasonable to assume that the main uncertainty in the SPICE model originates from the difficulty to determine $r_{BE}$ from the input impedance. At higher frequency (> 7 GHz) the associated gain is very low. In this frequency range the conventional approach and the thermodynamical model show better agreement. This is mainly due to the presence of an additional collector thermal noise source, which was not included in the SPICE model.

The data from the conventional measurements show significantly more scatter, compared to the Y-parameter based methods. This is explained by the uncertainty in the source admittance tuning procedure, which can be avoided using the SPICE or thermodynamical extraction methods. The investigated device exhibits a relatively large input reflection factor. The electronic tuner cannot provide enough admittance states with large reflection factors to allow an accurate determination of $\Gamma_{OPT}$. In Fig. 14 the magnitude of $\Gamma_{OPT}$, for the different extraction methods is shown. The
magnitude of $\Gamma_{OPT}$ is similar for the SPICE method and the conventional approach. On the other hand, for this bias it is found that the thermodynamical method gives a smaller magnitude of $\Gamma_{OPT}$, which is due to the difference in the input thermal noise source configuration. The angle of $\Gamma_{OPT}$ is shown in Fig. 15. The SPICE method is found to give slightly smaller values, compared to the conventional method. It should be noted that the extraction of $r_{BE}$ also influences $\Gamma_{OPT}$, which could explain some of the difference, especially at lower frequency. The thermodynamical method gives a larger angle, which is also related to the input thermal noise source configuration. In conclusion we find that even though both Y-parameter based methods can be used to extract $NF_{MIN}$ the thermodynamical method is not well suited to find $\Gamma_{OPT}$. This limits the use of the method for circuit design but the method is still useful for noise optimization.

![Fig. 14](image1.png)  
**Fig. 14** Magnitude of the optimum source reflection factor from conventional measurements and extracted using Y-parameter based models.

![Fig. 15](image2.png)  
**Fig. 15** Magnitude of the optimum source reflection factor from conventional measurements and extracted using Y-parameter based models.
5.3 Noise figure simulation

In this section we discuss optimization of the noise figure using device simulation. It has been demonstrated that the Ge-profile in the epitaxial base can be optimized for minimum noise [50, 54]. The basic low noise design principle is to increase the low current $\beta$ and $f_T$ by placing a larger part of the total Ge-content in the base close to the base-emitter pn-junction. This might degrade the high current operation but the peak $f_T$ and $f_{MAX}$ values can be kept almost unchanged. The increased low-current $f_T$ reduces the influence of the collector shot noise. For a constant $I_C$ the increased $\beta$ leads to lower $I_B$ and thus lower base shot noise. In Fig. 16 we show simulated results, using the SPICE method. Similar results are found in paper II, where the thermodynamical method was used. Both methods give similar results for $NF_{MIN}$. On the other hand the thermodynamical method predicts a different $\Gamma_{OPT}$ and hence $G_A$ will be changed. For simplicity we compare box-like Ge-profiles where the Ge-concentration is approximately uniform inside the neutral base. For this choice of Ge-profile it is found that $f_T$ as well as $\beta$ increase with increasing Ge-concentration. As expected this leads to reduced $NF_{MIN}$. In this context it is also of interest to investigate the associated gain $G_A$. It is observed, that profiles with a low $NF_{MIN}$ also have a slightly lower $G_A$. For circuit applications, where high $G_A$ is needed, this might introduce an additional trade-off between low noise and sufficient gain.

