Device Simulation of High-Performance SiGe Heterojunction Bipolar Transistors

vorgelegt von
M.Sc.
Julian Korn
geb. in Heidelberg

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
-Dr.rer.nat.-

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Martin Schneider-Ramelow
Gutachter: Prof. Dr. Bernd Tillack
Gutachter: Prof. Dr.-Ing. Christoph Jungemann
Gutachter: Prof. Dr.-Ing. Wolfgang Heinrich

Tag der wissenschaftlichen Aussprache: 27. Februar 2018

Berlin 2018

Abstract

Silicon-germanium (SiGe) heterojunction bipolar transistors (HBT) are well suited for high-frequency applications. Their performance has been improved continuously in recent years. Today's SiGe HBT technologies show transit frequencies f_T up to $300\,\text{GHz}$ and maximum oscillation frequencies up to $500\,\text{GHz}$.

Numerical device simulation plays an important role in the development of SiGe HBTs. Possible optimizations of the transistor can be evaluated by simulation, which reduces the number of necessary test wafers. Furthermore, device simulation helps to explore the physical mechanisms that govern the performance of the SiGe HBTs.

The benefit of device simulations depends on their predictive power. Limitations in the underlying model of charge transport can lead to false simulation results. Device simulation based on the hydrodynamic transport model is still the workhorse for the optimization and investigation of SiGe HBTs. More rigorous models such as the Boltzmann transport equation are computationally very expensive, which considerably limits their use.

In this thesis, the ability of state-of-the-art hydrodynamic simulation to predict the RF-performance of advanced SiGe HBTs is evaluated. For this purpose, a comprehensive comparison between measured and simulated electrical characteristics is made. SiGe HBTs with a scaled vertical doping profile and a transit frequency above $400\,\text{GHz}$ are used in this investigation. The impact of variations of the vertical doping profile on the transit frequency is investigated by simulation and experiment. Possible optimizations of the lateral architecture of the SiGe HBT are explored by means of simulation.

Zusammenfassung

Silicium-Germanium (SiGe) Heterobipolartransistoren (HBT) sind für Höchstfrequenz-anwendungen gut geeignet. Ihre Leistungsfähigkeit bei höchsten Frequenzen wurde in den letzten Jahren stetig verbessert. Heutige SiGe HBT Technologien verfügen über Transitfrequenzen f_T von bis zu $300\,GHz$ und maximale Oszillationsfrequenzen von bis zu $500\,GHz$.

Numerische Bauelementsimulation nimmt in der Entwicklung von SiGe HBTs eine wichtige Rolle ein. Mögliche Optimierungen des Transistors können vorab mit Hilfe der Simulation getestet werden, was zu einer Reduzierung der Anzahl der benötigten Testwa-fer führt. Zusätzlich trägt die Bauelementsimulation zum Verständnis der physikalischen Mechanismen bei, welche die Leistung des SiGe HBTs bestimmen.

Der Nutzen der Bauelementsimulation für die Entwicklung neuer Generationen von SiGe HBTs hängt von deren Vorhersagekraft ab. Unzulänglichkeiten des zugrundeliegen-den Modells des Ladungstransports können zu falschen Simulationsergebnissen führen. Bauelementsimulation, welche auf dem hydrodynamischen Modell des Ladungstrans-ports basiert, ist die meistverwendete Methode zur numerischen Untersuchung und Op-timierung von SiGe HBTs. Exaktere Modelle des Ladungstransports wie die Boltzmann-Transportgleichung erfordern einen sehr hohen Rechenaufwand, welcher ihre Anwendung deutlich einschränkt.

In dieser Arbeit wird die Fähigkeit der hydrodynamischen Bauelementsimulation zur Vorhersage der Hochfrequenz-Leistungsfähigkeit moderner SiGe HBTs untersucht. Hier-zu wird ein umfassender Vergleich zwischen simulierten und gemessenen elektrischen Kenngrößen angestellt. Für diesen Vergleich werden SiGe HBTs mit einem skalierten vertikalen Dotierungsprofil verwendet, welche Transitfrequenzen über $400\,GHz$ aufwei-sen. Der Einfluss des vertikalen Dotierungsprofils wird experimentell und simulativ un-tersucht. Mögliche Optimierungen der lateralen Transistorarchitektur werden mit Hilfe der Simulation evaluiert.

Contents

Contents

1. Introduction

Silicon-germanium (SiGe) heterojunction bipolar transistors (HBT) are well suited for radio-frequency (RF) applications. They provide high cut-off frequencies, can handle high power densities and have a high current drive capability and low noise. The performance of SiGe HBTs has been improved continuously in recent years. Today, SiGe HBTs are widely used in applications in the mm-wave range, which have traditionally been the domain of III-V compound semiconductors [1, 2]. Modern SiGe HBT technologies such as IHPs SG13G2 reach frequencies of several hundred GHz [3]. Major drivers for this development are applications like broadband communication, automotive radar or millimeter-wave sensing and imaging [4, 5, 6, 7].

Numerical device simulation plays an important role during the development of new technology generations. Variations and optimizations of the device can be evaluated by simulation which helps to reduce the number of test wafers. For example, simulation can be used to predict the impact of optimized doping profiles, device geometries and material compositions on the electrical characteristics of the transistor. The benefit of such simulations depends on their predictive power. Limitations of the physical models which describe the carrier transport can lead to false simulation results. For this reason, a lot of effort has been put into the improvement of the simulation tools. The validity of conventional device simulation methods based on the drift-diffusion and the hydrodynamic transport models has been extended continuously by including sophisticated models for transport parameters like the mobility [8, 9, 10]. Moreover, advanced simulation methods, based on the solution of the Boltzmann transport equation have been developed and applied to SiGe HBTs [11].

The main objective of this work is to evaluate how accurate such simulation methods can predict the performance of modern SiGe HBTs. This evaluation is based on a comparison of measured and simulated electrical characteristics. Dedicated reference transistors with an advanced vertical doping profile are used for this comparison. Device simulation of the transistors is performed using hydrodynamic transport with calibrated parameter models.

In the following section, the basic principles and key properties of SiGe HBTs are shortly described. Some remarks on TCAD (technology computer-aided design) for SiGe HBTs are provided in Section 1.2 and an overview of the content of this thesis is given in Section 1.3.

1.1. **The SiGe HBT**

The silicon-germanium (SiGe) heterojunction bipolar transistor (HBT) can be regarded as an advanced version of the conventional silicon bipolar junction transistor (Si BJT). In a SiGe HBT, the base region is formed by a SiGe layer, which is placed between collector and emitter. The resulting Si-SiGe-Si heterostructure leads to a reduced bandgap in the base. The bandgap of germanium is significantly smaller than the bandgap of silicon (0.66 eV compared to 1.12 eV at room temperature). Moreover, the relatively small lattice mismatch of 4.2 % allows pseudomorphic growth of strained SiGe on a silicon substrate. The critical thickness of a SiGe film with 30 % Ge is about 8 nm. By capping the SiGe layer with Si, it is possible to deposit pseudomorphic layers with a thickness exceeding the critical value [12].

A reduction of the bandgap in the base of the HBT leads to an exponential increase of electron injection from the emitter into the base because it reduces the potential barrier at the emitter-base junction. This results in a strong increase of the current gain β compared to a Si BJT with the same doping:

$$\frac{\beta_{SiGe}}{\beta_{Si}} \propto \exp\left(\frac{\Delta E_g}{kT}\right) \tag{1.1}$$

A HBT with a Ge mole fraction of 20 %, for example, has a band gap difference ΔE_g of about 140 meV which results in an enhancement of β by a factor of 148.

By tailoring the alloy composition of the SiGe base, one can optimize the characteristics of the transistor for particular applications. This approach is often referred to as bandgap engineering. Figure 1.1 shows the schematic band diagrams of a graded-base npn SiGe HBT and a corresponding npn Si BJT with uniform doping concentrations in base, emitter and collector. The bandgap difference ΔE_g between Si and $Si_{1-x}Ge_x$ depends on the Ge mole fraction x. Thus, the desired spatial variation of the bandgap can be attained by shaping the Ge profile accordingly. The triangular Ge profile shown in Fig. 1.1 results in a reduction of the conduction band barrier as well as in a conduction band gradient in the neutral base.

The maximum performance of conventional Si BJTs is fundamentally limited by the inherent trade-off between current gain and base transit time. To operate at high frequencies, a short base transit time is necessary which can be achieved by a small base width. A thinner base requires a higher doping concentration to avoid punch-through breakdown and to maintain a sufficiently low base resistance. However, a higher doping concentration in the base also reduces majority carrier injection from the emitter into the base, which results in a degradation of the current gain. The lower bandgap in the base of the SiGe HBT allows a high doping concentration while maintaining a high current. This enables device designs with a reduced base width leading to a strong performance improvement compared to Si BJTs. An additional reduction of the base transit time can be achieved by a graded Ge profile as indicated in Fig. 1.1. The additional drift field that is caused by the Ge gradient accelerates the diffusive transport of electrons in the neutral base.
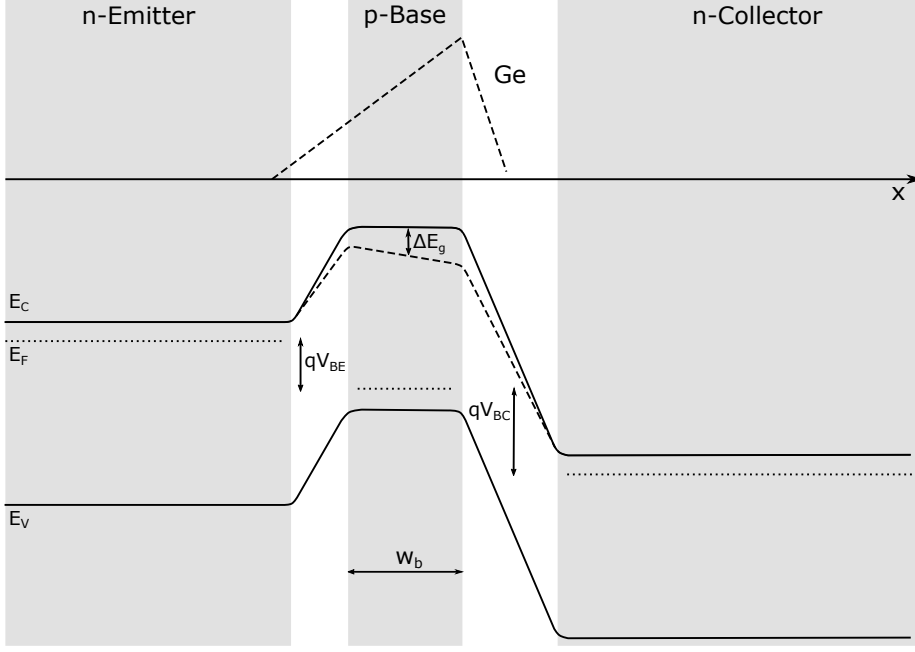
Figure 1.1.: Schematic band diagram of a SiGe HBT with a linearly graded Ge profile (dashed lines) and a corresponding Si BJT (solid lines), biased in forward active mode. The grey background highlights the neutral regions in contrast to the space charge regions.

Furthermore, a high Ge concentration at the collector side of the base leads to a high Early voltage (which characterizes the dependence of the collector current on the base-collector voltage) [13].

1.1.1. Figures of Merit

The most important figures of merit to characterize the high-frequency performance of SiGe HBTs are the cutoff or transit frequency f_T and the maximum oscillation frequency f_{max}. The cutoff frequency f_T is defined as the frequency at which the current gain of the transistor becomes unity. Circuit designers typically use operating frequencies much lower than f_T because it becomes increasingly difficult to design a circuit at a frequency near f_T. An analytical expression that relates f_T to the relevant device parasitics can be derived from a small-signal equivalent circuit model [14]:

$$f_T = \frac{1}{2\pi \left(\tau_f + \frac{C_{jEB}+C_{jBC}}{g_m}\right)} \tag{1.2}$$

Here, τ_f is the forward transit time, C_{jEB} and C_{jBC} are the depletion capacitances of the base-emitter and base-collector junction respectively and g_m is the transconductance. The forward transit time describes the delay that is associated with the storage of minority charge carriers. For further analysis, τ_f can be separated into individual transit times for the different regions of the transistor (cf. Section 4).

The maximum oscillation frequency f_{max} is defined as the frequency at which the power gain of the transistor becomes unity. An expression for f_{max} as a function of f_T can also be derived from the equivalent circuit [15]:

$$f_{max} = \sqrt{\frac{f_T}{8\pi C_{jBC} R_B}} \tag{1.3}$$

Here, one can see that f_{max} benefits from an increase of f_T and that it also depends on the base resistance R_B and the capacitance C_{jBC}.

1.1.2. SiGe HBT Performance Factors

The technological measures that have led to continuous improvements in the performance of SiGe HBTs can be divided into two categories: lateral scaling and vertical scaling. Vertical scaling mainly refers to the optimization of the vertical doping profile with the aim to reduce the transit time τ_f. On the other hand, lateral scaling refers both to the shrinkage of lateral device dimensions as well as to improvements of the lateral device architecture.

Vertical scaling involves optimizations of the emitter, the base and the collector of the transistor. Scaling of the emitter region aims at producing a shallow base-emitter junction to reduce the transit time across the BE depletion region. Moreover, a mono-crystalline emitter with high in-situ doping helps to reduce the emitter resistance R_E [16]. The base transit time is reduced by a shrinkage of the base width w_b and by an optimized concentration gradient in the Ge profile [17]. A significant progress in reducing the base width was possible by introducing carbon to suppress boron diffusion during the annealing process [18, 19]. Recently, further improvement has been achieved by using millisecond flash anneal to reduce the thermal budget [20, 21]. Scaling of the collector region aims at reducing the transit time through the base-collector depletion region and at suppressing degradation at high current densities (Kirk effect). Vertical scaling has lead to a strong increase of f_T in the past, but it also resulted in an increased current density, increased local electric fields and increased impact ionization [22]. For scaled vertical profiles, it might become difficult to maintain sufficient breakdown voltages and a low base resistance. Furthermore, self-heating becomes a serious issue with increasing current density.

Lateral scaling aims at a reduction of device parasitics by a combination of structural optimizations and shrinking device dimensions. The main purpose is to reduce the base resistance R_B and the external base-collector capacitance C_{BCx} in order to counteract the degradation of these parameters due to vertical scaling. An illustration of the lateral device geometry of a high-performance HBT is shown in Figure 1.2. This geometry corresponds to a HBT from IHPs $0.13\,\mu$m technology SG13G2 [3]. Specific features of this transistor architecture are the elevated extrinsic base, the implanted collector without deep-trench isolation and that the whole transistor is formed in a single active area, which means that there is no shallow trench isolation (STI) between emitter and collector contacts [23].
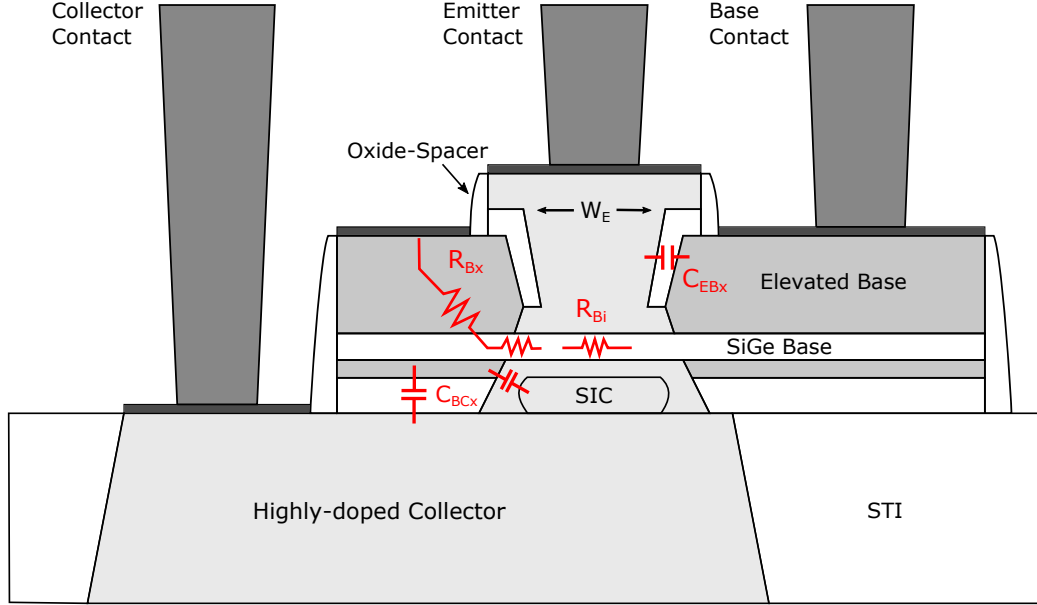
Figure 1.2.: Schematic cross section of a high-performance SiGe HBT from IHPs $0.13\,\mu$m technology SG13G2.

Lateral scaling involves a number of trade-offs between different parasitics. The internal base resistance R_{Bi}, for example, can be reduced by shrinking the emitter width W_E, which itself would lead to an increase of f_{max} according to (1.3). However, at the same time a smaller emitter window increases the relative external contributions to the base-emitter and base-collector capacitances C_{EBx} and C_{BCx}, which can lead to a degradation of f_T. Similar trade-offs between base resistance and external capacitance exist for the width of the base-emitter spacers and for the width of the external base. A guideline for the design of the lateral device dimensions is the desired ratio of f_T and f_{max}. Usually, high speed SiGe HBTs have a ratio f_{max}/f_T between 1 and 2.

1.1.3. Recent Developments

The development of SiGe HBTs during the last decade was primarily pushed forward by the EU-funded projects DOTFIVE and DOTSEVEN. Prior to the beginning of the DOT-FIVE project, the record value of f_{max} was 350 GHz [24]. This value was first increased to 500 GHz during DOTFIVE and finally to 720 GHz at the end of DOTSEVEN. One important result of DOTFIVE was that the conventional DPSA-SEG transistor structure, which is most commonly used in high-performance SiGe HBT technologies today, has a limited potential for further improvements [23]. The acronym DPSA-SEG stands for Double-Poly-silicon Self-Aligned emitter/base with Selective Epitaxially Grown base. All attempts during DOTFIVE to push f_{max} of conventional DPSA-SEG HBTs beyond 400 GHz have been unsuccessful. The main reason for the limited performance of this architecture is the high resistance of the vertical base link [25]. Two alternative HBT concepts with a lateral base link had been developed by IHP, one with selective and

another one with non-selective base epitaxy. Both achieved much higher f_{max} than the conventional architecture. The transistor with non-selective base epitaxy has then been integrated into IHPs $0.13\,\mu m$ BiCMOS technology SG13G2. Continued vertical and lateral scaling during DOTSEVEN have led to a record f_T/f_{max} of $505\,GHz/720\,GHz$, demonstrating the high potential of this transistor architecture [21]. A summary of published values of f_T and f_{max} from recent SiGe HBT technologies is shown in Figure 1.3.
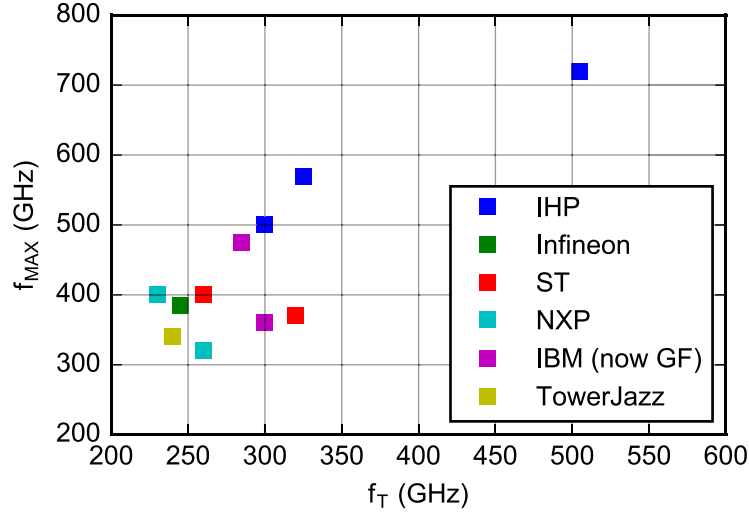
Figure 1.3.: Published values of f_T and f_{max} from several high-speed SiGe HBT technologies. Values are taken from [26, 27, 21] (IHP), [25] (Infineon), [23, 28] (ST), [29] (NXP), [30, 20] (IBM/Globalfoundries) and [31] (TowerJazz).

1.2. TCAD for SiGe HBTs

The term TCAD (technology computer-aided design) usually refers to a set of software tools to support the development of semiconductor technologies [32]. Commercially available TCAD packages typically include tools for process and device simulation, tools to define simulation structures, meshing tools for the creation of simulation grids and tools for parameter extraction. The core part of TCAD is physics-based process and device simulation. Process simulation software is used to simulate the fabrication of semiconductor technologies. Such tools include models for various process steps such as etching, deposition, oxidation, ion implantation and diffusion.

The main purpose of device simulation tools is to calculate the electrical characteristics of semiconductor devices. Additional features are the simulation of optical and thermal properties. The classical TCAD approach to simulate carrier transport is based on the drift-diffusion model. The drift-diffusion model is valid for devices with a minimal feature size in the micrometer range. The underlying assumption is that the mean free path of the carriers between two scattering events is much smaller than typical device dimensions. In this case, one can assume that the carriers are in thermal equilibrium with the lattice. High electric fields, which frequently appear in scaled devices, can

lead to a significant deviation from thermal equilibrium. This leads to non-local effects such as velocity overshoot [33], which occurs for example in the base-collector depletion region of SiGe HBTs. Such effects can be captured within TCAD by hydrodynamic or energy-balance transport models, which include an additional balance equations for the carrier energy. A more sophisticated description of carrier transport is given by the semiclassical Boltzmann transport equation (BTE) which supports detailed models for the band structure and scattering mechanisms and allows to simulate ballistic transport in nanometric devices. In extremely scaled devices, quantum effects become relevant. The semiclassical transport models can be augmented with quantum models to account for effects such as tunneling through the gate oxide or quantization in a 2D electron gas in the channel of a MOSFET. Several approaches to full quantum transport simulation exist, but these models are not an established part of commercial TCAD frameworks yet [34].

State-of-the-art simulation of SiGe HBTs is based on a hierarchy of different simulation approaches. The most rigorous approach is based on the Boltzmann transport equation which can be solved by stochastic Monte Carlo (MC) simulation [35] or by deterministic methods such as the spherical harmonics expansion (SHE) [36, 11]. Simulations based on the BTE are computationally very expensive, which limits the details of simulated devices to 1D or small 2D domains with only few parasitics. For this reason, classical TCAD based on the drift-diffusion or the hydrodynamic model is still the workhorse in scientific and industrial research. It makes it possible to simulate large realistic device structures with a reasonable computational expense.

In order to get reliable results with the classical TCAD approaches, well calibrated models for the transport parameters are required. Such models have been developed during the DOTFIVE project. These models have been calibrated to the results of advanced simulations based on the BTE [8, 9, 10]. However an experimental verification of these parameter models has not yet been done.

1.3. Thesis Content

Calibrated parameter models for the simulation of SiGe HBTs with the hydrodynamic transport model have been developed during the DOTFIVE project. In this work these models are applied in the investigation of advanced SiGe HBTs. The aim of this investigation is twofold: firstly to evaluate the accuracy of state-of-the-art TCAD for SiGe HBTs; and secondly to understand the mechanisms that limit the RF-performance of SiGe HBTs and explore possible measures to optimize them. For this purpose, simulations are compared with experimental results from transistors with an advanced vertical doping profile. These transistors have been fabricated during the development of a 700 GHz SiGe HBT within the context of the DOTSEVEN project. Special attention has been paid to the characterization of the doping profile, which goes beyond the usual method of basic SIMS measurements.

This thesis is organized as follows. In Chapter 2, the device simulation framework that is used in this work is described. An overview of the hydrodynamic transport model

and its derivation from the Boltzmann transport equation is given and the calibrated parameter models are described. Furthermore, the effective bandgap in the SiGe base of the HBT is determined experimentally for different SiGe alloy compositions using dedicated reference transistors.

In Chapter 3, a comprehensive comparison of simulated and measured electrical characteristics for a set of HBTs with an advanced vertical doping profile is presented. The vertical doping profile of these transistors is determined by a combination of different experimental techniques, including secondary ion mass spectroscopy (SIMS), energy dispersive X-ray spectroscopy (EDX) in a transmission electron microscope (TEM) and X-ray diffractometry (XRD).

In Chapter 4, the impact of different variations of the vertical doping profile is investigated by simulation and experiment. The 1D quasi-static transit time analysis is used to understand how the vertical profile affects the RF-performance of the SiGe HBTs.

In Chapter 5, possible optimizations of the lateral device geometry are explored by simulation and a comparison between the lateral geometry of the SG13G2 HBT and the 700 GHz HBT from DOTSEVEN is made. Furthermore, the effect of boron diffusion from the extrinsic base into the inner region is investigated.

2. Device Simulation Framework

The simulations in this work have been performed with the commercially available TCAD software Sentaurus Device by Synopsys [37] using the hydrodynamic transport model. A description of the model equations and an overview of their derivation from Boltzmann's transport equation is given in Section 2.1. The accuracy of HD simulations strongly depends on the quality of the parameters that describe the physical properties of the carrier transport. These parameters are the mobility, the energy relaxation time, the effective density of states and the bandgap. In this work, calibrated models of the mobility, the energy relaxation time and the effective density of states are used, which have been developed by Sasso et al. [8, 9] within the context of the EU project DOTFIVE [38]. These parameter models are described in Section 2.2.

The simulated collector current of the transistor is mainly determined by the bandgap E_g. Realistic simulation results can only be achieved if the dependence of E_g on the germanium mole fraction x is known accurately. Measured values for the effective bandgap in SiGe and SiGe:C HBTs can be found in the literature [39, 40, 41, 42, 18, 43, 44]. However, these values have been determined using certain assumptions about the mobility and effective density of states which may be inconsistent with the models that are used in this work. For this reason, the effective bandgap is determined in Section 2.3 from measurements of the collector current. For this purpose, a set of HBTs with box-shaped base profiles and varying $Si_{1-x}Ge_x$ alloy compositions has been fabricated.

2.1. Semiconductor Equations

In this section, a derivation of the hydrodynamic as well as the drift-diffusion transport model from the Boltzmann equation is briefly outlined. A comprehensive discussion of the various approximations that are used in the derivation of transport models from the Boltzmann equation can be found in the textbook by Lundstrom [45] and in the review by Grasser et al. [46].

2.1.1. Boltzmann Equation

The Boltzmann equation is the fundamental equation of semiclassical transport theory. It describes the state of the carriers in the device by a distribution function $f(\mathbf{r}, \mathbf{k}, t)$ in the six-dimensional phase space, which is spanned by the spatial coordinate $\mathbf{r}$ and the wave vector $\mathbf{k}$. The BTE is a balance equation for the distribution function f, describing

the kinetics of carriers under influence of an electric field $\mathbf{E}$. It is given by [45]

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \frac{q}{\hbar} \mathbf{E} \cdot \nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t} \right)_{coll}, \tag{2.1}$$

where $\mathbf{v}$ is the group velocity, q is the positive electron charge and $\hbar$ is Planck's constant divided by 2π. The right hand side of this equation is the collision integral, which describes the change of the distribution function due to scattering processes. Generally, it includes all possible scattering processes such as collisions with ionized impurities, phonons, crystal defects and other carriers as well as generation and recombination processes. In the case of intraband single-electron processes, the collision integral is given by

$$\left(\frac{\partial f}{\partial t} \right)_{coll} = \sum_{k'} \left\{ W_{kk'} f(\mathbf{r}, \mathbf{k}', t)[1 - f(\mathbf{r}, \mathbf{k}, t)] - W_{k'k}[1 - f(\mathbf{r}, \mathbf{k}', t)] f(\mathbf{r}, \mathbf{k}, t) \right\} \tag{2.2}$$

Here, $W_{kk'}$ denotes the transition probability from $\mathbf{k}'$ to $\mathbf{k}$, which is given by Fermi's golden rule in first order quantum mechanical perturbation theory. A high computational effort is required to solve the BTE numerically. The most common method to solve the BTE is the stochastic Monte Carlo method, which simulates the motion of individual carriers, subject to random scattering events. A deterministic solution, which has also been used to simulate realistic device structures, is based on the spherical harmonics expansion of the BTE [47, 11].

2.1.2. Method of Moments

Simpler transport models than the BTE, such as the drift-diffusion and the hydrodynamic model, are still desirable for the analysis and optimization of semiconductor devices. These models are based on balance equations which can be derived from the BTE. For this purpose, one considers moments of the distribution function f, which represent macroscopic quantities such as the carrier density n or the carrier's mean energy w. For any function $\Phi(\mathbf{k})$ of the momentum $\mathbf{k}$, the moment $\langle \Phi \rangle$ is defined as

$$\langle \Phi \rangle = \frac{1}{4\pi^3} \int \Phi(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) d^3 k. \tag{2.3}$$

A balance equation for the moment $\langle \Phi \rangle$ can be derived from the BTE. For this purpose, the BTE has to be multiplied with Φ and integrated over $\mathbf{k}$ space. This yields an equation of the form [48]

$$\frac{\partial}{\partial t} \langle \Phi \rangle + \nabla_{\mathbf{r}} \langle \mathbf{v} \otimes \Phi \rangle - \frac{q\mathbf{E}}{\hbar} \langle \nabla_{\mathbf{k}} \otimes \Phi \rangle = \left(\frac{\partial}{\partial t} \langle \Phi \rangle \right)_{coll}. \tag{2.4}$$

By a proper choice of $\Phi(\mathbf{k})$, the moments may represent the carrier density n, the momentum $\mathbf{p}$, the energy w and the energy flux $\mathbf{F}_w$:

$$n = \langle 1 \rangle = \frac{1}{4\pi^3} \int f(\mathbf{r}, \mathbf{k}, t) d^3k \tag{2.5}$$

$$\mathbf{p} = \langle \hbar\mathbf{k} \rangle = \frac{1}{4\pi^3} \int \hbar\mathbf{k} f(\mathbf{r}, \mathbf{k}, t) d^3k \tag{2.6}$$

$$w = \langle \mathcal{E}(\mathbf{k}) \rangle = \frac{1}{4\pi^3} \int \mathcal{E}(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) d^3k \tag{2.7}$$

$$\mathbf{F}_w = \langle \mathbf{v}\mathcal{E}(\mathbf{k}) \rangle = \frac{1}{4\pi^3} \int \mathbf{v}\mathcal{E}(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) d^3k \tag{2.8}$$

Here, $\mathcal{E}(\mathbf{k})$ is the electron energy and $\mathbf{v} = \hbar^{-1}\nabla_{\mathbf{k}}\mathcal{E}(\mathbf{k})$ is the electron group velocity. In the following an isotropic, parabolic band structure is assumed. The dispersion relation is then given by $\mathcal{E}(\mathbf{k}) = \hbar^2\mathbf{k}^2/(2m)$. Inserting (2.5) into (2.4), one obtains the balance equation for the carrier density as

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla\mathbf{j} = \left(\frac{\partial n}{\partial t}\right)_{coll}, \tag{2.9}$$

with the current $\mathbf{j} = -qn\langle\mathbf{v}\rangle = -qn\mathbf{v}_d$. The momentum balance equation reads:

$$\frac{\partial \mathbf{p}}{\partial t} + 2\nabla\hat{w} - nq\mathbf{E} = \left(\frac{\partial \mathbf{p}}{\partial t}\right)_{coll}, \tag{2.10}$$

with the energy tensor $\hat{w} = \langle\mathbf{v}\otimes\hbar\mathbf{k}\rangle/2 = m/2\langle\mathbf{v}\otimes\mathbf{v}\rangle$. This relation holds for parabolic bands. The carrier energy w is given by the trace of the energy tensor. The momentum equation (2.10) can be converted into a balance equation for the electron current j:

$$\frac{\partial \mathbf{j}}{\partial t} - \frac{2q}{m}\nabla\hat{w} - \frac{q^2n}{m}\mathbf{E} = \left(\frac{\partial \mathbf{j}}{\partial t}\right)_{coll} \tag{2.11}$$

The energy balance equation reads

$$\frac{\partial w}{\partial t} + \nabla\cdot\mathbf{F}_w - \mathbf{j}\cdot\mathbf{E} = \left(\frac{\partial w}{\partial t}\right)_{coll}. \tag{2.12}$$

Each equation derived by the method of moments is coupled to the balance equation of the next higher order moment. The continuity equation (2.9) for example contains the current $\mathbf{j}$. The balance equation of the current in turn depends on the carrier energy and so on. This leads to an infinite hierarchy of balance equations. For a practical simulation of the carrier transport however, a closed set of equations is required. In order to obtain a closed set of transport equations, an appropriate closure relation has to be introduced to truncate the infinite hierarchy. For example, a closure relation for the first three balance equations, (2.9), (2.11) and (2.14) has to express the energy flux F_w as a function of variables n, $\mathbf{j}$ and w. This leads to the hydrodynamic transport model.

2.1.3. Hydrodynamic Transport Model

Transport models that take into account the first three moments of the BTE are called hydrodynamic transport models due to their analogy to the Euler equations of fluid dynamics. Here, the closure relation of Bløtekjær [49], which relates the heat flow to the gradient of the carrier temperature, is used to derive the hydrodynamic transport model. For this purpose, the electron velocity is split in an average drift component and a random component due to collisions: $\mathbf{v} = \mathbf{v}_d + \mathbf{c}$. With this definition, the energy can be expressed as a sum of a drift energy and a thermal energy:

$$w = \frac{nm}{2}v_d^2 + \frac{nm}{2}\langle \mathbf{c}^2 \rangle. \tag{2.13}$$

Based on the temperature of an ideal gas, the carrier temperature T_C is defined by

$$\frac{nm}{2}\langle \mathbf{c}^2 \rangle = \frac{3}{2}nkT_C. \tag{2.14}$$

The corresponding energy tensor $\hat{w}$ is then given by [46]

$$\hat{w} = \frac{nm}{2}\langle \mathbf{v}_d \otimes \mathbf{v}_d \rangle + \frac{1}{2}nkT_C\hat{I}, \tag{2.15}$$

with the identity matrix $(\hat{I})_{ij} = \delta_{ij}$. Similarly, an expression for the energy flux $\mathbf{F}_w$ can be derived [45]:

$$\mathbf{F}_w = \langle \mathbf{v}\mathcal{E} \rangle \tag{2.16}$$
$$= \langle (\mathbf{v}_d + \mathbf{c})\mathcal{E} \rangle \tag{2.17}$$
$$= \mathbf{v}_d w + \frac{m}{2}\langle \mathbf{c}v^2 \rangle \tag{2.18}$$
$$= \mathbf{v}_d w + \frac{m}{2}\langle \mathbf{c}(v_d + c)^2 \rangle \tag{2.19}$$
$$= \mathbf{v}_d w + \frac{m}{2}(v_d^2\langle \mathbf{c} \rangle + 2v_d\langle \mathbf{c}^2 \rangle + \langle c^2\mathbf{c} \rangle) \tag{2.20}$$

The first component of the second term is zero by definition. With the definition of the carrier temperature from (2.14) and definition of the heat flux $\mathbf{Q}$ as

$$\mathbf{Q} = \frac{nm}{2}\langle c^2\mathbf{c} \rangle, \tag{2.21}$$

the energy flux can be written as

$$\mathbf{F}_w = w\mathbf{v}_d + nkT_c\mathbf{v}_d + \mathbf{Q}. \tag{2.22}$$

The first term accounts for energy transport by the motion of the carriers, the second term describes the work to move the carriers against the pressure of the electron gas and the last term describes the energy loss due to heat flow. With $\mathbf{F}_w$ given by (2.22), the momentum balance equation is still related to the moment of third order by the heat

flux $\mathbf{Q}$. To get a closed set of equations for the first three moments, the phenomenological closure relation $\mathbf{Q} = \kappa \nabla T_C$ is introduced. Thereby, the heat flow is related to the gradient of the carrier temperature by the thermal conductivity κ. Following the Wiedemann-Franz law, κ is given by

$$\kappa = \left(\frac{5}{2} - p\right) \frac{k^2}{q} n\mu T_C \tag{2.23}$$

with the fitting parameter p. The energy flux then becomes

$$\mathbf{F}_w = (w + nkT_C)\,\mathbf{v}_d - \kappa \nabla T_C. \tag{2.24}$$

If the contribution of the drift energy is neglected compared to the thermal energy in the first term on the rhs ($w \approx 3/2nkT_C$), this equation can be written as

$$\mathbf{F}_w = -\frac{5}{2}\frac{kT_C}{q}\mathbf{j} - \kappa \nabla T_C. \tag{2.25}$$

Furthermore, hydrodynamic device simulation requires suitable approximations for the collision terms. Therefore the collision terms are written in the relaxation time approximation. The collisions can be separated into inter-band and intra-band contributions [50]. Inter-band processes exchange carriers between the valence and conduction band and are usually described by generation and recombination rates. Intra-band collisions lead to a randomization of the carrier momentum and to energy transfer to the lattice. These effects can be described by a corresponding relaxation time. The collision terms of the first three moments can then be expressed as

$$\left(\frac{\partial n}{\partial t}\right)_{coll} = G - R, \tag{2.26}$$

$$\left(\frac{\partial \mathbf{p}}{\partial t}\right)_{coll} = -\frac{\mathbf{p}}{\tau_p}, \tag{2.27}$$

$$\left(\frac{\partial w}{\partial t}\right)_{coll} = -\frac{w - w_0}{\tau_w} + \frac{w}{n}(G - R), \tag{2.28}$$

with the net carrier generation rate $G - R$, the momentum relaxation time τ_p and the energy relaxation time τ_w. In order to cover the significant effects of the various scattering processes as good as possible in the device simulation, appropriate models for G, R, τ_p and τ_w are required. The dominating generation and recombination mechanisms in silicon transistors are trap-assisted Shockley-Read-Hall (SRH) recombination, Auger recombination at high carrier densities and impact ionization which occurs at high electric fields. Models which describe these effect as a function of the carrier density, energy and the electric field are contained in commercial device simulators [37]. Usually the carrier mobility μ is used for the formulation of the balance equations rather than the momentum relaxation time. The mobility μ is defined by $\mu = q\tau_p/m$. The mobility and energy relaxation time models used in this work are discussed in Section 2.2.

With (2.15) and (2.22) as well as (2.26)-(2.28) a closed set of balance equations can be obtained from (2.9), (2.12) and (2.22), which is referred to as the hydrodynamic (HD) transport model:

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla\mathbf{j} = G - R \tag{2.29}$$

$$\mathbf{j} - \tau_p\frac{\partial\mathbf{j}}{\partial t} + \tau_p\frac{2q}{m}\nabla\left(\frac{\mathbf{j}\otimes\mathbf{j}}{n}\right) = \mu\nabla(nkT_C) + nq\mu\mathbf{E} \tag{2.30}$$

$$\frac{\partial w}{\partial t} + \nabla\mathbf{F}_w - \mathbf{j}\mathbf{E} = \frac{w - w_0}{\tau_w} + \frac{w}{n}(G - R) \tag{2.31}$$

A further simplification of the HD transport model is obtained by neglecting the second and third term on the lhs of (2.30), which describes the convective current. The resulting set of equations is often referred to as energy transport (ET) model, however there is no consensus on nomenclature in the literature and often the terms *energy transport* and *hydrodynamic* are used synonymously. The full set of the HD/ET equations is given by:

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla\mathbf{j} = G - R \tag{2.32}$$

$$\mathbf{j} = \mu\nabla(nkT_C) + n\mu\nabla E_C - \mu\frac{3}{2}nkT_C\frac{\nabla m}{m} \tag{2.33}$$

$$\frac{3}{2}k\frac{\partial(nT_C)}{\partial t} + \nabla\mathbf{F}_w - \mathbf{j}\nabla E_C = \frac{3}{2}nk\frac{T_C - T_L}{\tau_w} + \frac{3}{2}kT_C(G - R) \tag{2.34}$$

$$\mathbf{F}_w = -\frac{5}{2}\frac{kT_C}{q}\mathbf{j} - \kappa\nabla T_C \tag{2.35}$$

In (2.33) an additional term has been added to account for a spacial variation of the effective mass as it occurs in heterostructures [45]. The variation of the effective mass acts as an additional driving force on the carriers. In this case, the term $-q\mathbf{E}$ has to be replaced by the effective force $\mathbf{F}$, which is given by

$$\mathbf{F} = -\nabla E_C + \frac{3}{2}kT_C\frac{\nabla m}{m}, \tag{2.36}$$

where E_C denotes the potential energy of the carriers.

2.1.4. Drift-Diffusion Model

The drift-diffusion model only takes into account the zeroth and first order moment of the BTE. It can be derived from the hydrodynamic equation by assuming that the carriers are in thermal equilibrium with the crystal lattice ($T_C = T_L = $ const.). From (2.32) and (2.33) one obtains

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla\mathbf{j} = G - R \tag{2.37}$$

$$\mathbf{j} = \mu kT\nabla(n) + n\mu\nabla E_C - \mu\frac{3}{2}nkT\frac{\nabla m}{m} \tag{2.38}$$

2.1.5. Poisson Equation

For a complete description of the semiconductor device, the transport model has to be coupled to Poisson's equation, which relates the electrostatic potential Φ to the charge density ρ.

$$\nabla \cdot \epsilon \nabla \Phi = q(n - p - N_D^+ + N_A^-) \tag{2.39}$$

Here, both the electron and hole density n and p, as well as the density of ionized donors and acceptors N_A^- and N_D^+ contribute to the net charge density. With the potential Φ the conduction and valence band energies are given by

$$E_C = -\chi - q\Phi \tag{2.40}$$
$$E_V = -\chi - E_g - q\Phi, \tag{2.41}$$

where χ denotes the electron affinity and E_g the bandgap energy. Here the vacuum energy is chosen as reference level.

2.1.6. Implementation in Device Simulators

Many variations of the hydrodynamic model have been developed, which treat the carrier energy in different ways [46]. The implementation of the hydrodynamic model in the device simulator *Sentaurus Device* includes some parameters which allow to adjust the equations to cover many of the models described in the literature. The current equation contains the parameter f^{td} which determines the influence of the thermal diffusion on the total current.

$$\mathbf{j} = \mu \left(k T_C \nabla n + f^{td} k n \nabla T_C + n \nabla E_C - \frac{3}{2} n k T_C \frac{\nabla m}{m} \right) \tag{2.42}$$

The thermal diffusion parameter f^{td} originates from the derivation of the hydrodynamic model by Stratton [51], which is based on the assumption that the distribution function is a heated Maxwellian. The equation for the current density derived by Stratton differs from the equation derived by Bløtekjær by the formulation of the diffusion term. After Strattons approach, the diffusion term is given by $k\nabla(n\mu T_C)$ instead of $k\mu\nabla(nT_C)$. The fact that the mobility is inside the gradient in the former case leads to an additional term in the current equation. The current equation after Stratton can be written as [46]

$$\mathbf{j} = \mu \left(q n \mathbf{E} + k T_C \nabla n + k n \left(1 + \nu_C \right) \nabla T_C \right) \tag{2.43}$$

with

$$\nu_C = \frac{T_C}{\mu} \frac{\partial \mu}{\partial T_C}. \tag{2.44}$$

The parameter $\nu_C = f^{td} - 1$ is usually used as a fit parameter with values between -0.5 and -1.