![Fig. 16 Simulated minimum noise figure and associated gain for different box-like Ge-profiles.](image-url)
5.4 Noise parameter de-embedding

Due to the relatively high losses in a low resistivity Si-substrate de-embedding or pad-compensation might be needed to extract the noise parameters of the intrinsic device. This is especially important for small devices, whereas the influence of parasitic coupling, from the pads to the substrate becomes less important for larger devices. Using deep trenches also de-couples the pads from the intrinsic device. De-embedding of the noise parameters is achieved by calculating the noise correlation matrices [55] for the device and for open and short pad-compensation structures. For the Y-parameter based noise parameter extraction methods conventional pad-compensation can be used [56]. This is illustrated here by results from a 25 GHz Si-BJT technology. The minimum noise figure $NF_{MIN}$ was extracted before and after de-embedding for devices with emitter areas of $10 \times 0.6 \ \mu m^2$ and $30 \times 0.6 \ \mu m^2$ respectively. Figure 17 illustrates the area dependence. It is found that the smaller device is slightly more affected by the pad-parasitics. For a technology without deep trenches, such as the SiGe HBT process in paper VI, the influence of pad de-embedding on the noise figure is more pronounced.

![Fig. 17 Minimum noise figure before and after de-embedding at 2 GHz for different emitter areas. 50 Ω source impedance.](image-url)
6 Harmonic and intermodulation distortion

We have investigated SiGe HBTs intended for use in RFICs. These circuits should be able to receive, transmit or amplify the RF-signal without any distortion of the original waveform. In order to accomplish this the SiGe HBT, which is used in the circuit, should ideally have a linear transfer function. For small signal conditions the transfer function can be assumed to be approximately linear. However, for increasing input power or signal amplitude the behavior of any semiconductor device is highly non-linear. This is valid for diodes and bipolar transistors as well as for field-effect transistors e.g. CMOS. The non-linearity of the device results in distortion of the RF-signal. The distortion limits the maximum signal amplitude or power, which can be used. Hence, the RF-distortion determines the upper limit of the dynamic range, whereas the noise floor gives the lower limit for the dynamic range. In the following section we give the definitions of harmonic and intermodulation distortion. This is followed by a short discussion about a novel mixed-mode simulation strategy, which was developed to investigate harmonic distortion. In the next section we discuss the different non-linear effects, which contribute to the RF-distortion in SiGe HBTs. Finally the harmonic distortion results for devices with different Ge-profiles and collector doping profiles are summarized.

6.1 Basic definitions

Non-linear effects inside the device generate output power at other frequencies than the fundamental input frequency. This is called harmonic and intermodulation distortion. The RF-distortion is affected by the circuit environment or impedance matching. The circuit topology is also important, e.g. a balanced topology is used for cancellation of even order harmonics. The generation of non-linear currents inside the device depends on both bias current and bias voltage. Hence it is important to choose a bias point, which minimizes the RF-distortion.

Figure 18 shows the simulated output power spectrum of a Si BJT in a common-emitter configuration. A two-tone RF input signal was used:

\[ \bar{v}_{\text{in}}(t) = v_1 \sin(2 \pi f_1 \cdot t) + v_2 \sin(2 \pi f_2 \cdot t) \]  

Eq. 50

where \( f_1 \) and \( f_2 \) are the two fundamental frequencies. The simulations were performed using the harmonic balance simulator HP-ADS and a MEXTRAM compact model for the Si BJT. The results illustrate the generation of harmonic and intermodulation products in the output power spectrum. The harmonic products occur at integer multiples of the fundamental frequencies - \( 2f_1, 2f_2, 3f_1, 3f_2 \) and so forth. For closely spaced input signals non-linear mixing generates output power in sidebands to the carrier. The third order intermodulation products are the most important and occur at \( 2f_1 - f_2 \) and \( 2f_2 - f_1 \).
6.2 Mixed mode circuit and device simulation

To analyze the harmonic distortion a simulation test-circuit similar to a conventional measurement set-up was used. Figure 19 shows a schematic of the test-circuit, including bias-tees for DC-feed and AC-coupling of the RF-input signal. The same configuration is used for S-parameter measurements as well as for linearity measurements, using a signal source and a spectrum analyzer. The active device (inside the dashed box) is modeled by finite element device simulation in SILVACO/ATLAS. Both the source (not shown) and the load impedance were chosen as $50 \, \Omega$. In typical RF-circuit applications the LNA will drive the input of the mixer stage, which means that the load impedance might also have a capacitive or inductive part. No specific tuning for the load impedance at the different harmonic frequencies was used. The influence of source impedance on the harmonic generation was found to be small. On the other hand changing the load impedance affects both power gain and the current dependence of the higher order harmonics.

Transient simulations (in the time domain) were used to investigate harmonic distortion. The output power at the different harmonic frequencies was calculated from a Fourier transform of the output current, delivered to the $50 \, \Omega$ load. It would also be possible to investigate intermodulation distortion using transient simulations and a two-tone input signal. However very long transient times are needed to resolve closely spaced spectral components at e.g. the fundamental frequency and third order intermodulation frequency. Therefore a systematic study to investigate intermodulation distortion becomes difficult. In general it is more efficient to solve the problem in the frequency domain. Using a harmonic balance solver, coupled to the device simulator drastically reduces the simulation time [57, 58]. Harmonic
balance simulations are mainly limited by the memory requirements, since a very large equation system needs to be solved, to include a sufficient number of harmonics.

**Fig. 19** Schematic of common emitter test-circuit used for harmonic distortion mixed mode circuit and device simulations.

### 6.3 Device physics - non-linear effects

In this section we will consider the device physics related effects on linearity. The device physics constitute the fundamental limitations for the RF-distortion. The following non-linear effects are present in bipolar devices:

The depletion capacitances $C_{BE}$ and $C_{BC}$ exhibit a non-linear bias and current dependence. At RF-frequencies these non-linear effects will be of major importance. The non-linear bias dependence is caused by the variation of the depletion region width with bias, and can be reduced by increasing the doping level. Also, for high $I_C$ when base push-out occurs $C_{BC}$ exhibits a strong current dependence [59], this effect is further discussed in the section about non-uniform collector doping profiles.

Another non-linearity is related to minority charge storage in the neutral base and also in the collector. The base-emitter diffusion capacitance $C_{dBE}$ is proportional to the collector current $I_C$, which has an exponential bias dependence. Minority charge storage in the collector becomes significant only for quasi-saturation conditions at high $I_C$.

For low $I_C$ the transconductance $g_m$ determines linearity, where:

$$g_m = \frac{\partial I_C}{\partial V_{BE}}$$

Eq. 51
Finally, for operation at medium $I_C$ and high $V_{CE}$ close to breakdown, the avalanche current generated at the base-collector junction will contribute significantly to the non-linear transfer function [60].

![Graph showing output power vs collector current]

**Fig. 20** Measured output power of fundamental (1 GHz) 2\textsuperscript{ND} and 3\textsuperscript{RD} harmonics vs. collector current.

Due to cancellation effects between these different mechanisms overall linearity is improved in the useful current range. The presence of cancellation effects have been demonstrated by turning on or off individual non-linearities in numerical simulations based on compact models or the Volterra series approach [61-63]. The relative importance of the different non-linear effects depends strongly on $I_C$. In Fig. 20 the output power at the fundamental and harmonic frequencies are shown as a function of $I_C$. A distinct local minimum in the second harmonic output power is observed for $I_C$ about 4 mA (region II). This is close to the current needed to achieve peak $f_T$. For lower $I_C$ (region I) the output power and the harmonic distortion generation are mainly determined by $g_m$. It is found that both the second and the third harmonic exhibit an exponential current dependence. In region III the high current effects become prominent. The gain at the fundamental frequency starts to drop and the second harmonic increases rapidly. In the third harmonic a small local minimum is observed. The gain roll-off for very high current operation tends to linearize the device and therefore the higher order harmonics decrease.

### 6.4 Ge-profile optimization

In this section the influence of the Ge-profile on harmonic distortion is investigated. First the choice of different simulated Ge-profiles is discussed. The starting point for the simulations was a reference device, based on SIMS measurement of the doping and Ge-profiles. This structure had a box-like Ge-profile, i.e., with approximately
constant Ge-concentration inside the neutral base. The reference was compared to other box-profiles, with higher and lower Ge-concentration respectively, to quantify the influence of the Ge-concentration at the base-emitter (BE) heterojunction. At current levels corresponding to peak $f_T$ the base-collector (BC) heterojunction must also be considered. Therefore retrograded Ge-profiles were used to study high current effects. Finally, the influence of minority charge storage in the neutral base was considered, using a graded Ge-profile.