The energy flux contains two empirical parameters. The heat flux diffusion factor f^{hf} and the energy flux coefficient r. They can be used to change the convective and diffusive contributions to the total energy flux.

$$\mathbf{F}_w = -\frac{5}{2}r\left(\frac{kT_C}{q}\mathbf{j} + f^{hf}\frac{k^2}{q}n\mu T_C\nabla T_C\right) \tag{2.45}$$

With this notation, the empirical factor p of the Wiedemann-Franz law used in (2.23) is given by $p = 5/2\left(1 - rf^{hf}\right)$. The hydrodynamic model parameters f^{td}, f^{hf} and r can have a significant impact on the simulation results. A pragmatic approach to choose these parameters is to calibrate them to the results of Monte Carlo simulation or to experiments.

2.1.7. Carrier Statistics

Assuming a parabolic and isotropic band structure with the well known square root dependence of the density of states on the energy and a Fermi distribution of the carriers, the electron and hole densities are given by

$$n = N_C F_{1/2}\left(\eta_n\right), \tag{2.46}$$
$$p = N_V F_{1/2}\left(\eta_p\right), \tag{2.47}$$

with

$$\eta_n = \frac{E_{fn} - E_C}{kT}, \tag{2.48}$$
$$\eta_p = \frac{E_V - E_{fp}}{kT}. \tag{2.49}$$

N_C and N_V are the effective densities of states in the conduction and valence band respectively. E_{fn} is the quasi-Fermi energy for electrons and E_{fp} is the quasi-Fermi energy for holes. $F_{1/2}$ denotes the Fermi integral of order $1/2$.

For non-degenerate semiconductors, the Fermi distribution functions can be approximated by a Maxwell-Boltzmann distribution so that a closed form solution can be given for the carrier densities:

$$n = N_C \exp\left(\frac{E_{fn} - E_C}{kT}\right) \tag{2.50}$$
$$p = N_V \exp\left(\frac{E_V - E_{fp}}{kT}\right) \tag{2.51}$$

In the derivation of the balance equations above, a Boltzmann distribution of the carriers has been implicitly assumed by using $3/2nkT_C$ for the thermal energy of the carriers in (2.14). However, if Fermi statistics are applied, the thermal energy density u of the carriers is given by [52]

$$u = \frac{3}{2}nkT_C\gamma_C \tag{2.52}$$

with

$$\gamma_C = \frac{F_{3/2}(\eta_C)}{F_{1/2}(\eta_C)}. \tag{2.53}$$

If this definition is used in the derivation of the balance equations, one arrives at the following expression for the current [37]:

$$\mathbf{j} = \mu \left(kT_C \nabla n - nkT_C \nabla \ln(\gamma_C) + \xi_C f^{td} kn \nabla T_C + n \nabla E_C - \frac{3}{2} nkT_C \frac{\nabla m}{m} \right) \tag{2.54}$$

with

$$\xi_C = \frac{F_{1/2}(\eta_C)}{F_{-1/2}(\eta_C)}. \tag{2.55}$$

The corresponding energy flux is given by

$$\mathbf{F}_w = -\frac{5}{2} r \xi_C \left(\frac{kT_C}{q} \mathbf{j} + f^{hf} \frac{k^2}{q} n\mu T_C \nabla T_C \right). \tag{2.56}$$

The simulations in this work are performed using Boltzmann statistics.

2.2. Transport Parameters

Device simulation based on the drift-diffusion or hydrodynamic transport models requires specification of several parameters. The parameters can be divided into band-structure-related and scattering-related parameters. If the band structure is approximated by a single-valley parabolic band, the corresponding parameters are the bandgap E_g, the electron affinity χ and the effective masses of electrons m_e and holes m_h. Alternatively, one can use the valence and conduction band energies E_V and E_C instead of E_g and χ and the effective densities of states in the valence and conduction band N_V and N_C instead of the effective masses. The collision-related parameters are the electron and hole mobilities μ_e and μ_h as well as the energy relaxation times for electrons and holes, $\tau_{w,e}$ and $\tau_{w,h}$

In this section the models for the mobility, the energy relaxation time and the effective DOS used in this work are described. These models have been developed by Sasso et al. within the DOTFIVE project [8, 9, 53]. They are matched to the results of full-band Monte Carlo simulations.

2.2.1. Energy Relaxation Time

The model of the energy relaxation time used in this work is described in [8]. Similar to the model introduced by Gonzalez et al. [54], it describes τ_w as a function of the carrier and lattice temperatures:

$$\tau_w = \tau_{w,0} + \tau_{w,1} \cdot \exp\left[C_1 \cdot \left(\frac{T_C}{300\mathrm{K}} + C_0 \right)^2 + C_2 \cdot \left(\frac{T_C}{300\mathrm{K}} + C_0 \right) + C_3 \cdot \left(\frac{T_L}{300\mathrm{K}} \right) \right] \tag{2.57}$$

The composition dependence of the energy relaxation time in the $Si_{1-x}Ge_x$ alloy is included by composition dependent parameters $\tau_{w,0}$ and C_1:

$$\tau_{w,0} = \tau_{w,0,Si} \cdot (1 - x_n) + \tau_{w,0,Si_{0.7}Ge_{0.3}} \cdot x_n + C_\tau (1 - x_n) x_n \tag{2.58}$$

$$C_1 = C_{1,Si} \cdot (1 - x_n) + C_{1,Si_{0.7}Ge_{0.3}} \cdot x_n + C_c (1 - x_n) x_n \tag{2.59}$$

Here, $x_n = x/0.3$ is the normalized Ge mole fraction. The values of the electron energy relaxation time $\tau_{w,e}$, for parameters as in [8], are plotted in Figure 2.1 for different SiGe compositions. Monte Carlo simulation predicts a strong increase of τ_w for very low electron temperatures [54], which is not reproduced by the analytical model. This behavior can usually be neglected because for electron temperatures close to the lattice temperature, the term $(T_C - T_L)/\tau_w$ in (2.34) becomes very small. However, if electron temperatures which are significantly lower than T_L occur in the device, this model might be inadequate. The parameter values used in the simulation are given in Table A.1 in the appendix.

Figure 2.1: Electron energy relaxation time as a function of the normalized electron temperature for different Ge mole fractions x.
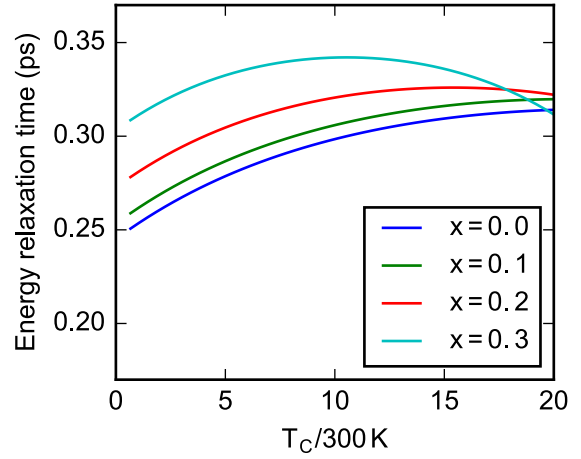
2.2.2. Mobility

The electron and hole mobilities μ_e and μ_h account for scattering processes during the motion of carriers. These parameters have a strong impact on the hydrodynamic simulation, because both current and energy flux are functions of the mobility. For this reason, an accurate description of the mobility is essential to obtain reasonable simulation results. Mobility models for device simulation are usually empirical models which are fitted to experiments or to the results of Monte Carlo simulation. These models are typically separated into a part which describes the low-field mobility and a part that accounts for high-field effects.

A model which describes the low-field mobility in doped silicon as a function of the doping concentration $N = N_A + N_D$ was given by Thomas and Caughey [55]:

$$\mu(T) = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + (N/N_{ref})^\alpha} \tag{2.60}$$

For $N \to 0$, this formula yields the lattice mobility μ_{max} which accounts for phonon scattering. For $N \gg N_{ref}$, when the mobility is governed by impurity scattering, it yields μ_{min}. The dependence of the lattice mobility on the temperature is usually modeled as

$$\mu_{max} = \mu_L \left(\frac{T}{300K} \right)^{-\gamma}, \tag{2.61}$$

where μ_L is the lattice mobility at $300\,\text{K}$. The low-field mobility model has to be combined with a high-field model to account for the effect of drift velocity saturation at high driving fields. The model given in [55] reads

$$\mu(F) = \frac{\mu_{low}}{\left[1 + (F/F_c)^\beta \right]^{1/\beta}}, \tag{2.62}$$

with the low-field mobility μ_{low} and the driving force F. For drift-diffusion simulations, one usually uses the electric field or the gradient of the quasi-Fermi potential as driving force F. In hydrodynamic simulations, the following expression [56, 37] is usually used:

$$F = \sqrt{\frac{3k(T_C - T_L)}{2q\tau_w\mu_{low}}} \tag{2.63}$$

The mobility model of Thomas and Caughey served as basis for a variety of extended mobility models. Based on comprehensive measurements on arsenic- and boron-doped samples, Masetti et al. [57] added an additional term to fit the observed drop of the mobility at very high doping concentrations. The model of Masetti was extended by Reggiani et al. [58]. They included a temperature dependence of the doping related parameters as well as an explicit functional dependence on the donor and acceptor concentration N_D and N_A. The latter allows to distinguish between the mobility of minority and majority carriers.

The mobility model used in this work was developed by Sasso et al. [8, 9] within the DOTFIVE project (in the following referred to as *dotfive model*). It is an extension of the model by Reggiani et al. which holds only for silicon. Based on Monte Carlo simulations it includes the dependence of the mobility on the germanium mole fraction x up to $x = 0.3$. The model equations are given in the following sections.

Low Field Mobility

The main equations of the low-field mobility model are borrowed from the Reggiani model.

$$\mu = \mu_0 + \frac{\mu_{max} - \mu_0}{1 + (N_D/C_{r1})^{\alpha_1} + (N_A/C_{r2})^{\alpha_2}} - \frac{\mu_1}{1 + (N_D/C_{s1} + N_A/C_{S2})^{-2}} \tag{2.64}$$

The lattice mobility is given by

$$\mu_{max} = \mu_{max,0} \left(\frac{T_L}{300\text{K}} \right)^\gamma. \tag{2.65}$$

2. Device Simulation Framework

In order to distinguish between majority and minority mobility, the parameters μ_0 and μ_1 are expressed as

$$\mu_0 = \frac{\mu_{0d} \cdot N_D + \mu_{0a} \cdot N_A}{N_D + N_A} \tag{2.66}$$

$$\mu_1 = \frac{\mu_{1d} \cdot N_D + \mu_{1a} \cdot N_A}{N_D + N_A}. \tag{2.67}$$

The dependence of the mobility on the germanium mole fraction x is modeled by interpolation between the silicon mobility μ_{Si} and the mobility in $\mathrm{Si}_{0.7}\mathrm{Ge}_{0.3}$ $\mu_{Si_{0.7}Ge_{0.3}}$:

$$\frac{1}{\mu_{SiGe}} = \frac{1 - x/0.3}{\mu_{Si}} + \frac{x/0.3}{\mu_{Si_{0.7}Ge_{0.3}}} + \frac{[1 - (x/0.3)^\alpha](x/0.3)^\alpha}{C_\mu} \tag{2.68}$$

The minority electron mobility as a function of the acceptor concentration is plotted in Fig. 2.2 for different temperatures as well as in Fig. 2.3 for different germanium fractions. The parameter values are given in Tab. A.2, Tab. A.3 and Tab. A.4.
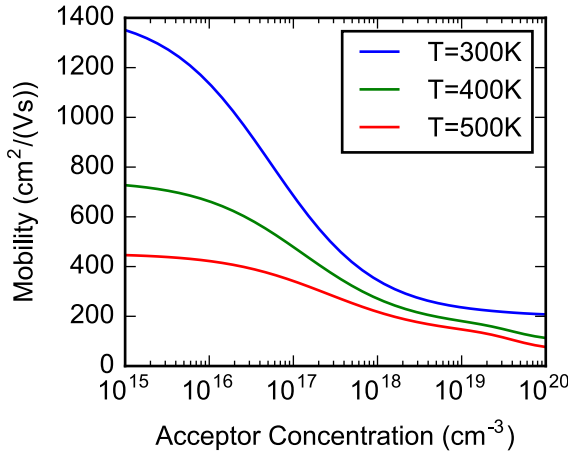

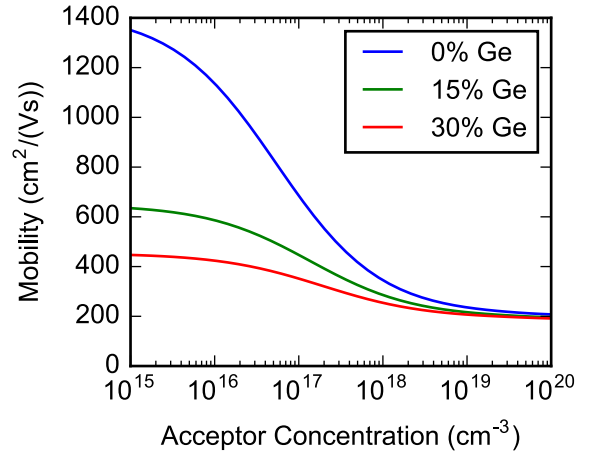**Figure 2.2.:** Minority electron mobility in silicon as function of acceptor doping concentration.

Figure 2.3.: Minority electron mobility for different germanium percentages ($T = 300\,\mathrm{K}$).

High Field Mobility

The high-field mobility model [9] that was developed along with the dotfive low-field mobility model cannot be implemented in Sentaurus Device, because the physical model interface of Sentaurus Device does not support the carrier temperature (or more specifically (2.63)) as a driving force for high-field mobility models. Instead, the standard model (2.62), which is implemented in the device simulator by default, is used here [37]. If the critical field strength F_c in (2.62) is replaced by v_{sat}/μ_{low} one obtains

$$\mu_{high} = \frac{\mu_{low}}{\left[1 + (\mu_{low}F/v_{sat})^\beta\right]^{\frac{1}{\beta}}}. \tag{2.69}$$

The exponent β depends on the temperature according to

$$\beta = \beta_0 \left(\frac{T}{300K}\right)^{\beta_{exp}}. \tag{2.70}$$

Similarly, the saturation velocity v_{sat} is modeled by

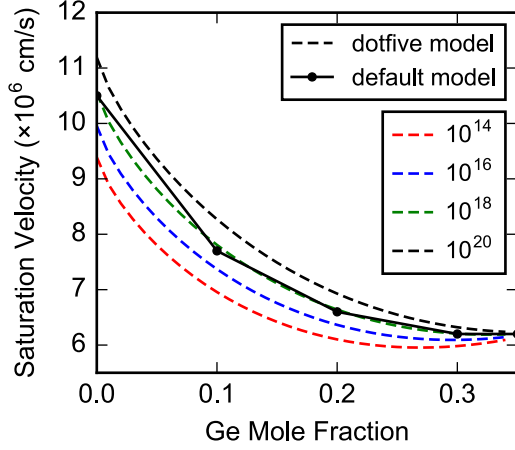$$v_{sat} = v_{sat,0} \left(\frac{300K}{T}\right)^{v_{sat,exp}}. \tag{2.71}$$

Figure 2.4.: Electron saturation velocity as a function of the Ge mole fraction ($T = 300$ K).

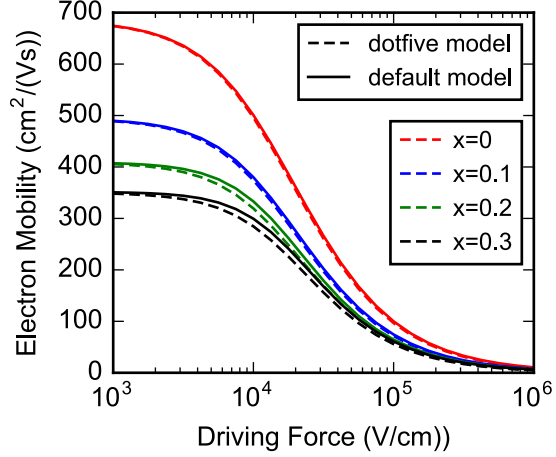**Figure 2.5.:** Field dependence of the minority electron mobility ($T = 300$ K, $N_A = 10^{17}$ cm^{-3}).

The model parameters were adjusted according to the model of Sasso et al. [9]. The dependence upon the germanium mole fraction is included by parameter tables which are interpolated linearly. With the fitted parameters, a good accordance of the default high-field mobility model with the dotfive model could be achieved, at least in the range of doping concentrations which are of interest for the simulation of SiGe HBTs. In Fig. 2.5, the electron mobility is shown as a function of field strength for different Ge mole fractions. However, the dotfive model includes the dependence of the saturation velocity on the the doping concentration, which is not included in the default model (see Fig. 2.4). Here, the parameters of the default saturation velocity model have been fitted to the values of the dotfive model for a doping concentration of 10^{18} cm^{-3}. The fitted parameter values are given in Tab. A.5.

2.2.3. Effective Density of States

The splitting of the valley degeneracy in strained SiGe modifies the effective density of states. To account for this effect, the effective DOS has to be modeled as a function of the Ge mole fraction x. Analytical models from [9] are used in this work. The effective

2. Device Simulation Framework

DOS in the SiGe conduction band is calculated by

$$N_C^{SiGe}(T, x) = N_C^{Si}(T_L)\frac{4 + 2\exp\left(\frac{\Delta E_C \cdot x}{kT}\right)}{4 + 2}, \tag{2.72}$$

where N_C^{Si} is the effective DOS in silicon. N_C^{SiGe} is modeled in accordance to the valley splitting and the relative occupation of the upper bands. $\Delta E_C \cdot x$ is the energy difference between the 4-fold degenerate lower band and the 2-fold degenerate upper band. A similar model is used for effective DOS in the valence band:

$$N_V^{SiGe}(T, x) = N_V^{Si}(T)\frac{1 + \exp\left(\frac{\Delta E_{V1} \cdot x}{kT}\right) + \exp\left(\frac{\Delta E_{V2} \cdot x}{kT}\right)}{1 + 1 + \exp\left(\frac{\Delta E_{V3} \cdot x}{kT}\right)} \tag{2.73}$$

The temperature dependence of the silicon effective DOS is modeled as

$$N_{\{C,V\}}^{Si} = N_{\{C,V\}}^{Si}(300\text{K})\left(\frac{T}{300\text{K}}\right)^{\alpha_{\{C,V\}}}, \tag{2.74}$$

using the exponents α_C and α_V as fit parameters. The result of both models are shown in Figure 2.6 for parameter values as given in Table A.6.

Figure 2.6: Effective density of states in the conduction and the valence band as a function of the Ge mole fraction.
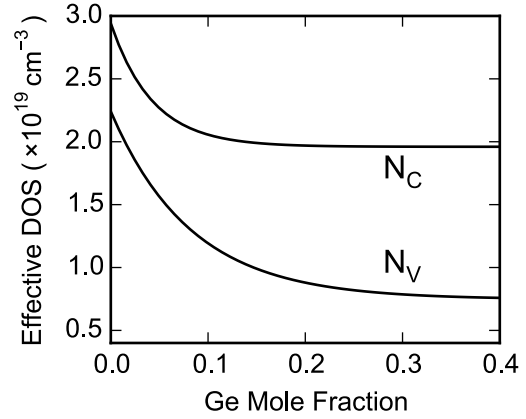
2.2.4. Hydrodynamic Model Parameters

Besides the parameters described above, also the hydrodynamic model parameters f^{hf}, f^{td} and r have a significant impact on the simulation results. A comprehensive investigation of the influence of the individual parameters has been carried out along with the development of the analytical transport models which are used here. In [59], different calibration strategies for the HD model parameters have been evaluated by a comparison of the HD simulation with results from device simulators based on the BTE. It turned out, that the best calibration strategy is to fix r and f^{td} to 1 and use f^{hf} as fitting parameter. Setting f^{td} to 1 corresponds to Bløtekjærs formulation of the current

equation. If the value of the parameter f^{hf} is too high, an unphysical negative output conductance and an exaggerated spurious velocity overshoot in the collector region can be observed. Thus, a sufficiently small value of f^{hf} has to be used to obtain reasonable simulation results. Optimal values of f^{hf} are given in [59] for three different reference transistors. These reference devices represent different HBT generations with a peak f_T of 100 GHz, 450 GHz and 700 GHz. The corresponding optimum values of f^{hf} are 0.2, 0.07 and 0.05, respectively.

A similar calibration of the HD parameters to BTE simulations has been performed by Wedel el al. [10]. They also used three referencs HBTs (100 GHz, 500 GHz and 1000 GHz), but only adjusted a single set of parameters in order to obtain a universal result for all technology generations. They obtained $r = 1$, $f^{td} = 1$ and $f^{hf} = 0.295$, which is similar to the results of Sasso et al.

In this work the calibration strategy suggested in [59] has been adopted. The parameters r and f^{td} are set to 1 and f^{hf} is used to fit the output conductance of the simulation to measured characteristics. A good agreement between measurement and simulation has been obtained for $f^{hf} = 0.4$ (cf. Section 3.4.1).