The major influence of the Ge-profile at low to medium current operation can be readily understood by remembering that for a high Ge-concentration at the base-emitter junction a lower $V_{BE}$ is needed to achieve the same $I_C$. Therefore $C_{BE}$ will also be reduced, which can also be seen in an increased low-current $f_T$. Changing $C_{BE}$ affects both power gain and harmonic distortion. In Fig. 21 the effect on the power gain is demonstrated. Box-like profiles with varying Ge-content (8.75 % - 16.25 %) are compared.

![Graph showing simulated fundamental output power vs. base-emitter voltage for SiGe HBTs with box-like Ge-profiles.](image)

**Fig. 21** Simulated fundamental output power vs. base-emitter voltage for SiGe HBTs with box-like Ge-profiles.

The output power is shown as a function of $V_{BE}$. This illustrates that a lower $V_{BE}$ is needed to achieve high power gain for a device with high Ge-concentration at the base-emitter junction. This also translates to higher low-current $f_T$ and $\beta$. We also find that devices with lower Ge-concentration exhibit higher maximum power gain. This is explained by reduced high-injection heterojunction barrier effects and hence depends mainly on the Ge-concentration (valence band offset) at the BC heterojunction.

To illustrate the influence on harmonic distortion, in a circuit application, it is more convenient to analyze harmonic distortion as a function of $I_C$. These results are shown in Fig. 22 and are also discussed in Paper I.
In these simulations profiles with lower Ge-concentration exhibit higher power gain as a function of $I_C$. The gain roll-off due to high-injection heterojunction barrier effects appears at a lower $I_C$ for profiles with higher Ge-concentration. The influence on harmonic distortion is more complex. For low $I_C$, indicated as region I, the transconductance $g_m$ determines the linearity of the SiGe HBT. It is found that the output power of the second harmonic relative to the fundamental output power is suppressed for increasing Ge-concentration. For medium $I_C$ cancellation effects govern the total harmonic distortion. The cancellation effects become less pronounced for high Ge-concentration due to the influence of the high-injection effects. Therefore it is found that using a higher Ge-concentration at the base-emitter junction is beneficial for low-current operation but not suitable for higher currents.

The base-emitter diffusion capacitance $C_{dBE}$ is proportional to $I_C$ or more correctly the minority carrier gradient inside the neutral base. Thus devices with a constant and a graded Ge-profile will have a different minority carrier profile in the neutral base. This affects the diffusion capacitance and a difference in the harmonic distortion will be observed at medium $I_C$, where cancellation between different non-linear mechanisms in the device is important. For the investigated profiles the influence of $C_{dBE}$ on linearity was comparatively small. The influence of the Ge-grading is further discussed in Paper II.

For operation at a collector current $I_C$ close to peak $f_T$ the device performance is very sensitive to the design of the Ge-profile at the BC heterojunction. In a double heterojunction device such as the SiGe a valence band offset is present at the base-collector heterojunction. It is well known that devices with a high Ge-concentration, i.e., large valence band offset exhibit a very rapid fall-off of $\beta$ and $f_T$ at the onset of
the Kirk effect [64]. At the same time power gain and harmonic distortion will be affected. This is due to high-injection heterojunction barrier effects (HBE). During normal operation the base-collector junction is reversed biased and the band bending masks the valence band barrier. As base push-out occurs the valence band barrier becomes exposed since the junction is no longer reverse biased. This leads to increased hole storage in the neutral base. The charge neutrality condition in the base leads to the formation of an electron barrier in the conduction band. Using a retrograding of the Ge-profile towards the collector side of the base reduces the valence band offset. A high Ge-concentration at the base-emitter junction can still be utilized, in order to improve the low current properties, including noise figure, harmonic distortion and $f_T$. We illustrate the effect of the Ge-retrograding by comparing a box-like profile and a profile with a retrograding. A relatively high Ge-concentration (16.25 %) at the base-emitter junction was used for both cases. The simulated harmonic distortion is shown in Fig. 23.