2.3. Calibration of the Effective Bandgap in SiGe

In this section, the bandgap parameters are calibrated against measurements of the collector current. In order to determine the dependence of the effective bandgap on the Ge mole fraction x, a set of dedicated HBTs with box-shaped base profiles and varying $Si_{1-x}Ge_x$ alloy compositions has been fabricated.

A relation between the effective bandgap $E_{g,eff}$ and the collector current J_C can be derived from the generalized Moll-Ross relation [60]. In the ideal region of operation, the collector current of a bipolar transistor with non-uniform base doping and bandgap is given by:

$$J_C = \frac{\exp\left(\frac{qV_{BE}}{kT}\right)}{\int_0^{w_b} \frac{p(x)}{kT\mu_n(x)n_{i,eff}^2(x)}dx} \tag{2.75}$$

The integral runs over the neutral base with width w_b. The effective intrinsic carrier density $n_{i,eff}$ is given by

$$n_{i,eff} = \sqrt{N_C N_V} \exp\left(\frac{-E_{g,eff}}{2kT}\right). \tag{2.76}$$

The HBTs which are used here have nearly uniform B and Ge concentrations in the base (see Fig. 2.10), so it can be assumed that both μ_n and $n_{i,eff}$ are constant. In this case, the collector current can be written as

$$J_C = -q\frac{\mu_n kT n_{i,eff}^2}{Q_B} \exp\left(\frac{qV_{BE}}{kT}\right), \tag{2.77}$$

with the total hole charge in the base

$$Q_B = -q\int_0^{w_b} p(x)dx. \tag{2.78}$$

23

Q_B can be approximated by SIMS measurements of the base doping or it can be determined from measurements of the base sheet resistance. Equation 2.77 shows, that measurements of J_C as a function of V_{BE} deliver the value of $\mu_n n_{i,eff}^2$. To determine the effective bandgap from such measurements, both the mobility and the effective DOS have to be known. Here, the models decribed above are used to calculate μ_n as well as N_C and N_V. In this way, consistent bandgap parameters for the simulation are obtained.

In order to model the effective bandgap in SiGe HBTs, two effects are usually considered: first, the reduction of the $Si_{1-x}Ge_x$ bandgap with the mole fraction x and second, the heavy-doping induced bandgap narrowing. One usually assumes that these effects can be separated. This means on one hand, that the doping induced bandgap narrowing $\Delta E_{g,dop}$ does not depend on the Ge mole fraction and on the other hand, that the bandgap reduction due to germanium $\Delta E_{g,Ge}$ is independent of doping. With the silicon bandgap $E_{g,Si}$, the effective bandgap in SiGe can then be written as:

$$E_{g,eff} = E_{g,Si} - \Delta E_{g,Ge} - \Delta E_{g,dop} \tag{2.79}$$

In the following, $\Delta E_{g,Ge}$ is determined experimentally while the values of $E_{g,Si}$ and $\Delta E_{g,dop}$ are taken from the literature. For $\Delta E_{g,dop}$ the results of Klaassen et al.[61] are used. These results have also been derived from measurements of the collector current using certain assumption on the mobility. To obtain values which are consitent with the mobility model used here, the values of Klaassen et al. are recalculated according to the following relation:

$$\mu_{old} \exp\left(\frac{\Delta E_{g,dop,old}}{kT}\right) = \mu_{new} \exp\left(\frac{\Delta E_{g,dop,new}}{kT}\right). \tag{2.80}$$

Both original and recalculated values of $\Delta E_{g,dop}$ are plotted in Fig. 2.7. Next, the vertical profiles of the reference transistors have to be determined.

Figure 2.7: Doping-induced bandgap narrowing. Original values from Klaassen et al. [61] and the recalculated values which are consitent with the dotfive mobility model.
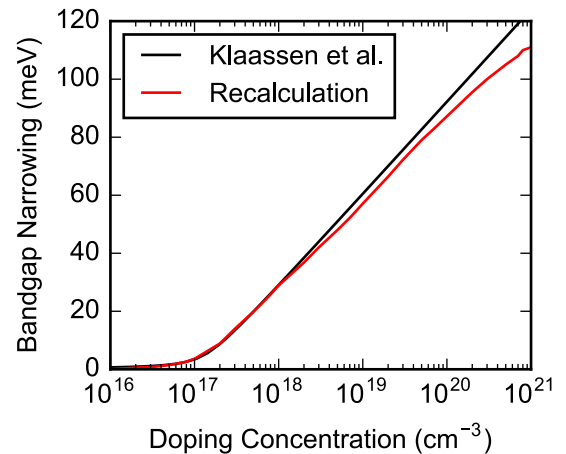
2.3.1. Characterization of the Box-Shaped Reference Profiles

For the calibration of the effective bandgap, a set of transistors with box-shaped boron and germanium profiles has been fabricated. The base layers of these transistors contain carbon at concentrations up to $10^{20}\,cm^3$. Carbon is introduced in the base to suppress boron diffusion. Additionally, substitutional carbon also influences the bandgap by a compensation of strain.

A thickness of 20 nm was targeted for the SiGe:C layers as well as a constant boron doping in the center of the base with 5 nm undoped SiGe:C spacers at each side. Four different SiGe:C alloy compositions with a Ge percentage of 0 %, 8.4 %, 21 % and 30.7 % were realized. A boron concentration of about $5 \times 10^{19}\,cm^{-3}$ was targeted for all variations. Table 2.1 summarizes the measured Ge percentages and doping concentrations of the four samples.

The transistors have been fabricated in a dedicated process flow, which is based on IHPs SG13G2 technology [3]. However, process steps not relevant for the HBT are omitted and only one metal layer is used. The process module of the HBT is also changed compared to the original process by fabricating the HBT without an elevated external base. In this way, the process complexity and the thermal budget can be reduced significantly. The reduced thermal budget minimizes dopant diffusion and helps to maintain the deposited Ge profile. A TEM cross section of a HBT fabricated in this process is shown in Figure 2.8.
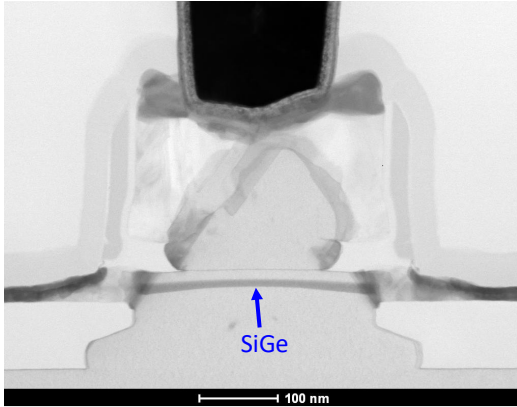

Figure 2.8.: TEM cross section of a SiGe HBT fabricated in the simplified process.
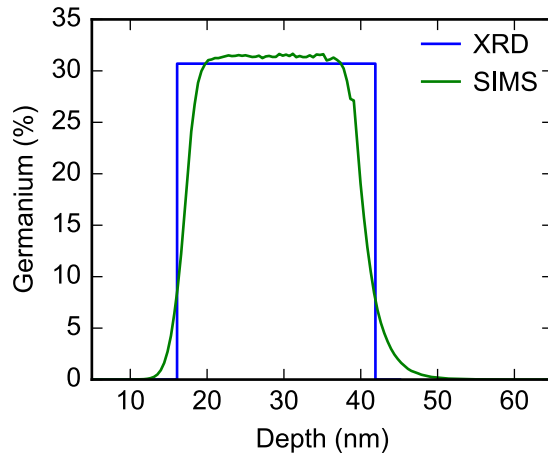
Figure 2.9.: Ge concentration from SIMS and X-ray diffraction.

An accurate characterization of the material composition and the doping profile in the base of the HBT is essential for the calibration of the effective bandgap. Two different methods have been used to determine the germanium concentration: X-ray diffraction (XRD) and secondary ion mass spectroscopy (SIMS) (see Appendix B for a short description of the methods). Basically, XRD is the most accurate method to determine the material composition because it directly measures the lattice strain. However, it is not possible to discriminate between germanium and carbon by measuring the strain. For

this reason, Ge profiles without carbon have also been deposited, which were used to determine the Ge concentration. Figure 2.9 shows Ge depth profiles determined by SIMS and X-ray diffraction. XRD cannot measure the depth profile directly. Assumptions on the shape of the profile have to be made to evaluate the measurements. Here, a simple box profile is assumed. The Ge concentrations resulting from SIMS deviate from the XRD result by less then $1\,\mathrm{at.\%}$.
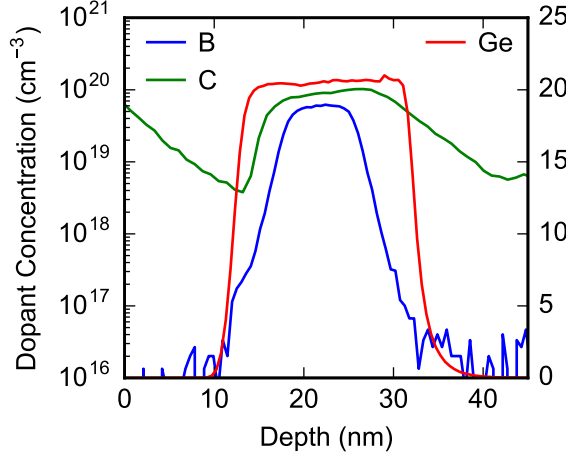

Figure 2.10.: Typical SIMS profile of the box-like SiGe base. This profile corresponds to wafer 16 from Tab. 2.1.
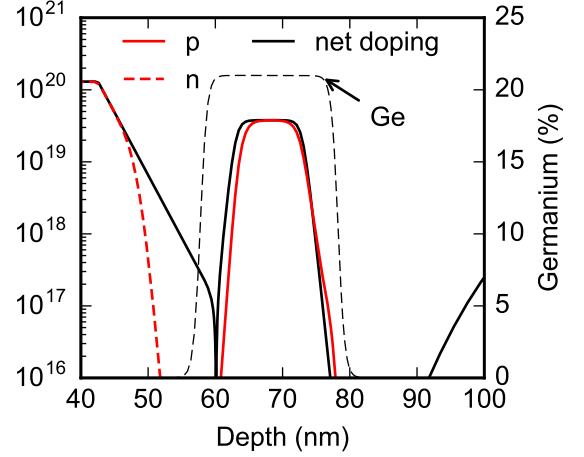
Figure 2.11.: Vertical transistor profile and simulated electron and hole density at zero bias.

A typical SIMS profile, measured after base epitaxy is shown in Figure 2.10. The Ge concentration from the SIMS measurement has been calibrated to the results of XRD. In contrast, it was not possible to obtain reliable results for the total boron dose from SIMS. A comprehensive calibration of the SIMS setup is necessary for a reliable quantification of the boron concentration in SiGe [62], which was not available at the time. For this reason, the hole charge in the base Q_b is determined from the measured base sheet resistance R_{sbi}, which is given by

$$R_{sbi} = \left(q \int_0^{w_b} \mu_p(x)p(x)dx \right)^{-1} \approx -\frac{1}{\mu_p Q_b}. \tag{2.81}$$

The hole mobility μ_p in the base is required to determine Q_b. It can be calculated by the model given in Sect. 2.2.2. However, the mobility itself depends on the boron concentration in the base. For this reason, Q_b cannot simply be determined from the measured R_{sbi} using (2.81). Instead, device simulation has been used to determine a consistent set of values for both the base charge Q_b and the boron concentration N_A. The vertical doping profile determined from SIMS has been used as starting point for the 1D simulation. For this purpose, the measured boron profile has been fitted by a double sigmoid function:

$$y(x) = A \cdot \left(1 + \exp\left[-\frac{x + w_1/2}{w_2} \right] \right)^{-1} \cdot \left(1 - \left(1 + \exp\left[-\frac{x - w_1/2}{w_2} \right] \right)^{-1} \right) \tag{2.82}$$

The simulation yields $\mu_p(x)$ and $p(x)$, which can be used to calculate the base sheet resistance of the simulated profile. Then, the boron concentration of the simulated profile has been changed iteratively, until the calculated sheet resistance equaled the measured R_{sbi}. The boron concentration has been adjusted via the parameter A in (2.82), while leaving the other parameters unchanged. This procedure yields consistent values for N_A and Q_B in accordance with the mobility model that is used here.

The vertical profile used in the simulation as well as the simulated electron and hole concentrations are plotted in Figure 2.11 for the HBT with 21 % Ge. The emitter and collector profiles have also been determined by SIMS. The widths of the base-emitter and base-collector junction have been adjusted to measured capacitances of large HBTs with an emitter area of $100 \times 50 \, \mu\mathrm{m}^2$. The base sheet resistance is determined using transistor tetrodes with different emitter widths (see Section 3.2.4 for a description of the tetrode structure). Measured values of R_{sbi} as well as the results of the boron concentration N_A and the hole charge Q_b are summarized in Table 2.1.

Table 2.1.: Parameters of wafer splits with different material composition in the HBT base.

Parameter	Unit	Method	Wafer 13	14	16	17
Ge	%	XRD	0	8.4	21	30.7
N_A	$10^{19}\mathrm{cm}^{-3}$	R_{sbi}	2.83	3.25	3.79	5.36
Q_b	$10^{13}\mathrm{cm}^{-2}$	R_{sbi}	2.03	2.58	3.17	4.83
Carbon	%	SIMS	0.03	0.1	0.22	0.3
R_{sbi}	kΩ	Tetrode	5.6	4.5	3.6	2.4

2.3.2. Extraction of the Effective Bandgap

The effective bandgap can be extracted from measurements of the collector current density J_C as a function of the base-emitter voltage V_{BE}. Measured Gummel plots for the different alloy compositions are shown in Fig. 2.12. Relatively large transistors with an emitter area of $5 \times 5 \, \mu\mathrm{m}^2$ have been evaluated here, in order to reduce perimeter effects. The collector current shows an ideal behavior up to a V_{BE} of about 0.7 V for all alloy compositions. The base current is also shown in Fig. 2.12. At low and medium V_{BE}, I_B increases with higher Ge, which can be explained by enhanced recombination in the neutral base [63].

In order to determine the effective bandgap, the saturation current J_{C0} is extracted from the ideal region of the Gummel plot. Assuming an idealiy factor of 1.0, J_{C0} is given by

$$J_{C0} = J_C \exp\left(\frac{-qV_{BE}}{kT}\right). \tag{2.83}$$

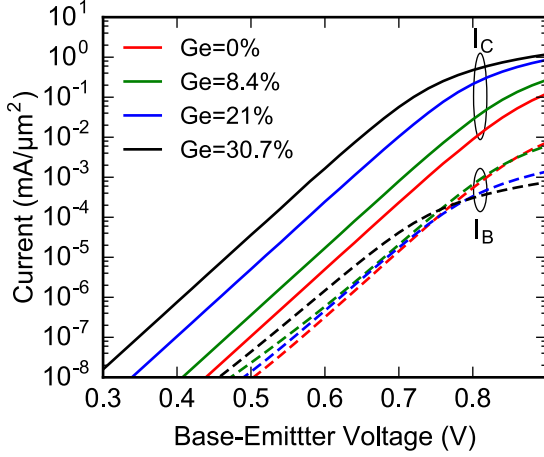Combining this relation with (2.77) and (2.76), the effective bandgap can be expressed

Figure 2.12.: Gummel characteristics for different Ge percentages. Solid lines represent the collector current and dashed lines represent the base current.
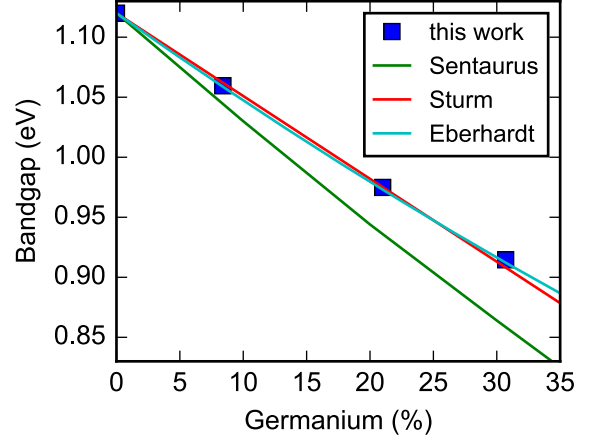
Figure 2.13.: Bandgap as a function of the germanium mole fraction. Symbols show data obtained in this work. Lines show values from Sturm [64] and Eberhardt [44] as well as default values from Sentaurus Device.

as a function of the saturation current:

$$E_{g,eff} = -kT \log \left(\frac{-q\mu_n kT N_C N_V}{J_{C0} Q_B} \right). \tag{2.84}$$

The bandgap reduction due to Ge can be determined from the difference of the effective bandgap of the Si BJT and the effective bandgap of the SiGe HBT. In this case, one has to consider that the transistors have different doping concentrations in the base, which lead to a different doping induced bandgap narrowing:

$$\Delta E_{g,Ge} = (E_{g,eff}^{Si} + \Delta E_{g,dop}^{Si}) - (E_{g,eff}^{SiGe} + \Delta E_{g,dop}^{SiGe}) \tag{2.85}$$

Inserting (2.84) into this equation, one obtaines

$$\Delta E_{g,Ge} = -kT \log \left(\frac{\mu_n^{Si}}{\mu_n^{SiGe}} \frac{N_C^{Si} N_V^{Si}}{N_C^{SiGe} N_V^{SiGe}} \frac{Q_b^{SiGe}}{Q_b^{Si}} \frac{J_{C0}^{SiGe}}{J_{C0}^{Si}} \right) + \Delta E_{g,dop}^{Si} - \Delta E_{g,dop}^{SiGe}. \tag{2.86}$$

The mobility and the effective DOS product are calculated using the models described in sections 2.2.2 and 2.2.3. The total base charge is given in Table 2.1. The results are presented in Figure 2.13, where $E_g = 1.12\text{eV} - \Delta E_{g,Ge}$ is plotted as a function of the Ge percentage together with data from the literature.

In a review of Sturm [64] the published SiGe bandgap data available at that time is evaluated. A linear fit to the data shows a bandgap reduction of $6.9\,\text{meV}$ per $1\,\%$ Ge. Sturm used data of the minority electron mobility in Si from [65] for the mobility in SiGe. This is a reasonable approximation at high doping levels, where the minority electron mobility is only weakly dependent on the Ge content (cf. Fig. 2.2). For the ratio of the effective densities of state in Si and SiGe $(N_C N_V)^{SiGe}/(N_C N_V)^{Si}$ the model of

[66] is used. Additional data on the SiGe effective bandgap was published by Eberhardt et al. [44]. They analyzed HBTs with germanium contents ranging from 15.4 % to 28.4 % and a base doping concentration of about $5 \times 10^{19}\,\mathrm{cm}^{-3}$. Using electron mobility values calculated by Bufler [67] and a constant ratio of 0.4 between the effective DOS in SiGe and Si, they obtained a bandgap reduction similar to the results of Sturm. A linear fit to the results of this work provides a bandgap reduction of 6.8 meV/%, which is in good agreement with the results of Sturm and Eberhardt et al. The default bandgap values from the parameter file for SiGe HBTs of Sentaurus Device are also plotted in Fig. 2.13. These values are calculated according to the theoretical work of Martin and Van de Walle [68]. The use of these values would lead to a strong overestimation of the collector current.

The HBTs which are used here contain up to 0.3 % carbon. Substitutional carbon in the SiGe base compensates strain and increases the bandgap. Bandgap changes of 20-30 meV/%C have been reported [69, 18, 70]. In order to investigate the impact of carbon on the effective bandgap, carbon-free transistors with the same germanium and doping profiles as the samples in Table 2.1 have been fabricated. However, no significant impact on the collector current was observed. For this reason, the impact of carbon on E_g is neglected in the following.

2.3.3. Temperature Dependence of the Effective Bandgap

The temperature dependence of the bandgap is usually described by the empirical equation of Varshni [71]:

$$E_g(T) = E_g(0K) - \frac{\alpha T^2}{\beta + T}. \tag{2.87}$$

The parameter values for silicon are $\alpha = 4.73 \times 10^{-4}\,\mathrm{eV/K}$ and $\beta = 636\,\mathrm{K}$. Typically, these values are also used to model the temperature dependence of the SiGe bandgap. Optical absorption measurements by Braunstein et. al. [72] support this approach. However, these measurements were done at unstrained bulk SiGe and only up to a temperature of 300 K. Therefore, it is also necessary to calibrate the temperature dependence of the effective bandgap. For this purpose, Gummel plots at various temperatures between 233 K and 473 K have been measured as shown in Figure 2.14. This temperature range corresponds to the valid range of the parameter models.

Experimental values of the effective bandgap are shown in Fig. 2.15 as a function of the temperature. The solid lines represent $E_g(T)$ as given by the Varshni model using the silicon values of α and β. $E_g(0\,\mathrm{K})$ is set in such a way that the curve agrees with the measured values at 300 K. In case of 0 % and 8.4 % Ge, there is a good agreement between the Varshni model and the experimental results at 293 K, 353 K and 413 K. The measured bandgaps at 233 K and 473 K are larger then predicted by the Varshni model. This is an indication that the temperature dependence of the parameters (μ, $N_{C/V}$) which are used for the extraction of the bandgap is not described accurately over the whole temperature range. However, in the temperature range relevant for most (room temperature) applications, there is no such discrepancy.
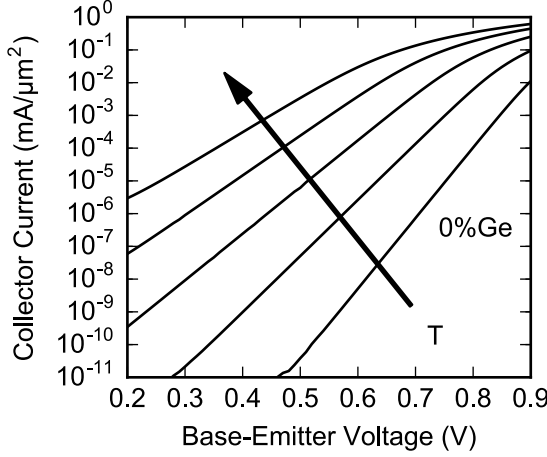
Figure 2.14.: Collector current of the all-silicon transistor measured at 233 K, 293 K, 353 K, 413 K and 473 K.
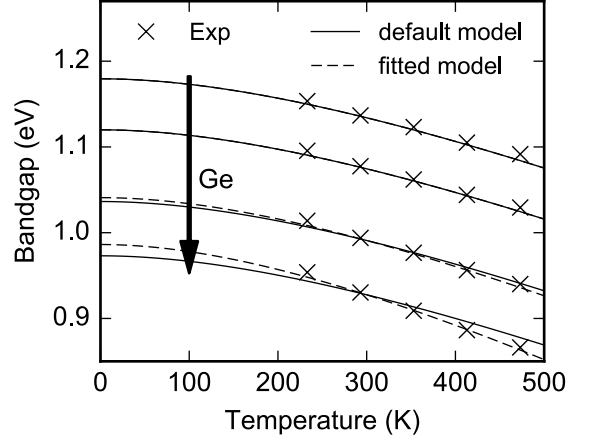

Figure 2.15.: Temperature dependence of the effective bandgap for 0 %, 8.4 %, 21 % and 30.7 % germanium.

In the case of 21 % and 30.7 % Ge, the measured effective bandgap decreases stronger with increasing temperature than in case of the Si BJT. The parameters of the Varshni model have to be changed to obtain a better agreement with the measurements. A good fit of the data is achieved with $\alpha = 5.2 \times 10^{-4}$ eV/K for 21 % Ge and with $\alpha = 6.1 \times 10^{-4}$ eV/K for 30.7 % Ge, except for the bandgap at $T = 233$ K. The fitted curves are plotted in Fig. 2.15 as dotted lines. A better fit of the measured bandgap can be achieved by a linear dependence ($\beta = 0$ K), however this would lead to strong overestimation of E_g at lower temperatures.

The generalized Moll-Ross relation has been used here to determine the effective bandgap. It is derived from the drift-diffusion model of the current. So, the effective bandgap, which is determined on the basis of (2.75), is strictly valid only for drift-diffusion simulations. In the hydrodynamic transport model, thermal diffusion additionally contributes to the current so that the simulated I_C also depends on the hydrodynamic model parameters f^{td}, f^{hf} and r. Therefore the validity of the parameters for HD simulations has to be verified by a comparison of measured and simulated gummel characteristics. Figure 2.16 shows the measured collector current and the results of hydrodynamic simulations which use the bandgap parameters that have been determined in this section as well as the parameter models described in the previous sections. The calibrated bandgap leads to a good agreement between simulation and measurement. The deviation at high V_{BE} is a result of series resistance which has not been considered in this simulation.

2.4. Summary

In Chapter 2, the hydrodynamic model framework which is used for the simulation of SiGe HBTs has been described. The derivation of the hydrodynamic transport model has been outlined and the models of the transport parameters, which have been de-
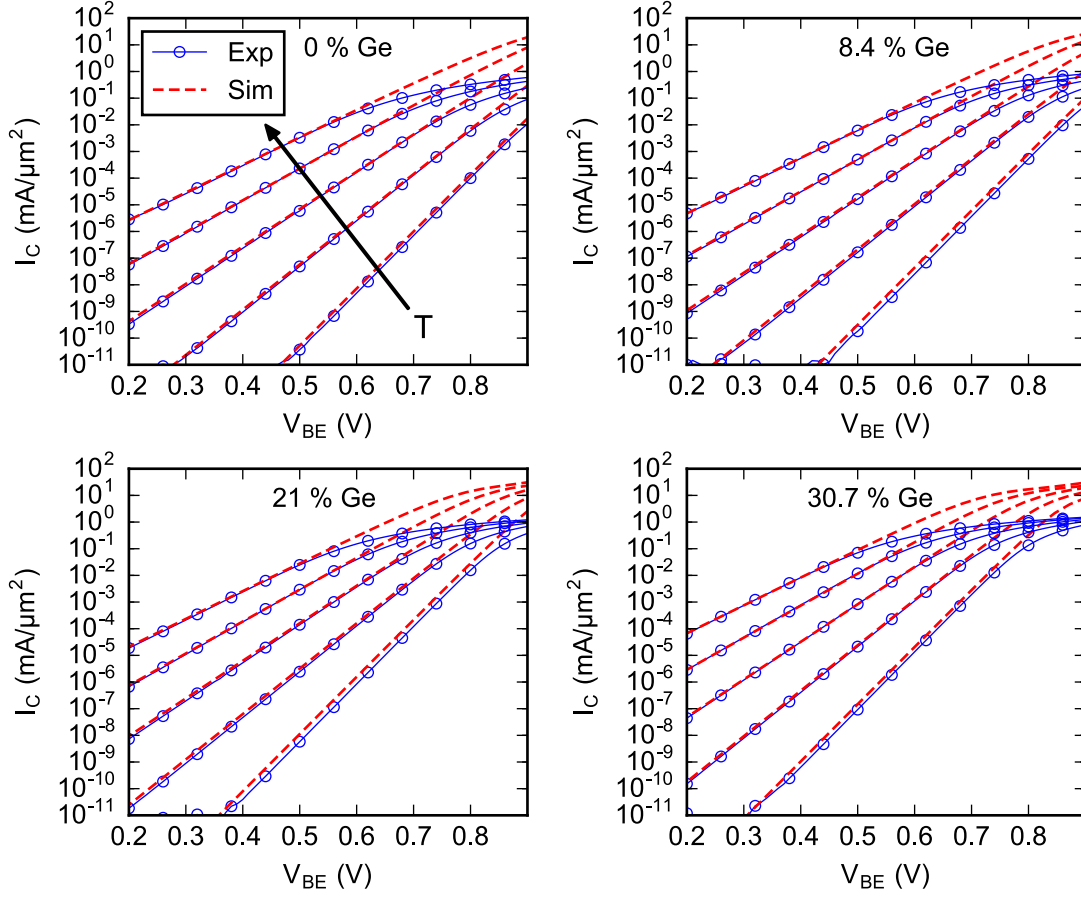
Figure 2.16.: Measured and simulated collector current as a function of V_{EB} for the four wafer splits with different Ge fraction. The current is plotted for a temperatures of 233 K, 293 K, 353 K, 413 K and 473 K.

veloped within the DOTFIVE project, are presented. The effective bandgap has been determined experimentally from the collector current of dedicated reference transistors with various SiGe:C alloy compositions. The results agree well with published data from SiGe HBTs. A clear impact of the carbon on the effective bandgap could not be found for the investigated devices. So, the published bandgap values for SiGe HBTs can also be used to simulate SiGe:C HBTs with typical carbon concentrations of about $10^{20}\,\mathrm{cm}^{-3}$.

3. Comparison of 2D Simulation and Experiment

The TCAD based roadmap for SiGe HBTs by Schröter et al. [73], which is also part of the 2013 ITRS roadmap, suggests that f_T values of more than $1\,\mathrm{THz}$ may be achieved with vertically scaled doping profiles and reduced paracitics. However, the reliability of such predictions has not yet been verified. There are essentially two conditions for a realistic estimation of the performance of future SiGe HBT generations. On the one hand, realistic assumptions for process specific parameters of future technologies have to be made. Examples include the achievable steepness of doping profiles and the possible reduction of the contact resistivity with advanced processes. On the other hand, the simulation tools that are used to predict the performance of hypothetical devices have to be sufficiently accurate. They must capture all relevant physical effects that determine the performance of advanced SiGe HBTs.

In order to evaluate the accuracy of state-of-the-art simulation of SiGe HBTs, a comparison of measured and simulated electrical characteristics of transistors with an advanced vertical doping profile is presented in this chapter. For a quantitative comparison between measurement and simulation, a realistic 2D model of the device which comprises all relevant parasitics has to be used. Such simulations are not practicable with device simulation based on Boltzmann's transport equation (MC, SHE) due to their high computational expense. Hydrodynamic simulations with calibrated parameter models, however, are well suited for this task because they are computationally less expensive. Therefore, the hydrodynamic model is still the workhorse for technology development.

Another important prerequisite for a meaningful comparison of simulation and measurement is that the transistors which are used in such a study are well characterized. Precise knowledge of the vertical doping profile is particularly important to draw conclusions on the accuracy of device simulation. Vertical doping profiles for device simulation are usually derived from SIMS profiles or from process simulation. These methods are generally not accurate enough to decide, if discrepancies between measurement and simulation are caused by errors in the doping profile or by an insufficient description of the carrier transport. The usual way, to deal with the lack of precise knowledge on the vertical doping profile, is to adjust the profile to electrical measurements [74]. Obviously, this method is not the best option if the accuracy of device simulation shall be evaluated by a comparison of measurement and simulation. In this work, a combination of different experimental techniques has been used to determine the vertical doping profile with the best available accuracy. The experimental techniques include secondary ion mass spectroscopy (SIMS), energy dispersive X-ray spectroscopy (EDX) in a transmission electron microscope (TEM) and X-ray diffractometry (XRD).

This chapter is organized as follows: Section 3.1 deals with the experimental characterization of the HBTs. In Section 3.2, the definition of the 2D simulation domain is presented. In Section 3.3, the calculation of the figures of merit f_T and f_{max} is described and in Section 3.4, the measured and simulated electrical characteristics are compared and analyzed.

3.1. Experimental Characterization of the Reference Transistors

HBTs with a scaled vertical doping profile are used for the comparison of simulation and experiment. They were fabricated in an experimental process based on IHPs SG13G2 technology [3]. This preparation aimed at realizing a vertical doping profile that was suggested by a TCAD based optimization (Node N3 from the SiGe HBT roadmap) [75]. The process flow in this experiment was simplified by skipping the steps which are only necessary for CMOS and by reducing the number of metal layers from 5 to 3. Further, the width of the base-emitter spacers and the minimal width of the emitter window were increased compared to the original process. This helps to realize homogeneous doping profiles over the whole emitter width. Relaxed lateral dimension lead to a reduction of f_{max}, however the focus of this study is on the impact of the vertical profile on f_T. The peak temperature of the final rapid thermal annealing (RTA) was reduced to 1030 °C in order to minimize broadening of the deposited doping profiles. The structure of the device is shown in a TEM cross section in Figure 3.1
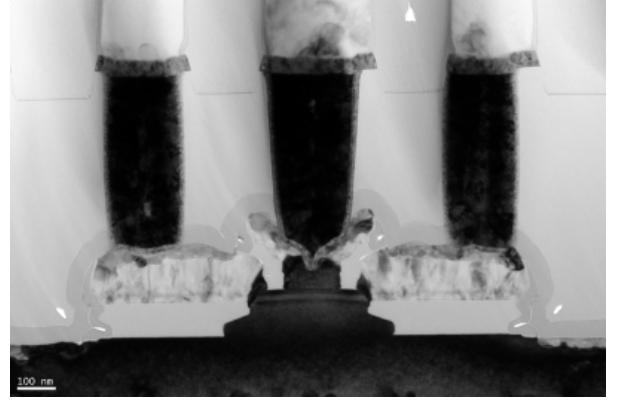

Figure 3.1: TEM cross section of the HBT.

3.1.1. Electrical Characteristics

DC and RF measurements have been performed at different transistor layouts. The highest f_T achieved in this preparation is 430 GHz [76]. This value has been measured at RF transistors with a BEC layout and eight emitters in parallel. The emitters have a length of 1.01 µm and a width of 0.17 µm. In the following, this device is denoted as D0. The BEC layout configuration is shown in Figure 3.2. Base and collector contacts are at

opposite ends of the emitter, which minimizes the overlap between the base polysilicon and the collector resulting in a reduction of C_{BC}.
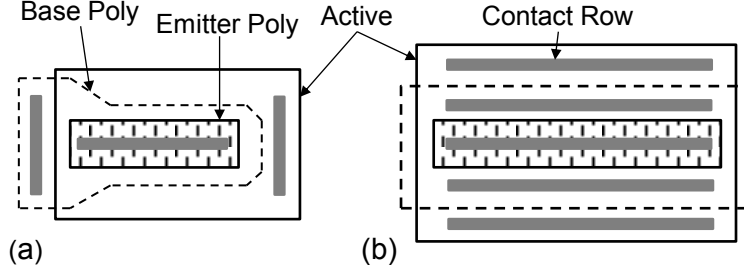

Figure 3.2.: Different transistor layouts: (a) BEC layout with minimal overlap between base poly and active. (b) Symmetrical CBEBC layout which is well sited for comparison with 2D simulation.

For a comparison with 2D device simulation, the symmetrical CBEBC layout with base and collector contacts parallel to the emitter is better suited. Here, three devices with such a layout configuration and different emitter geometries were available. These devices, in the following denoted as D1, D2 and D3, have an emitter length of $5\,\mu$m and a width of $0.14\,\mu$m, $0.28\,\mu$m and $0.41\,\mu$m, respectively. Due to the fact that the length of these transistors is much larger than their width, the effect of the short edges on the overall operation of the transistor can safely be neglected. As a consequence, electron transport in these transistors is assumed to be accurately described by a 2D simulation.
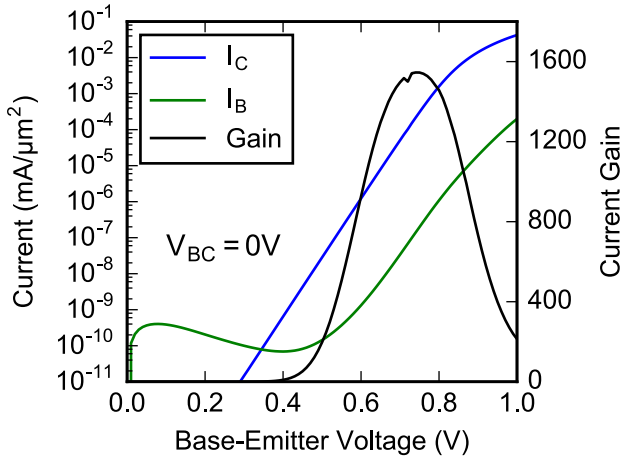

Figure 3.3.: Measured Gummel characteristics of device D0. Non-ideal base current is mainly caused by band-to-band tunneling.
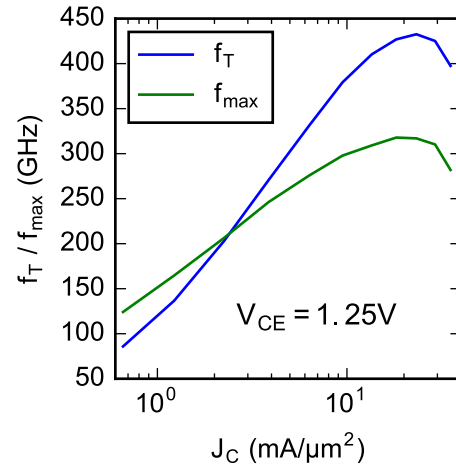
Figure 3.4.: f_T and f_{max} as a function of the collector current density.

Measured Gummel characteristics of the device D0 are shown in Fig. 3.3. The non-ideal base current at low V_{BE} is caused by band-to-band tunneling at the base-emitter junction. A narrow base-emitter junction helps to reduce the base-emitter transit time while, on the other hand, it leads to an increase of band-to-band tunneling. The measured transit frequency f_T and the maximum oscillation frequency f_{max} of D0 are plotted

in Fig. 3.4 showing peak values of 430 GHz and 315 GHz, respectively. Here, f_T and f_{max} have been determined from extrapolation of the small-signal current gain h_{21} and the unilateral gain U, respectively, at a frequency of 40 GHz (cf. Section 3.3).

Further electrical parameters of devices D0 to D3 are summarized in Table 3.1. The breakdown voltages of emitter-base and base-collector junctions (BV_{EBO} and BV_{CBO}) as well as the collector-emitter breakdown voltage (BV_{CEO}) are still high enough for many applications. The base sheet resistance R_{sbi} is about 2.7 kΩ, which is sufficiently low for achieving high f_{max} values with reduced lateral device dimensions.

Table 3.1.: Measured electrical parameters of the HBTs.

Parameter	Method	D0	D1	D2	D3
Layout		8×BEC	CBEBC	CBEBC	CBEBC
W_E (μm)		0.17	0.14	0.28	0.41
L_E (μm)		1.01	5	5	5
peak f_T (GHz)	$V_{CE} = 1.25$V	430	371	404	395
peak f_{max} (GHz)	$V_{CE} = 1.25$V	315	278	233	195
j_C (mA/μm^2)	at peak f_T	22	19	16	15
R_{TH} (K/W)	as in [77]		3100	2900	2700
β	$V_{BE} = 0.7$V		1550		
BV_{EBO} (V)	$I_E = 9\,\mu$A/μm^2		1.0		
BV_{CBO} (V)	$I_C = 1\,\mu$A/μm^2		4.0		
BV_{CEO} (V)	I_B reversal		1.45		

3.1.2. Vertical Base Profile

An accurate characterization of the doping profile and the material composition in the base layer is essential for a quantitative comparison of device simulation and experiment. A combination of various experimental techniques has been used to analyze the vertical doping profile. SIMS measurements were performed at IHP and by Evans Analytical Group (EAG) to reduce uncertainties related to the SIMS analysis. Additionally, the germanium profile in the base has been analyzed by energy dispersive X-ray spectroscopy (EDX) in a transmission electron microscope and by X-ray diffractometry (XRD). A short description of the methods is given in Appendix B.

The measured Ge and B profiles of an as-grown HBT structure are shown in Fig. 3.5. Both SIMS measurements from IHP and EAG are shown together with the EDX results of the Ge profile. The depth scale of the IHP SIMS measurements has been fitted to the EDX results. EDX in a TEM provides a direct measurement of the depth scale with a high resolution, while dynamic SIMS records the count rate of secondary ions as a function of time, which has to be converted into a depth profile. The adjustment of the SIMS depth scale has been done by fitting the sputter rate: In order to convert the time axis into depth, the final depth of the sputter crater is measured by a profilometer after
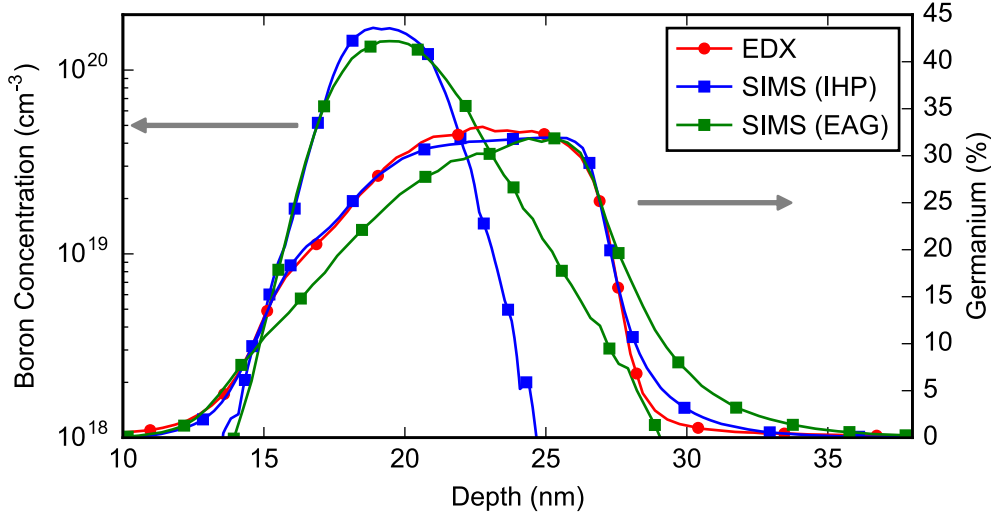
Figure 3.5.: Depth profile of the base doping. SIMS measurements of boron and germanium from IHP (blue, square) and EAG (green, square) as well as EDX profile of germanium (red, circle).

the SIMS analysis. The average sputter rate is then simply given by the crater depth divided by the total sputtering time. However, the sputter rate generally depends on the material composition of the sample. This effect must be considered when converting the ion signal over time into a depth profile. Here, a linear dependence of the $Si_{1-x}Ge_x$ sputter rate r on the Ge fraction x was assumed to calculate the depth scale:

$$r(x) = r_0 \cdot (1 + a \cdot x) \tag{3.1}$$

The sputter rate correction factor a was then used to fit the depth scale of the SIMS measurement to the EDX depth profile. The shape of the EDX profile and the SIMS profile from IHP are almost identical, as can be seen in Figure 3.5. Here, an agreement between SIMS and EDX was achieved with a correction factor of $a = 0.35$.

In order to achieve an accurate quantification of the germanium concentration, the EDX and SIMS measurements of the Ge profile were calibrated against the results of XRD [78]. Samples with a simple box-like Ge profile without any boron or carbon doping have been used for the calibration. At such samples XRD allows an accurate determination of the germanium mole percentage with an approximate absolute error of $\pm 0.3\%$ [79].

The Ge profile measured at EAG has a smaller gradient and a longer tail but the total Ge dose is nearly the same. The reason for the different shapes is that SIMS at IHP was performed using oxygen as primary ions with an energy of $0.5\,\mathrm{keV}$, whereas SIMS at EAG was carried out using $2\,\mathrm{keV}$ caesium sputtering for the Ge profile. In general, sputtering at higher energies leads to enhanced profile broadening.