![Graph](image)

**Fig. 23** Simulated harmonic distortion vs. collector current. Retrograding of the Ge-profile towards the collector shown as open symbols, box-like profile filled symbols.

It is found that the retrograding leads to small reduction of the maximum power gain. However, a clear suppression of both the second and third harmonic is observed for medium and high current operation. This example clearly illustrates the sensitivity of the harmonic distortion to changes in the Ge-profile. However, none of the profiles that have been discussed here are fully optimized for high $f_T$, low noise figure and suppressed harmonic distortion. One interesting conclusion, which is illustrated in Paper II, is that similar Ge-profiles, with increased Ge-concentration at the base-emitter junction, are useful for noise figure as well as harmonic distortion optimization.
6.5 Non-uniform collector profiles

The mixed-mode simulation methodology is also well suited to investigate the collector doping profile. The main reason for this is the possibility to investigate arbitrary collector doping profiles. However, due to the relatively long simulation time a simpler approach would be preferred. In this study 2D device simulations and mixed mode simulations were used. Other authors have used a 1D device simulator and a harmonic balance circuit simulator to investigate the collector doping profile for low- and high-voltage Si BJTs [65, 66].

In this section we discuss the advantages of using a non-uniform collector profile. As was described above a trade-off exists between $BV_{CEO}$ and $f_T$. By considering the different non-linear mechanisms in a bipolar transistor we can find a collector profile, which reduces the RF-distortion. The collector profile influences RF-distortion through four different mechanisms: Quasi-saturation, avalanche current generation, bias and current dependence of $C_{BC}$. Below we discuss the relative importance of these effects in more detail. The main conclusions will be illustrated by simulation results.

Quasi-saturation or forward biasing of the collector-base junction occurs either due to a resistive voltage drop in the collector region or due to the Kirk effect. In high-speed devices quasi-saturation occurs mainly due to the Kirk effect since the collector doping level is relatively high ($\sim 1 \times 10^{17}$ cm$^{-3}$ or higher). Using a highly doped collector layer also reduces the bias dependence of $C_{BC}$. This comes at the expense of a reduced $BV_{CEO}$. Using a non-uniform profile, allows both breakdown and $C_{BC}$ bias-dependence to be improved. The collector resistance for a retrograded profile is also reduced, compared to a uniform profile with comparable $BV_{CEO}$ and $f_T$.

The influence of the avalanche current on the SiGe linearity was treated in some detail by Niu using the Volterra series approach [60]. It was found that the avalanche multiplication factor decreased with increasing $I_C$. Hence, at typical operating currents the avalanche generation will be less important than the $C_{BC}$ non-linearity, which becomes dominant. In our study we have used a relatively low collector voltage $V_{CE}$ of 1.5 V for a device with $BV_{CEO}$ of approximately 3 V. Simulations with higher $V_{CE}$ (2 – 3.1 V) with and without impact ionization indicated that the influence of non-linear avalanche effects were small in the present devices.

For high current when base push-out occurs the base-collector depletion capacitance exhibits a strong current dependence. In Fig. 24 this effect is illustrated by plotting $C_{BC}$ and $f_T$ as a function of collector current. $C_{BC}$ was extracted from the simulated S-parameters, using the following relation:

$$\omega C_{BC} = -\text{imag}(Y_{12})$$

Eq. 52
The simulations have been verified with measurements on high-speed Si BJTs, where similar behavior could be observed. The rise in $C_{BC}$ occurs at the collector current corresponding to peak $f_T$ and can be explained by two different physical mechanisms. Firstly, hole injection into the collector epi-layer occurs as the base-collector junction becomes forward biased. This leads to a build-up of minority charge and hence results in an increased diffusion capacitance. The base-collector diffusion capacitance is insignificant during normal operation, i.e., before the onset of quasi-saturation. Accurate modeling of the minority charge is important in compact models of high-speed bipolar transistors [61, 67, 68]. Secondly, as the electric field at the base collector junction collapses the width of the base-collector depletion region is reduced, resulting in an increased depletion capacitance. This effect is most pronounced at higher currents, where a very steep slope of $C_{BC}$ is observed. The current dependence of $C_{BC}$ can be reduced, by using a higher collector doping. Changing the collector doping profile not only affects the current and bias dependence of $C_{BC}$. An increased $C_{BC}$ also leads to a reduced power gain due to the feedback from output to input in the common emitter configuration. Therefore the choice of profile is a trade-off between power gain and reduced harmonic distortion. The feedback also affects the generation of third order intermodulation products, resulting from the mixing of the second harmonic and the fundamental signal.

![Graph of Collector-base capacitance and cut-off frequency vs Collector current](image)

**Fig. 24** Simulated current dependence of base-collector capacitance and cut-off frequency.

For improved breakdown and high current $f_T$ a retrograded profile where the doping increases towards the buried collector and a profile with the peak doping placed closer to the collector-base junction were implemented in the simulations. A uniform collector doping was used as a reference. The different collector profiles are shown in Fig. 25.
In Table 1 the electrical characteristics for the different profiles are summarized.

<table>
<thead>
<tr>
<th>Profile</th>
<th>$f_T$ (GHz)</th>
<th>$BV_{CEO}$ (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>46.0</td>
<td>2.30</td>
</tr>
<tr>
<td>Retrograde</td>
<td>45.1</td>
<td>2.60</td>
</tr>
<tr>
<td>Peak</td>
<td>47.1</td>
<td>2.45</td>
</tr>
</tbody>
</table>

Figure 26 summarizes the harmonic distortion characteristics of devices with uniform and non-uniform collector doping profiles. The devices with a non-uniform profile have slightly lower power gain. For medium current operation lower than the peak $f_T$ current (1mA) the second harmonic is suppressed for the non-uniform profiles. This can be attributed to the cancellation between different non-linear effects. For high current operation beyond the onset of the Kirk effect the second harmonic is suppressed for the non-uniform profiles due to the reduction in the $C_{BC}$ current dependence. The third harmonic is also significantly suppressed at these high current levels. From the relatively small changes in harmonic distortion for different collector profiles we conclude that a non-uniform profile is to be preferred, due to the improved $BV_{CEO}$ and almost unchanged $f_T$. It was shown in Paper I that the harmonic distortion of SiGe HBTs could be improved by both tailoring the collector doping profile and the Ge-profile. However the harmonic distortion was more sensitive to changes in the Ge-profile and the leverage of a non-uniform collector profile was comparatively small.
Fig. 26 Simulated harmonic distortion for uniform and non-uniform collector doping profiles.
7 Low-frequency noise

In non-linear circuits such as oscillators and mixers the low-frequency noise should be minimized to avoid undesired phase noise or spectral broadening. In this Chapter we first consider the origin of the dominant low-frequency noise sources in polysilicon emitter SiGe HBTs and Si BJTs. It has been found that the polysilicon/monosilicon emitter interface plays an important role to determine the noise level. This will be discussed with emphasis on SiGe HBT structures. Further, in Paper VII reduction of the low-frequency noise by hydrogen passivation of interfacial traps was investigated. Finally, the main noise sources in a SiGe HBT structure with a non-selectively grown base are identified. This was studied in Paper III, by comparing devices with different designs of the extrinsic base region.

The power spectral density of the low-frequency noise in bipolar transistors exhibits a characteristic $1/f^{\gamma}$ frequency dependence, where $\gamma$ is close to unity. Figure 27 shows the power spectral density of the input referred noise ($S_{IB}$) for a SiGe HBT measured for four different base currents. The noise is found to increase with base current $I_B$. At frequencies higher than 10 kHz the shot noise level corresponding to $2qI_B$ is clearly visible.