The boron profiles were both measured with O_2^+ sputtering with an energy of $0.5\,\mathrm{keV}$ at IHP and $1\,\mathrm{keV}$ at EAG. The EAG profile shows a stronger broadening at the trailing

edge compared to the IHP profile. The total boron dose of the EAG profile is about $7.2 \times 10^{13}\,\mathrm{cm^{-2}}$, the IHP measurement delivered a dose of $5.2 \times 10^{13}\,\mathrm{cm^{-2}}$. It can be assumed that the dose measured by EAG is more accurate, because their proprietary SIMS analysis (PCOR-SIMS) includes a so called point-by-point correction for the relative sensitivity factors (RSF). The relative sensitivity factor relates the secondary ion intensities of the matrix material and the element to be determined, I_M and I_E to their respective concentrations C_M and C_E:

$$\frac{I_M}{C_M} = RSF_E \cdot \frac{I_E}{C_E} \tag{3.2}$$

The SIMS analysis available at IHP assumed a silicon matrix and a constant RSF for the quantification of the boron concentration. This can introduces a considerable error in the measured boron concentration in the base, where the real matrix material is SiGe with up to 30 % germanium. The EAG point-by-point correction claims to take into account this effect. Therefore it can be expected that the boron dose from the EAG measurement is more accurate than the IHP result. On the other hand, one can assume that the boron profile from IHP reflects the real width of the base more accurately than the EAG profile, due to the lower sputtering energy.
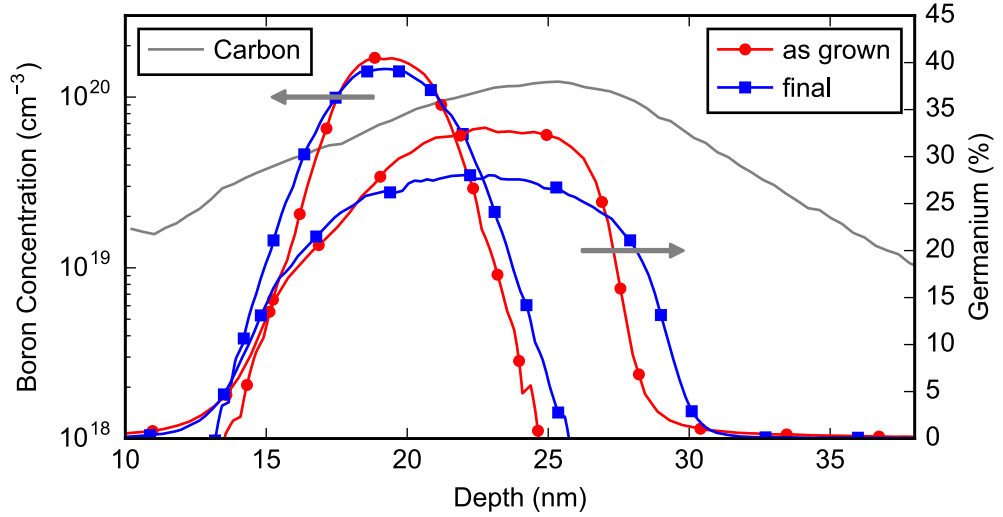

Figure 3.6.: Influence of the thermal budget on the base profile. As-grown (red, circles) and final (blue, squares) boron and germanium profile. Boron profile is obtained by SIMS, germanium profile by EDX. Grey line shows carbon concentration from SIMS.

Dopant diffusion due to thermal processing leads to a broadening of the doping profile. In Fig. 3.6, the final profile which has been measured after emitter deposition and final annealing, is compared to the as-grown profile, which has been measured before the emitter formation. Only minor broadening of the B profile can be observed due to the suppression of boron diffusion by both germanium and carbon. On the other hand, a considerable broadening of the Ge profile occurs, which leads to a reduction of the Ge peak concentration from 32 % to 28 %.

Another important effect has to be considered to obtain the correct vertical profile of the HBTs. On a patterned wafer, the thickness of epitaxial layers depends on the size of the patterned structure. SIMS measurements require a relatively large sample area compared to the typical area of a HBT . Here, a window of $400\,\mu\mathrm{m}^2 \times 600\,\mu\mathrm{m}^2$ has been used for SIMS. The Ge depth profile has been measured by EDX, both in the SIMS monitor and in a typical HBT structure. As can be seen in Figure 3.7, the width of the SiGe layer in the HBT is 14 % smaller than in the large SIMS monitor. It is assumed here, that the boron profile in the HBT structure scales in the same way and can thus be obtained from the SIMS profile by shrinking the depth scale by 14 %.
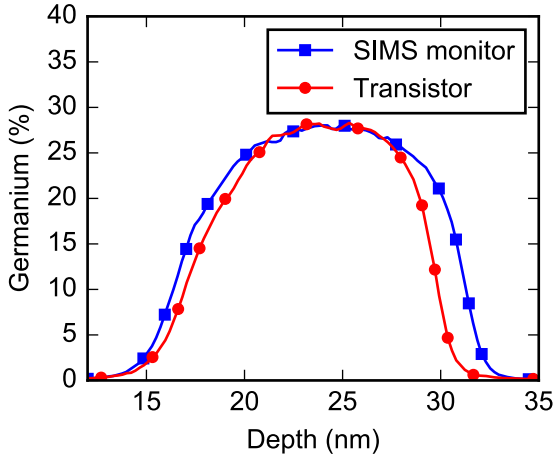

Figure 3.7: EDX line scan over the SiGe layer at a sample from the transistor (red, circles) and a sample from the large SIMS monitor (blue, squares).

3.2. Setup of the Simulation

The overall aim of this chapter is to compare the measured characteristics of the HBTs to the results of 2D device simulation. For a quantitative comparison between simulation and experiment, the two-dimensional structure which represents the cross-section of the transistor must comprise all features which affect the electrical characteristics. Some features, such as the vertical doping profile, can be determined very accurately by a direct measurements. Other features, such as the lateral doping profile, are not directly accessible by experiment and must be deduced indirectly for example from electrical measurements. In the following, the construction of the two-dimensional simulation structure based on the experimental characterization of the HBTs is described in detail. The resulting simulation structure is shown in Figure 3.8

3.2.1. Geometry of the 2D Simulation Domain

The geometry of the 2D simulation domain is based on TEM images of the transistor cross-section as shown in Fig. 3.8. It includes the complete active region of the device, the contacts and the first metal layer. Open-short de-embedding is used to eliminate the contact pad and interconnect parasitics from the RF-measurements. However, the

contact to the transistor which is formed in the first two metal layers is not included in the de-embedding structures and must be considered in the simulation domain. The base and collector contact lines are formed using only the first metal layer. Only the contact to the emitter uses a connection formed by the second metal layer. Simulation showed that the capacitive coupling of the second metal layer emitter connection to the base and collector is negligible. For this reason, only the first metal layer is included in the 2D simulation domain. Due to the symmetry, only half of the device has to be simulated.
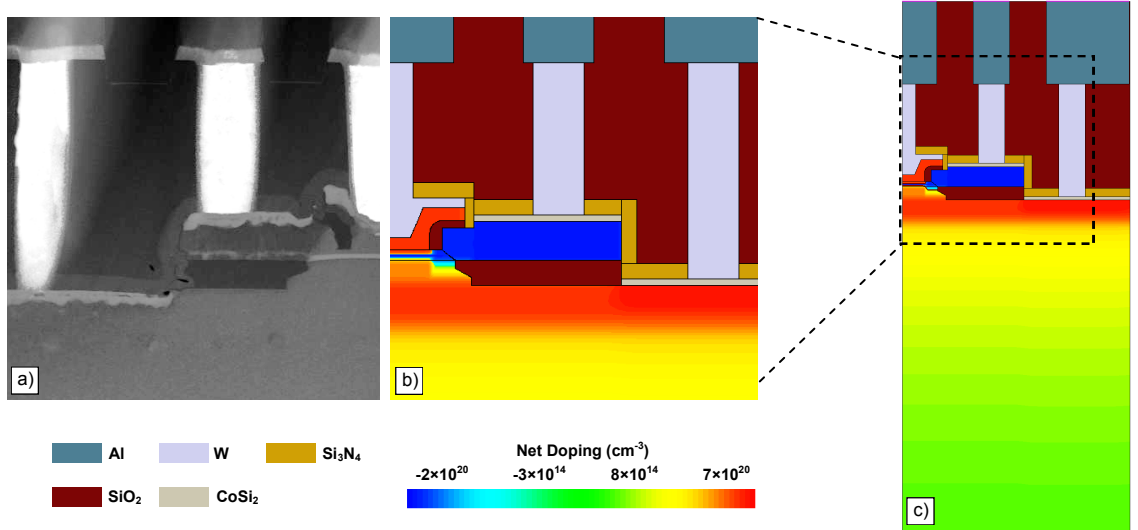

Figure 3.8.: a) TEM image of the HBT cross section. b) Detailed view of the simulation domain. c) Complete simulation domain. It includes the complete active region, the contact wiring and the first metal layer.

3.2.2. Doping Profile

The vertical doping profile used for simulation is shown in Fig. 3.9. The germanium profile is fitted to the final profile, measured by EDX in the HBT structure (see Fig. 3.7). The boron profile has been fitted to the SIMS profile obtained at IHP because it shows the weakest broadening. Additionally, the B concentration was scaled to fit the total dose of $7.2 \times 10^{13} \, \text{cm}^{-2}$ measured by EAG, which seems to be the most reliable result. In a last step, the depth scale of the boron profile was shrunken by 14% to account for the lower deposition rate in small structures.

SIMS measurements of the arsenic doping of the emitter and collector are also shown in Fig. 3.9. A significant As concentration peak can be observed near the base-emitter junction. Such a high concentration of As at the beginning of the emitter deposition is required to achieve the desired steep profile. However, it is believed that not all the As in the peak is electrically active due to the formation of neutral As clusters at impurity concentrations in excess of $10^{20} \, \text{cm}^{-3}$. For this reason, a constant As concentration in
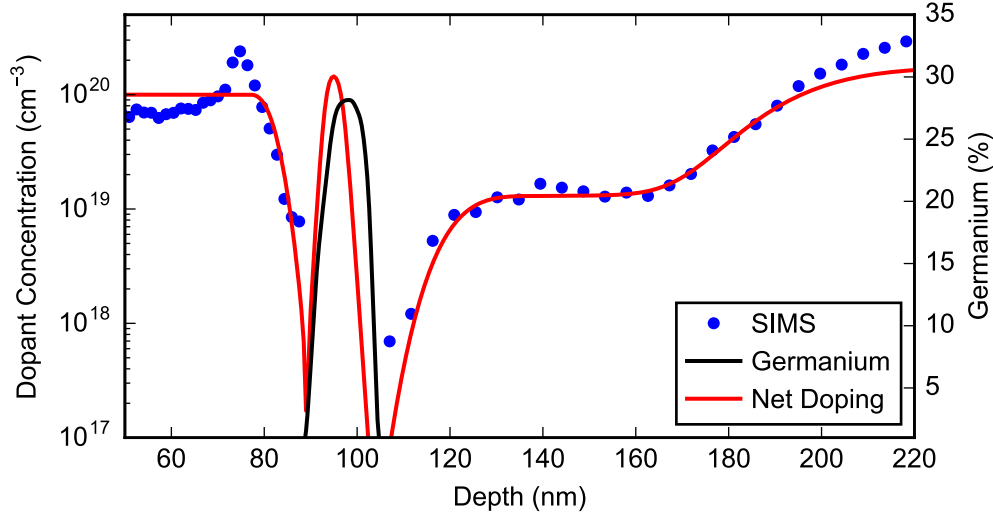
Figure 3.9.: Vertical doping profile used in the simulation. Red line is the net doping concentration, black line is the germanium percentage. SIMS measurements of the arsenic doping are indicated by blue dots.

the emitter is used in the simulation. The arsenic diffusion from the emitter into the base is modeled by a Gaussian function.

Incomplete dopant activation also occurs in the implanted collector. An indication to this effect is that the sheet resistance of the collector well, which results from a simulation with the measured SIMS profile, is significantly lower than the measured sheet resistance. To account for this effect, the peak As-concentration in the collector has been reduced compared to the SIMS profile from $2.1 \times 10^{20}\,\mathrm{cm}^{-3}$ to $1.7 \times 10^{20}\,\mathrm{cm}^{-3}$ (cf. Fig. 3.9). The fitted collector profile yields the correct sheet resistance of $48\,\Omega/\mathrm{sq}$.

The arsenic profile and the base profile were determined from different SIMS measurements, so they do not contain reliable information about their relative position. For this reason, the widths of the base-emitter and base-collector junctions are adjusted to the respective measured junction capacitance. The area-specific junction capacitance can be determined from the dependence on the emitter width. In general, the total capacitance between base and emitter or between base and collector terminals can be separated into two contributions: the inner junction capacitance C_i which is proportional to the junction area and the parasitic edge capacitance C_x, which is proportional to the perimeter length. The latter includes the parasitic capacitance contributions of the emitter perimeter, the overlap between collector and external base and the contact metal. As stated above, the emitter length L_E of the transistors used here is much larger than their width W_E, which means that the contributions of the short edge to C_x can be neglected. The total capacitance between two ports of such a transistor can then be described by

$$C = C_i \cdot L_E \cdot W_E + C_x \cdot L_E. \tag{3.3}$$

In Fig. 3.10c, the measured zero bias BE and BC capacitance per length is plotted

against the emitter width. The capacitance has been extracted from s-parameters and averaged over measurement frequencies from $5\,\mathrm{GHz}$ to $10\,\mathrm{GHz}$. Fitting a straight line to the measured data yields C_{EBi} of $11.7\,\mathrm{fF}/\mu\mathrm{m}^2$ and C_{BCi} of $5.3\,\mathrm{fF}/\mu\mathrm{m}^2$. The corresponding junction widths in the simulation structure are fitted to these values.
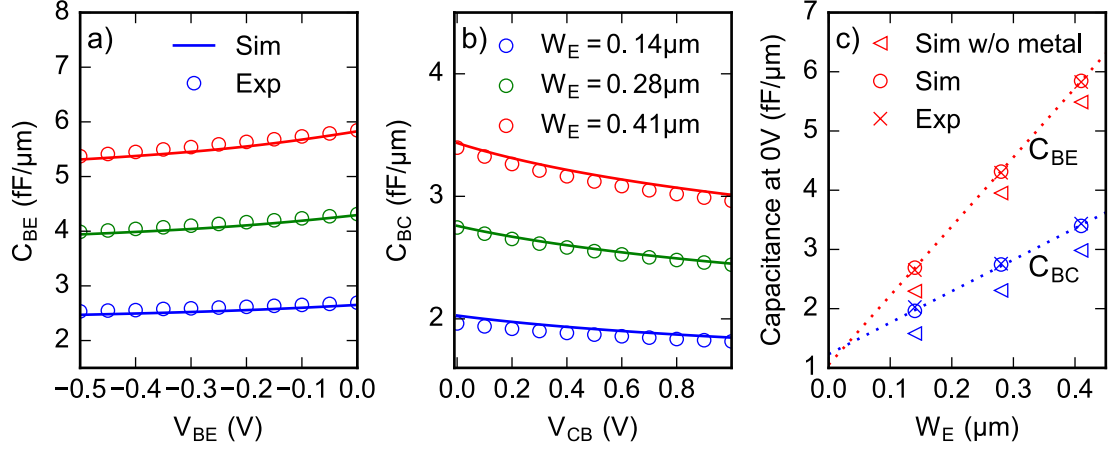

Figure 3.10.: Measured and simulated base-emitter (a) and base-collector (b) capacitance per length as a function of reverse bias voltage. c) Capacitance scaling: zero-bias capacitance as a function of W_E. Linear fit indicated by dotted lines.

The parasitic edge capacitances are $C_{EBx} = 1\,\mathrm{fF}/\mu\mathrm{m}$ and $C_{BCx} = 1.2\,\mathrm{fF}/\mu\mathrm{m}$. These values are mainly determined by the lateral geometry of the transistor, by the doping concentration in the emitter and in the external base and by the lateral extension of the selectively implanted collector. The device geometry is known from TEM images and has been reproduced quite detailed in the 2D simulation domain (cf. Fig. 3.8). The vertical profile of the implanted collector is also known from the SIMS. However, the doping profile in the external base and the lateral profile of the SIC are not directly accessible by measurements.

For the simulation domain, a constant boron doping in the external base of $10^{20}\,\mathrm{cm}^{-3}$ is assumed. The extent of boron diffusion into the internal transistor region and the lateral spreading of the SIC have been adjusted to the measured edge capacitance. Measured and simulated C_{EB} and C_{BC} as a function of the corresponding terminal voltages are shown in Fig. 3.10 for the three HBTs with different emitter widths. One can see that the scaling of the capacitance with W_E and the dependence on the junction voltage are well captured by the simulation.

The 2D simulation domain which is used to simulate the capacitance includes the metal contacts as described in Section 3.2.1. Due to a bug in the device simulator, it was not possible to perform simulations that include self-heating with a device structure that contains metal regions[1]. For this reason, the impact of the metal wiring is modeled by two lumped element capacitors $C_{EB,met}$ and $C_{BC,met}$ connected to the device structure. The resistance of the metal lines is negligible. The values of $C_{EB,met}$ and $C_{BC,met}$ can

---

[1]Bug has been reported and is fixed in a newer version of Sentaurus TCAD [80]

be determined by comparing the simulated capacitance of a device with metal and a device without metal, as shown in Fig. 3.10c. The simulations show that about 45 % of the C_{EBx} and 26 % of C_{BCx} are related to the metal contacts.

3.2.3. Effective Bandgap

Simulations of the HBT with the vertical profile from Fig. 3.9 and the bandgap parameters determined in Section 2.3, produce a collector current that is slightly smaller then the measured current. This difference can have several reasons. The most obvious reason is that the Ge concentration might be too small, although much effort has been put in the characterization of the Ge-profile. However, already a small error in the Ge concentration can lead to a significant change of I_C. For example, a reduction of E_g by about 7.8 meV leads to a doubling of I_C, which corresponds to a difference in the Ge concentration of less than 1.1 at.%. Furthermore, process-inherent stress acting on the base layer has a significant impact on the effective bandgap [81]. As a result the collector current also depends on the layout of the metal interconnects near the device. The bandgap parameters have been determined at relatively large transistors with only one metal layer. So it is also possible that the difference in the effective bandgap results from different stress conditions in the SiGe layer.
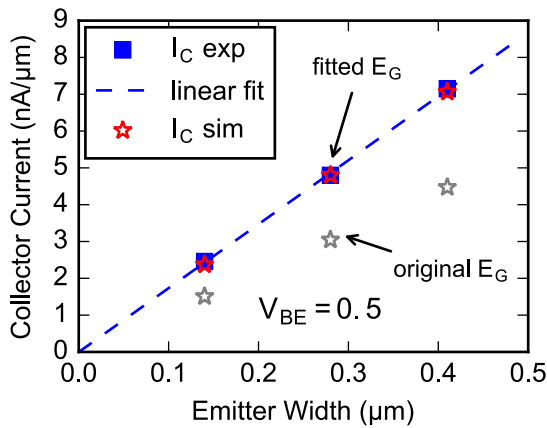

Figure 3.11.: Current scaling: Collector current ($V_{BE} = 0.5V$) as a function of the emitter width W_E.
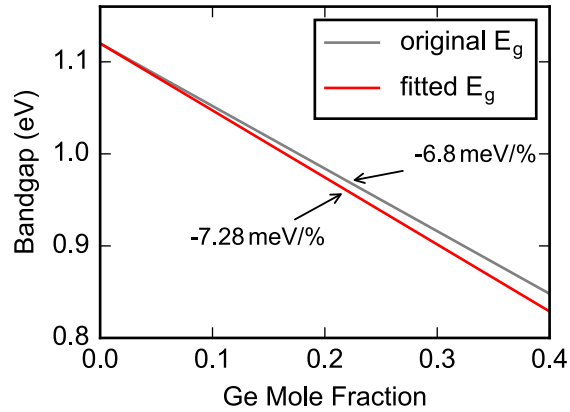
Figure 3.12.: Bandgap as a function of the germanium mole fraction.

Here, the bandgap parameters are adjusted to match the simulated and the measured collector current at $V_{BE} = 0.5$ V. Instead of the original parameters from Section 2.3, which are given by $E_g(x) = 1.12$ eV $- 0.68$ eV $\cdot x$, a slightly larger bandgap reduction of 7.28 meV/% was used here. Figure 3.11 shows the simulated and measured collector current at $V_{BE} = 0.5$ V as a function of the emitter width. The three devices show a nearly ideal scaling with W_E. With the fitted bandgap an excellent agreement between measurement and simulation is achieved. The corresponding bandgap is plotted in Figure 3.12 as a function of the Ge mole fraction.

Theoretical calculations of the band structure of strained SiGe on a Si substrate suggest a type II band alignment with a large offset in the valence band and a very small offset in the conduction band [82]. However, there is no consensus on the exact values of the band offsets. For the simulation in this work, a zero band offset in the conduction band is assumed, so that the valence band offset equals the bandgap difference ($\Delta E_C = 0$, $\Delta E_V = \Delta E_{g,Ge}$).

3.2.4. Series Resistance

Series resistance occurs at various regions of the HBT e.g. at the neutral device regions, at the metal wiring and at the various material interfaces. The former two contributions are naturally accounted for by the simulation structure (given that the doping concentration, the mobility and the metal conductivity are specified correctly). The resistance at material interfaces, especially at the silicide contacts strongly depends on process conditions and thus cannot be included in the simulation in a rigorous way. Instead, lumped element resistances connected to the contacts of the simulation domain are used to include the measured series resistance into the simulation.