![Figure 27](image)

**Fig. 27** Power spectral density of the input referred noise in a SiGe HBT (reproduced from Paper III).

The main source for the $1/f$-type low-frequency noise in polysilicon emitter bipolar transistors is a trapping-detrapping mechanism at the thin SiO$_2$ interface between the polysilicon and monosilicon emitter [69]. It has been found that the noise level can be related to thickness and quality of the SiO$_2$ layer [70]. Standard SiGe HBT processes utilize a polysilicon emitter technology. Therefore the noise performance
is expected to be similar to Si BJTs. It has been shown that the introduction of a strained SiGe epitaxial base layer does not alter the noise levels [71]. However, thanks to increased $\beta$ it is possible to remove the SiO$_2$ interface and utilize a regrown single crystalline emitter. This has been shown to considerably reduce the noise level [72]. In Si BJTs this approach leads to a trade-off between noise and $\beta$ [70]. In Paper VII we discuss passivation of noise sources at the thin SiO$_2$ interface by means of a hydrogen anneal. It is well known that a hydrogen anneal suppresses the non-ideal base current components. Measurement on transistors before and after an anneal showed that low-frequency noise is effectively reduced by hydrogen passivation. Since the noise source could be located to the emitter area this result suggests that electrical quality of the polysilicon emitter interface has been improved. The reduced noise after hydrogen anneal indicates that the number of defects, which contribute to carrier trapping-detrapping process has been reduced. Several types of electrical defects might be present at a SiO$_2$ interface and thus contribute to the low-frequency noise [73]. However, it is well known that the so-called Pb center is deactivated by hydrogen, which is consistent with the observed reduction in noise level.

The results from noise studies on SiGe HBTs as well as Si BJTs indicate that interfacial engineering is the most effective way to reduce the low-frequency noise. However, it is also important to consider other possible contributions to the total noise. A well-known example is the increased noise level in devices, which have been subjected to electrical stress (high fields) [74, 75]. In this case the excess noise contribution originates from traps, located at the Si/SiO$_2$ interface in the base-emitter depletion region. In Paper III we demonstrate that for certain designs of the SiGe HBT extrinsic base region additional noise sources will have a significant contribution. The investigated device designs differed in the position of the extrinsic base implantation. For some designs the implantation was placed inside the epitaxial base, whereas for other devices only the polycrystalline part of the SiGe base was implanted. This affects the low-frequency noise in at least two ways. Firstly, the interface between the epitaxial SiGe and the polycrystalline SiGe might contain a large number of defects, such as dangling bonds, acting as recombination centers. This leads to increased low-frequency noise [76]. If the region, containing the interface, is implanted with a high dose of boron the defects might be passivated and thus no longer contribute to the noise. Secondly, ion implantation into epitaxial SiGe creates a large number of point defects, Si self-interstitials or defect clusters. These defects are removed by a thermal anneal, which also activates the dopant atoms. However, for a low thermal budget such as in the process reported in Paper V not all defects may be annealed out. In devices where the epitaxial region had received an ion implantation a significant increase in the noise level was observed, suggesting that the implantation defects act as additional noise sources. For some of the transistor designs the noise power spectrum also deviates from the $1/f$ dependence. Several Lorentzian shaped generation-recombination (g-r) bumps are observed. The g-r bumps are superimposed on the $1/f$-type noise, which is also present in the transistors. The g-r noise sources are assumed to be located in the base-emitter depletion region. In some cases the shape of the noise spectrum might exhibit a bias dependence [77]. This is due to the band bending in the depletion region, which activates and deactivates the traps acting as noise sources [78, 79]. In conclusion we find that for an optimized design, where these excess contributions have been
removed, the dominant noise source can be located to the polysilicon monosilicon emitter SiO₂ interface.
8 Concluding remarks

The main focus of this thesis has been on device simulation and RF-optimization issues for SiGe HBTs. Physically based simulations have been calibrated to measurement and good agreement has been demonstrated. SiGe HBTs have been fabricated and characterized and devices with a cut-off frequency of more than 60 GHz were successfully demonstrated. We have also investigated high-frequency noise, low-frequency noise, and RF-distortion.