To adjust the base resistance in the simulation, measurements of R_B from a tetrode transistor structure as well as measurements of the sheet resistance of the base polysilicon are used. The external base resistance R_{sbx} strongly depends on the sheet resistance of the polycrystalline silicon on top of the oxide. The mobility in poly-Si generally differs from the mobility in crystalline Si, so that realistic values of R_{sbx} cannot be simulated with the default silicon mobility model described in Section 2.2.2. At least the majority mobility in the base-poly must be described properly to achieve reasonable results for R_B. Therefore, the hole mobility in the base poly is modeled by a constant value which is calculated from the sheet resistance R_{sbply} using

$$R_{sbply} = \frac{1}{q \cdot N_A \cdot \mu_{p,eff} \cdot d}. \tag{3.4}$$

Here, N_A denotes the acceptor concentration, $\mu_{p,eff}$ the effective hole mobility and d the thickness of the poly silicon layer. With the assumed boron concentration of $10^{20}\,\text{cm}^{-3}$, a thickness of $0.14\,\mu\text{m}$ and a measured R_{sbply} of $148\,\Omega$, one obtains an effective hole mobility of $30\,\text{cm}^2/(\text{Vs})$. This value is used in the simulation for the hole mobility in the external base.

The base resistance of the HBTs is further evaluated with a transistor tetrode [83]. The tetrode structure has two base terminals which can be biased independently. They are separated by the emitter which completely encloses the inner base (see Fig. 3.13). The tetrode structure can be biased with a small voltage drop ΔV_B between B1 and B2 which leads to a current I through the base. The total resistance R_{B12} between B1 and B2 can be expressed by the following relation

$$\frac{1}{R_{B12}} = \frac{2}{R_{sbi}\frac{W_E}{L_E} + 4\frac{R_{sbx}}{L_E}} + \frac{1}{R_{fs}(W_E)} \tag{3.5}$$

Here, R_{sbi} is the area-specific sheet resistance of the inner base, R_{sbx} is the length-specific resistance of the external base and R_{fs} is the resistance of the foreside of the structure. Tetrode structures with different emitter lengths L_E can be used to separate R_{fs} from R_{B12}. In this way, one obtains the length specific resistance between B1 and B2 without the impact of R_{fs}. The length specific resistance is given by (cf. Fig. 3.13)

$$R_{lr} = R_{sbi} \cdot W_E + 4 \cdot R_{sbx}. \tag{3.6}$$
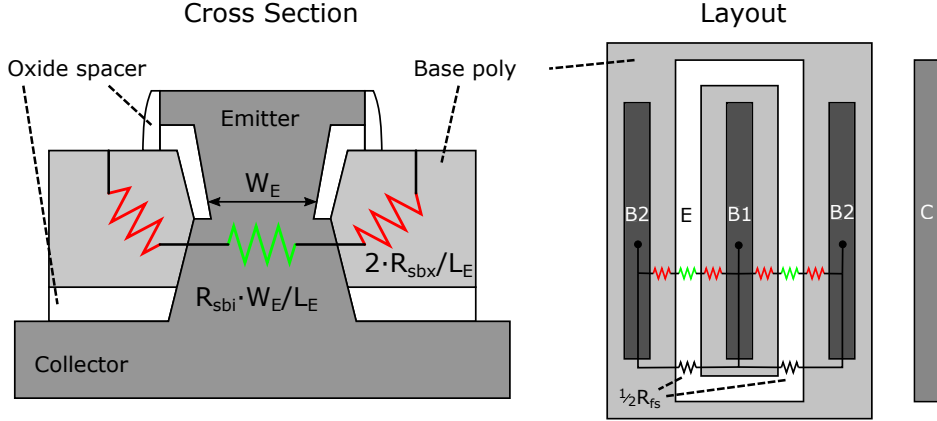
Figure 3.13.: Layout and schematic cross-section of the tetrode transistor structure.

A tetrode structure which corresponds to the layout of D1 ($W_E = 0.14\mu$m) has been available to compare the base resistance of the simulation structure to a measured value. For this purpose, a current flowing through the base from the left to the right contact is simulated. For such a simulation the domain cannot be restricted to one half of the device. The complete cross-section has to be considered. The simulated value of R_{lr} is $610\,\Omega\mu$m, whereas the tetrode structure yields $890\,\Omega\mu$m. Several effect increase the resistance of the external base which cannot be directly included in the simulation. These are for example the interface between single- and polycrystalline silicon or the contact resistance of the silicide. To account for such effects in the simulation, an additional lumped resistance ΔR_{sbx} of $70\,\Omega\mu$m is attached to the base contact.[2]

As in case of the base resistance, also the emitter resistance depends on various aspects and details of the fabrication process and cannot be accounted for by the simulation in a rigorous way. Thus, again a lumped resistance is added to the simulation domain to account for the total emitter resistance. The additional emitter resistance $R_{E,ex}$ is used to fit the simulated Gummel plots at high V_{BE}. A reasonable agreement between simulation and experiment is achieved with an $R_{E,ex}$ of $25\,\Omega\mu$m, $14\,\Omega\mu$m and $12.5\,\Omega\mu$m for D1, D2 and D3, respectively. An essential prerequisite to determine $R_{E,ex}$ in such a way is that self-heating is also treated properly in the simulation.

---

[2] $R_{lr}(\text{Measurement}) - R_{lr}(\text{Simulation}) = 4 \times \Delta R_{sbx} = 890\,\Omega\mu\text{m} - 610\,\Omega\mu\text{m}$

3.2.5. Self-Heating

The global self-heating model [84] is used to account for an increasing lattice temperature at high current densities. Within this model, a uniform lattice temperature T_L for the complete simulation domain is calculated from the dissipated power P_{diss}. With the current I_i and the voltage V_i at contact i, the total dissipated power is given by

$$P_{diss} = \sum_i I_i \cdot V_i \tag{3.7}$$

The global lattice temperature is calculated from the following heat balance equation:

$$T_L - T_0 = P_{diss} \cdot R_{TH} \tag{3.8}$$

Here, R_{TH} denotes the thermal resistance and T_0 is the temperature of the corresponding heat sink connected to the device. The main advantage of the global self-heating model is that measured values of R_{TH} can be used in the simulation, which results in the correct excess temperature. Measured values of R_{TH} from the reference transistors are given in Table 3.1.

Device simulation tools are also able to simulate the electrothermal behavior of the transistor by solving the heat equation for the lattice temperature coupled to the semiconductor equations. In principle, such simulations yield a more realistic temperature distribution than the global self-heating model, which assumes a uniform temperature over the whole device [85, 86]. However, using the global self-heating model reduces simulation time, improves the convergence of the iterative solver and allows to use measured values of R_{TH}. Furthermore, as pointed out by D'Alessandro et al. [87], realistic simulations of the thermal behavior are not possible with 2D device simulation tools, because heat diffusion is an intrinsically three-dimensional process. For these reasons, the global self-heating model is the preferred method to account for self-heating in this work.

In Sentaurus Device, the global self-heating model is implemented as a post-processing model [37]. The uniform lattice temperature T_L is computed by equation (3.8), after the coupled solution of the transport equations. Therefore, the step size of a quasi-stationary simulation must be sufficiently small to capture the dependence of T_L on the applied bias voltage correctly. Another consequence of the fact, that the temperature is not computed self-consistently with the other variables, is that the temperature response is not considered in the small-signal ac simulation. However, at frequencies in the GHz range, the temperature response is very small and can be neglected, as will be demonstrated below.

To test whether the global self-heating model can be used to simulate the RF figures of merit of the SiGe HBTs, a comparison with full electrothermal simulations is performed. Within these simulations, the balance equation for the lattice temperature T_L is solved in addition to the energy balance equation of the electrons [37]. A thermal contact is added at the bottom of the simulation domain assuming that most heat is dissipated to the substrate. The thermal resistance at this contact has been adjusted in such a way, that the collector current agrees with the results from the simulations
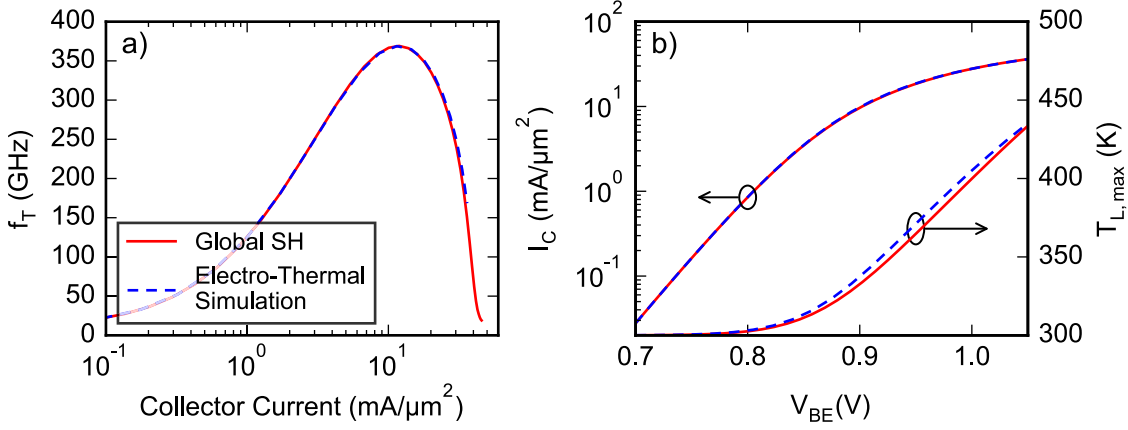
Figure 3.14.: Comparison between full electro-thermal simulation and the global self-heating model. a) f_T as a function of I_C. b) Collector current and maximum lattice temperature as a function of V_{EB}. $V_{CB} = 0\,\text{V}$.
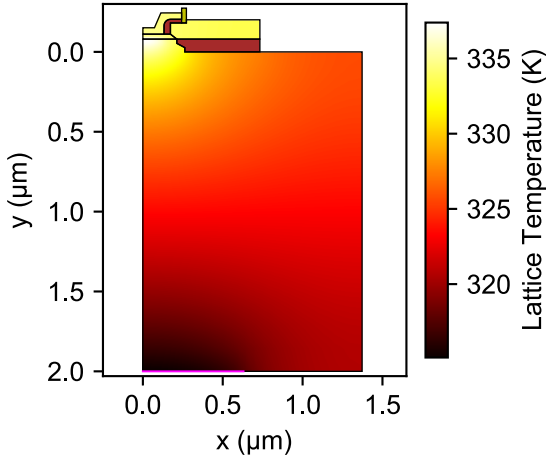

Figure 3.15.: Distribution of the lattice temperature T_L at $V_{BE} = 0.9\,\text{V}$. The magenta line indicates the thermal contact.
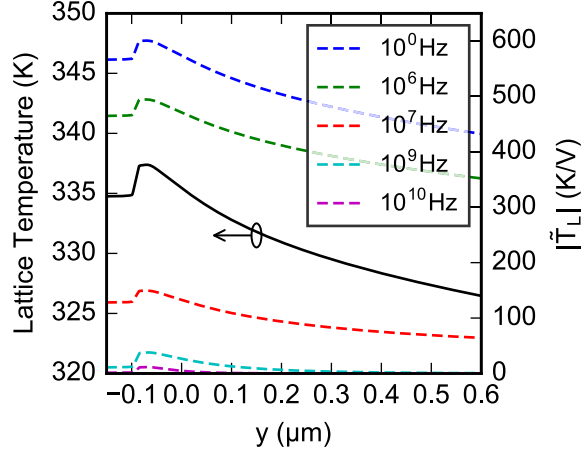
Figure 3.16.: Cross-sections of T_L and the small-signal temperature response ΔT_L at $x = 0.001\,\mu\text{m}$ for $V_{BE} = 0.9\,\text{V}$.

with the global self-heating model. The simulated collector current I_C and the lattice temperature T_L are plotted in Fig. 3.14b as a function of base-emitter voltage. In case of the full electrothermal simulation, the maximum value of T_L is plotted, which roughly agrees with the average junction temperature of the transistor. Both simulations agree very well. The spatial distribution of the lattice temperature is illustrated in Fig. 3.15. The simulated transit frequency f_T is plotted in Fig. 3.14a as a function of the collector current density for the different simulation approaches. It can be seen that both approaches yield the same f_T. This result shows that the small-signal response of T_L has no significant impact on f_T.

Small-signal ac simulation computes the linear response of the device to a small sinusoidal variation of the dc voltage. For this purpose, the system variables (n, p, Φ, T_n, T_L) are written in the from $\xi(t) = \xi_0 + \tilde{\xi}(\omega)e^{i\omega t}$, with the dc solution ξ_0 and the corresponding

small-signal response $\tilde{\xi}$ [37]. The magnitude of the small-signal response of the lattice temperature $|\tilde{T}|$ is plotted in Fig. 3.16 for different frequencies. A strong decrease of $|\tilde{T}|$ with increasing frequency can be observed. The transit frequency is determined from the small-signal current gain h_{21} at a frequency of 40 GHz (cf. Section 3.3). At such high frequencies, the impact of the lattice temperature response can be neglected. For this reason, the RF figures of merit can safely be calculated from simulations that utilize the global self-heating model.

3.3. Calculation of f_T and f_{max}

The transit frequency f_T is defined as the frequency at which the magnitude of the small-signal short-circuit current gain h_{21} becomes one. Correspondingly, the maximum oscillation frequency f_{max} is defined as the frequency where the power gain becomes one. There are different definitions for the power gain of the HBT. Here, Mason's unilateral gain U is used to determine f_{max}. U can be calculated from the impedance parameters by [88]

$$U = \frac{|z_{21} - z_{12}|^2}{4 \left[\text{Re}(z_{11}) \, \text{Re}(z_{22}) - \text{Re}(z_{12}) \, \text{Re}(z_{21}) \right]}. \tag{3.9}$$

Both h_{21} and U can be determined as a function of frequency using small-signal AC simulation, as shown in Fig. 3.17a. At low frequencies, the gain is constant. At high frequencies h_{21} can be approximated by (cf. Section 4.2.2)

$$h_{21} = \frac{g_m}{j\omega(C_{EB} + C_{BC})}. \tag{3.10}$$

Assuming that g_m and $(C_{EB} + C_{BC})$ are independent of the frequency, $|h_{21}|^2$ falls off with a slope of -20 dB per decade. The same ideal slope can be derived for the unilateral gain U. The results of small-signal simulation show that this assumption is valid for h_{21} up to a frequency of 100 GHz. At higher frequencies, the simulated gain deviates significantly from the ideal slope. In contrast, the simulated unilateral gain U shows an ideal behavior up to 1 THz. Note that the simulation domain does not include coupling of the collector to the substrate. Typical substrate coupling networks used in compact models cause U to decay with a steeper slope than -20 dB per decade.

In reality, reliable measurements of the gain at frequencies as high as f_T and f_{max} are not possible. Thus, f_T and f_{max} are usually determined by an extrapolation of the gain to higher frequencies, assuming an ideal slope of -20 dB per decade. Under this assumption, f_T and f_{max} can simply be calculated from the gain at a single frequency:

$$f_T = |f_{ex} \cdot h_{21}(f_{ex})| \tag{3.11}$$

$$f_{max} = f_{ex} \cdot \sqrt{U(f_{ex})} \tag{3.12}$$

Here, f_{ex} denotes the extrapolation frequency. Fig. 3.17b shows the extrapolated f_T and f_{max} as a function of f_{ex}. The transit frequency f_T increases significantly at higher extrapolation frequencies because the slope of the simulated h_{21} deviates from the ideal

$-20\,\mathrm{dB/dec}$. For example, the f_T value which results from an extrapolation at $40\,\mathrm{GHz}$ is $360\,\mathrm{GHz}$, compared to $420\,\mathrm{GHz}$ where the simulated $|h_{21}|$ equals $0\,\mathrm{dB}$.
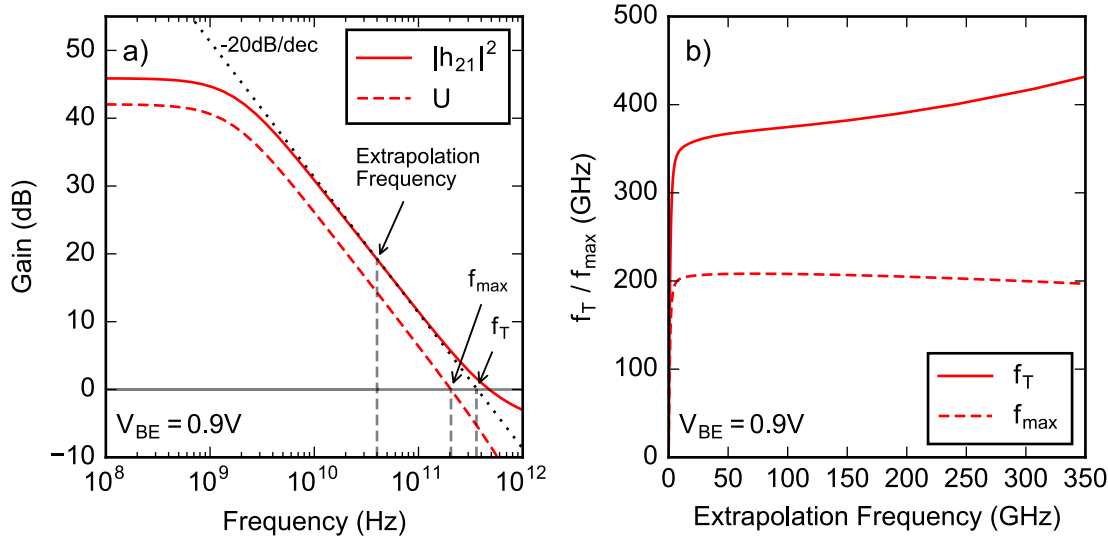

Figure 3.17.: a) Calculated current gain h_{21} and unilateral power gain U as a function of frequency at $V_{BE} = 0.9\,\mathrm{V}$. b) f_T and f_{max} as a function of the extrapolation frequency.

Experimentally, f_T and f_{max} are determined in this work by extrapolation of h_{21} and U at $40\,\mathrm{GHz}$, respectively. For this reason, the simulated figures of merit which are presented in the following are determined by (3.11) or (3.12) with $f_{ex} = 40\,\mathrm{GHz}$.

3.4. Simulation Results

Measured and simulated Gummel characteristics of D2 are shown in Figure 3.18. Due to the adjustment of the effective bandgap and the emitter contact resistance $R_{E,ex}$, a good agreement between measured and simulated collector current is achieved. The base current is also shown in this figure. It is determined by recombination in the base and in the emitter. Ohmic boundary conditions are used at the contacts corresponding to infinite recombination velocities. The following bulk recombination models have been activated in Sentaurus Device: Auger recombination, Shockley-Read-Hall (SRH) recombination with doping dependent recombination lifetimes, the trap-assisted tunneling model by Schenk as well as the band-to-band tunneling model by Hurkx [37, 50, 89]. The default model parameters have been used in the simulation. For $V_{BE} > 0.5\,\mathrm{V}$, there is a quite good agreement between simulated and measured I_B. At smaller V_{BE}, where the base current is dominated by band-to-band tunneling, the simulation deviates from the measurement. The base current at low V_{BE} is not relevant for a correct simulation of the high-frequency figures-of-merit f_T and f_{max}, so no adjustment of the recombination models has been carried out here.

The measured and simulated ideality factor of the collector current n_{I_C} is also shown in Fig. 3.18. The simulation yields a collector current that is less ideal than the measured

current. At $V_{BE} = 0.5\,\text{V}$ the measured n_{I_C} is 1.025, whereas the simulated n_{I_C} is 1.033. As a result, the simulated I_C at high V_{BE} is slightly smaller than the measured current because the simulated I_C has been fitted to the measurements at $V_{BE} = 0.5\,\text{V}$. An in-depth discussion of the simulated collector current ideality is given in the Section 3.4.2.
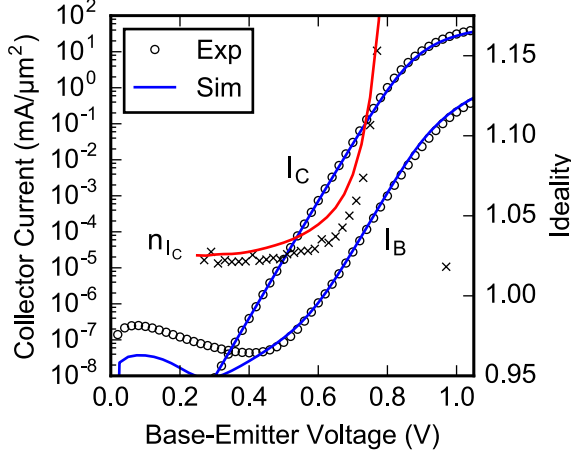

Figure 3.18.: Measured and simulated gummel characteristics and the ideality of the collector current n_{I_C} of device D2.
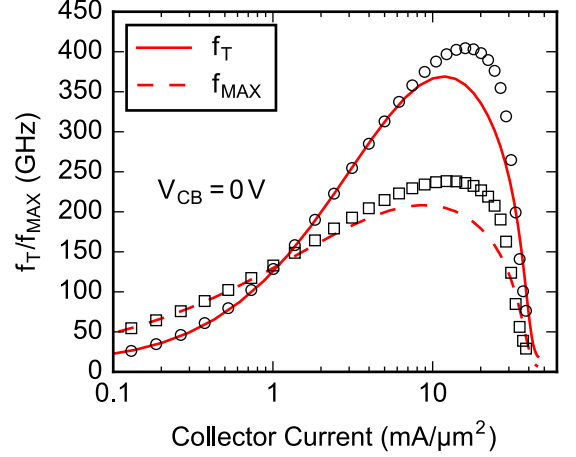
Figure 3.19.: Measured and simulated f_T and f_{MAX} of device D2 at $V_{CB} = 0V$.