SiGe HBTs with a non-selectively grown epitaxial base were investigated. In particular we have addressed base profile control, i.e., transient enhanced diffusion (TED) of boron. TED effects caused by the extrinsic base implantation were found to have a significant influence on DC and HF-characteristics. We have also discussed the influence of extrinsic base design on the low-frequency noise. It was found that the low-frequency noise in the SiGe HBTs could be reduced by an optimized design of the extrinsic base region and that the dominant noise sources could be located to polysilicon/monosilicon SiO₂ emitter interface. In addition it was shown that the hydrogen anneal at the end of the process line improved the low-frequency noise properties of polysilicon emitter bipolar transistors.

Furthermore, we have considered optimization of the collector profile for high cut-off frequency and breakdown voltage. A novel collector fabrication concept using low-energy antimony implantation was investigated experimentally and by process and device simulations. Devices with a double collector implantation using both an antimony implantation and a standard phosphorous SIC showed improved cut-off frequency. However, it was found that antimony segregation during the growth of the epitaxial base led to an undesired broadening of the implanted antimony profile.

We have also investigated the high-frequency noise of SiGe HBTs, using two recently developed Y-parameter based noise parameter extraction methods. Both extraction methods showed good agreement to conventional noise figure measurement. However, the so-called SPICE method was found to give significantly better agreement for the optimum source admittance, which is important for low noise circuit design.

In the final part of this work the influence of the circuit environment was also considered. A novel mixed-mode circuit and device simulation methodology was developed to investigate RF harmonic distortion. A relatively simple test circuit including a SiGe HBTs and passive components for DC-feed and AC-coupling of the RF-signal was used. The source and load impedance was chosen to 50 Ω. The calibrated simulations showed good agreement with measurement on a reference device. The influence of Ge-profile and collector doping profile tailoring on the RF harmonic distortion was quantified. Guidelines for choice of Ge-profile for high and low-current operations were discussed in some detail. In particular the influence of high-injection heterojunction barrier effects on SiGe HBT harmonic distortion was discussed for the first time.
SiGe HBTs have demonstrated a large potential for use in RF-applications and high-speed optical networks. The development has been very rapid and is expected to continue. An important driving force for the SiGe technology development is the continuing evolution of standard Si process technology. It should be pointed out that a SiGe HBT module can be integrated in a conventional Si CMOS process flow, which allows volume production with good yield, using state-of-the-art tools and equipment. Furthermore flexible circuit solutions for use in e.g. cellular phones can be realized by combining these two technologies. New application areas will also add to the demand for SiGe-based circuits, which offer high performance combined with relatively low cost. As an example the recent introduction of SiGe:C HBTs has significantly improved the performance. As a result SiGe:C HBTs have become a promising technology for use in high-speed 10 – 40 Gbit optical networks. Recently RF-CMOS have emerged as a strong competitor to SiGe HBTs. Due to the continuing downscaling of device dimensions Si-based 0.18 µm RF-CMOS now offers more than sufficient performance for many of the wireless applications where SiGe HBTs are presently used. The short gate length leads to $f_T$ and $f_{MAX}$ values close to 100 GHz. Important passive elements, such as spiral inductors with high quality factors, can be realized in a CMOS process flow. However, so far SiGe HBT circuits consume less power and also exhibit very good low-frequency and high-frequency noise performance.

Finally, we give some examples of important issues for future research. It is clear that dedicated compact (SPICE) models for SiGe HBT circuit simulations will be needed for accurate prediction of high current operation as well as harmonic distortion. For development of SiGe:C HBTs calibrated process simulations as well as physical models and parameters for device simulations will be very important.
References