Measured and simulated f_T and f_{max} of D2 are shown in Figure 3.19 as a function of the collector current density. Up to a current density of about $5\,\text{mA}/\mu\text{m}^2$, a good agreement between experiment and simulation can be observed. However, the simulated f_T begins to decrease at a significantly lower current density than the measured curve. It reaches its peak value of only $367\,\text{GHz}$ at a current density of $10\,\text{mA}/\mu\text{m}^2$ ($V_{BE} = 0.9\,\text{V}$), whereas the measured peak value of f_T equals $404\,\text{GHz}$ at $16\,\text{mA}/\mu\text{m}^2$ ($0.93\,\text{V}$). So, the simulated peak f_T is $9\,\%$ smaller than the real value. A similar behavior can be observed for f_{max}. At low current densities, simulation and measurement match well but the f_{max} roll-off also starts earlier in the case of the simulation. It yields a peak f_{max} of $208\,\text{GHz}$ at $9\,\text{mA}/\mu\text{m}^2$ compared to the measured value of $238\,\text{GHz}$ at $12\,\text{mA}/\mu\text{m}^2$. A detailed investigation of the difference between measured and simulated f_T is given in Section 3.4.4 and Section 3.4.5.

3.4.1. Hydrodynamic Transport Model Parameters

The transport model parameters r, f^{td} and f^{hf} have a significant impact on the simulation results. Following the suggestions of [59] and [10], the energy flux coefficient r and the thermal diffusion parameter f^{td} are set to 1. The heat flux diffusion factor has to be sufficiently small to avoid an unphysical negative Early-Voltage. Simulated Gummel and output characteristics are shown in Fig. 3.20 for different values of f^{hf}. For $f^{hf} > 0.5$ a negative Early-Voltage can be observed in the output characteristics with a fixed V_{BE} of $0.7\,\text{V}$. For values of f^{hf} smaller 0.5, this effect does not occur. Very small values of f^{hf} lead to a significant decrease of the collector current and to an increase of n_{I_C}.

Best agreement with the measured Gummel and output characteristics is achieved for $f^{hf} = 0.4$. This value is used throughout this work. The impact of f^{hf} on the simulated transit frequency is shown in Fig. 3.20c. One can see that f_T increases with higher values of f^{hf}, except for $f^{hf} = 0.05$ which yields approximately the same f_T as $f^{hf} = 0.6$. The maximum value of f_T ranges from 365 GHz for $f^{hf} = 0.2$ to 395 GHz for $f^{hf} = 1$.

In [59] and [10], the heat flux parameter has been calibrated to the results of BTE simulations. An optimal f^{hf} of 0.07 was obtained in [59] for a generic HBT with f_T of 450 GHz. A value of 0.295 was found in [10] to be suitable for different HBT generations.
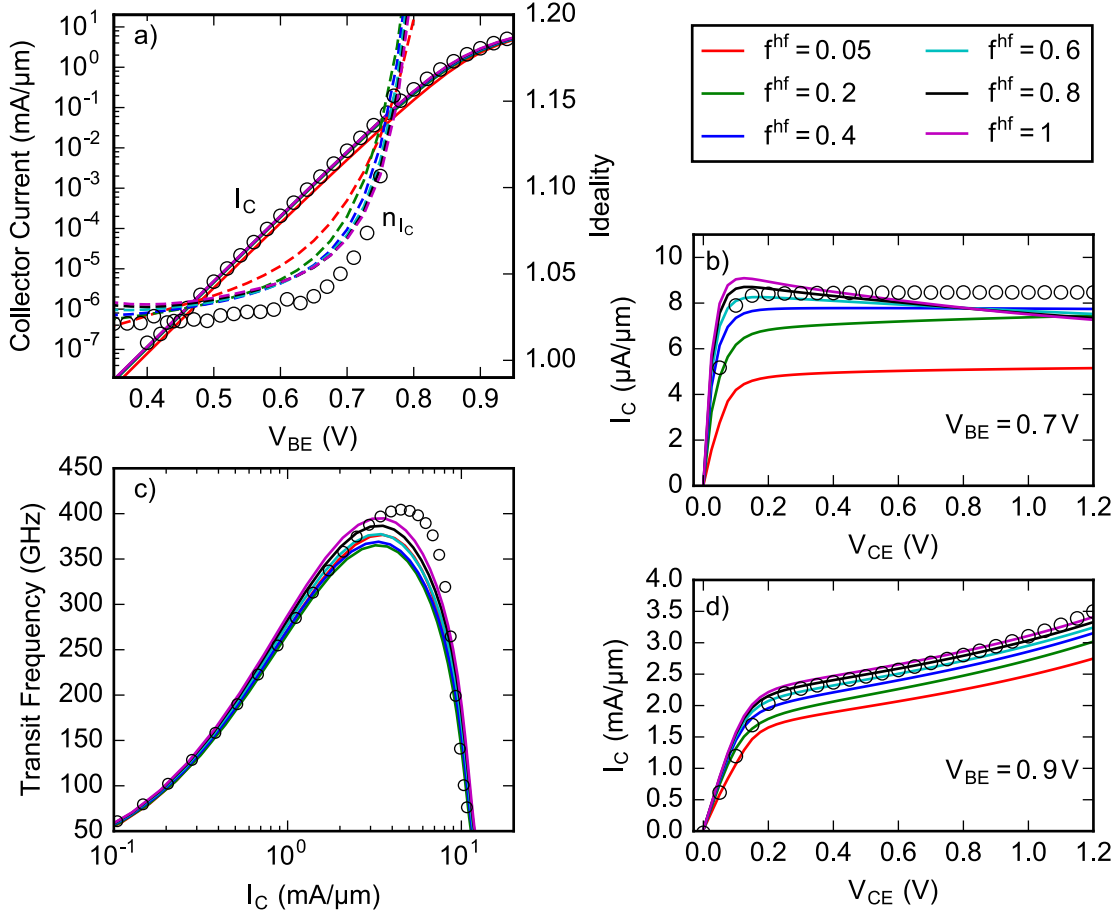
Figure 3.20.: a) Simulated collector current and corresponding ideality for different values of f^{hf}. Symbols indicate measurements. b, d) Measured and simulated output characteristics. c) Measured and simulated f_T.

3.4.2. Impact of the Bandgap on the Collector Current Ideality

The ideality of the collector current of a SiGe HBT is highly sensitive to the shape and the position of the Ge profile [90]. This can be explained by the Moll-Ross relation for the collector current which is given by (2.75). Assuming a constant electron diffusivity

3. Comparison of 2D Simulation and Experiment

D_n, the saturation current J_{C0} can be written as

$$J_{C0} = qD_n \left(\int_0^{w_b} \frac{p(x)}{n_{i,eff}^2(x)} dx \right)^{-1}. \tag{3.13}$$

The saturation current is determined by the integral of the base charge p weighted by the square of the effective intrinsic density $n_{i,eff}$. The effective intrinsic density depends exponentially on the bandgap (cf. (2.76)). Thus, if there are regions in the base in which the Ge concentration is small, these regions make a major contribution to the integral in (3.13) due to a small $n_{i,eff}^2$. In case of the doping profile which is investigated here, a considerable part of the base doping lies inside the gradient of the Ge profile where the Ge concentration is relatively low. A change of the base-emitter voltage changes the width of the BE space-charge region and shifts the boundary of the neutral base. This shift leads to a change of the hole concentration p in the particular region where the Ge concentration is small. For this reason, the HBTs used here have a relatively high n_{I_C} of about 1.02. In comparison, the HBTs with box-shaped Ge profiles used in Section 2.3 have a n_{I_C} of less than 1.002.

One possibility to obtain a better agreement between measured and simulated collector current ideality is to adjust the shape of the vertical Ge profile [74]. However, this approach does not necessarily yield a more realistic profile. At worst, one just compensates the insufficiency of the physical models by a wrong Ge profile. Another possibility to fit the simulated n_{I_C} is to adjust the model of doping induced bandgap narrowing. Assessing the shape of the bandgap at the emitter side of the base in more detail, it turns out that, besides the Ge gradient, doping-induced bandgap narrowing also has a strong impact on the ideality of the collector current.

In the following, these two possible sources for the poor agreement between measured and simulated ideality are investigated. Therefore, two examples are presented which demonstrate that the ideality of the simulated collector current can be matched to the measurement. First, this matching is accomplished by an adjustment of the Ge profile. In the second example, it is accomplished by changing the doping-induced bandgap narrowing model.

Measured and simulated Gummel characteristics and the corresponding n_{I_C} are shown in Fig. 3.21 for three different cases. Fig. 3.21a shows the results of the original simulation with the vertical profile from Section 3.2.2 and the bandgap narrowing model described in Section 2.3. The results of the simulations with the adjusted Ge profile and the changed bandgap narrowing are shown in Fig. 3.21b and Fig. 3.21c, respectively. The adjusted Ge profile is plotted in Fig. 3.22b. Compared to the original profile, the leading edge has been shifted towards the emitter by 0.8 nm. An error of a few nanometers can be assumed for the relative position of the measured boron and Ge profiles, so the adjusted profile still lies within the expected error range. The shift of the Ge profile leads to a higher Ge concentration at the base-emitter junction and thus to a higher collector current. To get the same collector current with the shifted Ge profile, the Ge-dependent bandgap has to be reduced. Agreement with the measured I_C is achieved for the new profile using a $\Delta E_{g,Ge}$ of 6.8 meV/%. The collector current ideality is reduced by the
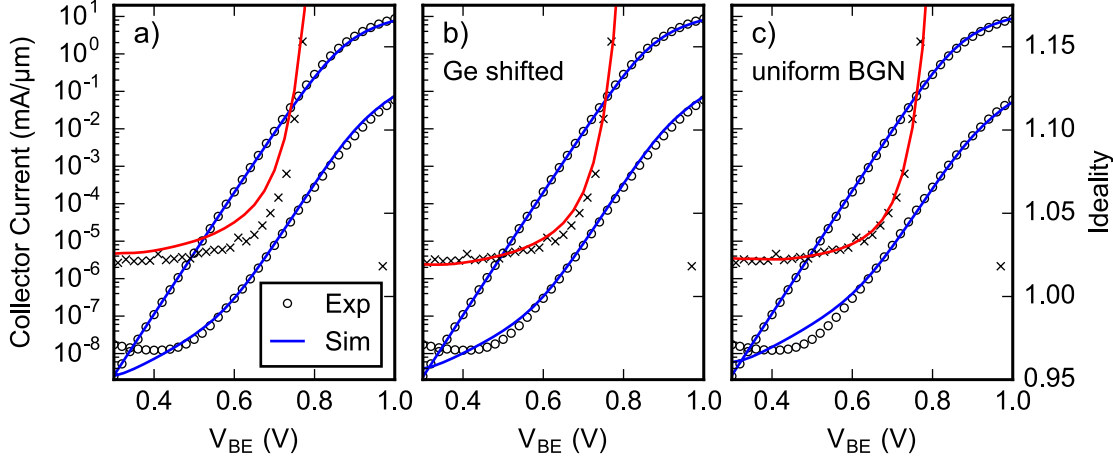
Figure 3.21.: Measured and simulated Gummel plots and collector current ideality for D2. a) Simulated with original profile and bandgap. b) Simulated with shifted Ge profile. c) Simulated with uniform bandgap narrowing.

profile shift to 1.025 at $V_{BE} = 0.5\,\text{V}$ which agrees with the measured n_{I_C}. However, the V_{BE} dependence of n_{I_C} still differs from the measured results. A better agreement could in principle be achieved by further adjustment of the Ge profile, but this would be in contradiction with the measured profile. The effect of the profile changes can also be observed in the conduction band, which is shown in Fig. 3.22. The conduction band peak at the BE junction, which becomes significant at $V_{BE} = 0.7\,\text{V}$, is less pronounced in case of the shifted Ge profile.

In the second row of Fig. 3.22 the bandgap is illustrated for the three different simulations. Both the SiGe bandgap $E_{g,SiGe}$ and the effective bandgap $E_{g,eff}$ are shown. $E_{g,SiGe}$ only includes the bandgap reduction due to germanium while $E_{g,eff}$ also includes the doping-induced bandgap narrowing:

$$E_{g,SiGe} = 1.12 - \Delta E_{g,Ge} \tag{3.14}$$

$$E_{g,eff} = 1.12 - \Delta E_{g,Ge} - \Delta E_{g,dop} \tag{3.15}$$

The difference between the two curves is the doping-induced bandgap narrowing $\Delta E_{g,dop}$. One can see that the shape of the bandgap at the edge of the BE space-charge region is not only determined by the Ge profile but also by the doping-induced bandgap narrowing. $\Delta E_{g,dop}$, which is modeled as function of the total doping concentration $N_A + N_D$, has a local minimum at the BE junction and a maximum in the center of the base. This causes the strong gradient of the bandgap which, in case of the original profile, leads to the higher ideality factor. The shift of the Ge profile towards the emitter reduces the impact of $\Delta E_{g,dop}$ on the shape of the bandgap in that region, and thus leads to a reduction of n_{I_C}.

Due to the fact that doping-induced bandgap narrowing is mainly responsible for the poor agreement between measured and simulated ideality, it is interesting to see how the ideality behaves if the bandgap is only determined by the Ge profile. Therefore, the doping dependence of the bandgap is switched off and a uniform bandgap narrowing is
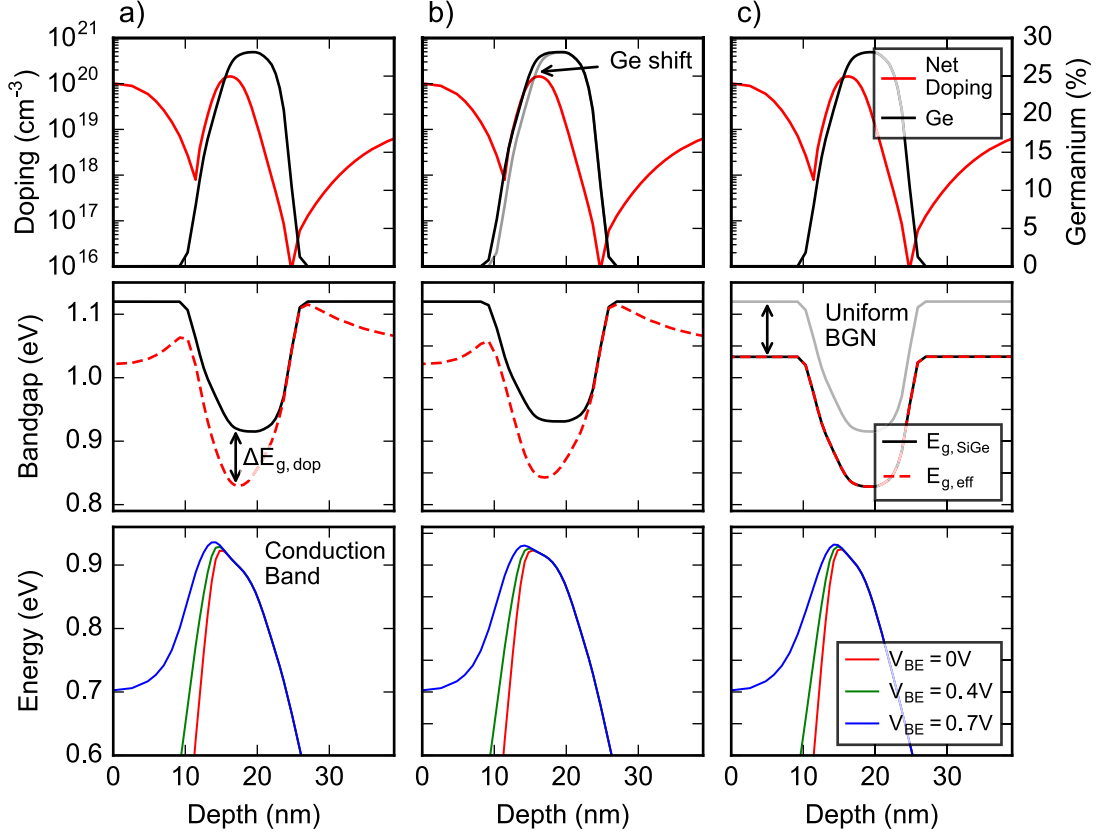
Figure 3.22.: Vertical profile, bandgap and conduction band energy as a function of depth. Left column (a) shows simulations with the original profile and bandgap narrowing. Center and right columns (b, c) show simulations with a shifted Ge profile and a uniform doping-induced bandgap narrowing, respectively.

applied to the whole device. The value of the uniform BGN equals the doping-induced BGN in the center of the base so that the collector current density is similar in both simulations. The resulting bandgap is illustrated in Fig. 3.22c and the corresponding simulation results are shown in Fig. 3.21c. A very good agreement between measured and simulated n_{I_C} can be observed in this case.

These results show that the simulated collector current ideality is extremely sensitive to the shape of the the band structure at the base-emitter junction, especially if the leading edge of the Ge profile is close to the leading edge of the base doping. A reasonable agreement between measured and simulated n_{I_C} can be achieved by a shift of the Ge profile by 0.8 nm which is justifiable by the inaccuracy of the profile measurement. However, it is also possible that the bandgap peak in the BE depletion region, which is caused by doping-induced bandgap narrowing, is overestimated in the simulation. In this case, the shifted Ge profile would compensate an insufficient physical model. The impact of the profile and bandgap changes on the simulated f_T is shown in Figure 3.27.

3.4.3. High Injection and Self-Heating

Self-heating has a significant impact on the RF-performance of SiGe HBTs. Hence, it is important that this effect is considered properly in the simulation. In this section we examine how accurate self-heating can be described by the global self-heating model that uses a uniform lattice temperature for the whole device. Figure 3.23 shows the collector current as a function of V_{BE} for different V_{CB} illustrating the effect of self-heating. According to (3.7), the dissipated power increases with V_{CB} leading to an increase of the lattice temperature T_L at a sufficiently high current density. The higher temperature in turn leads to an increase of I_C. So, the Gummel characteristics measured at different V_{CB} split up at high V_{BE} due to self-heating. This effect is most pronounced for D3 with $W_E = 0.41\,\mu\text{m}$ because more power is dissipated by a larger transistor. The dotted lines in Fig. 3.23 show the simulation results with the standard profile and a constant thermal resistance, the same setup that is used for the simulations in Figs. 3.18 and 3.19. The dashed lines result from simulations with the Ge profile which is shifted by $0.8\,\text{nm}$, as shown in Fig. 3.21b. The solid lines show the results of simulations which use a temperature-dependent thermal resistance $R_{TH}(T)$ (in addition to the shifted Ge profile).
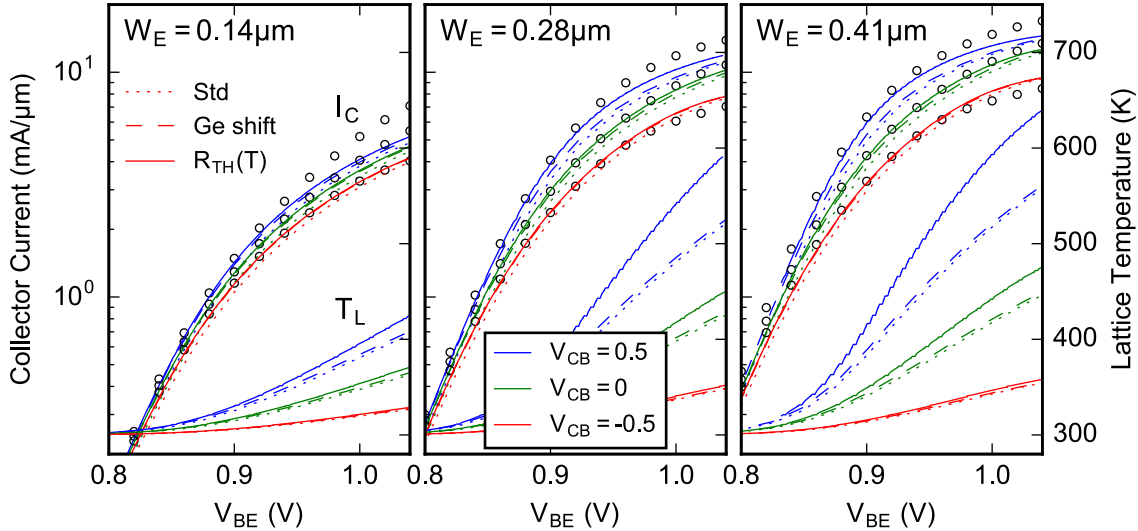

Figure 3.23.: Measured and simulated I_C as well as simulated lattice temperature T_L as a function of V_{BE} at a V_{CB} of $0.5\,\text{V}$, $0\,\text{V}$ and $-0.5\,\text{V}$.

The standard simulation reproduces the split-up of the Gummel plots quite well for V_{CB} of $-0.5\,\text{V}$ and $0\,\text{V}$. The absolute value of I_C is a bit too small at such high V_{BE} due to the overestimation of n_{I_C}. The shifted Ge profile gives a better agreement with the measurement. However, in the case of the strongest self-heating at $V_{CB} = 0.5\,\text{V}$, both the standard simulation and the simulation with the shifted Ge profile strongly underestimate the collector current at $V_{BE} > 0.9\text{V}$. The main reason for this underestimation is the assumption of a constant thermal resistance. At lattice temperatures in excess of $400\,\text{K}$ the temperature dependence of R_{TH} has to be taken into account. Here, $R_{TH}(T)$

3. Comparison of 2D Simulation and Experiment

is described by the phenomenological model

$$R_{TH}(T) = R_{TH}(300\,\mathrm{K}) \left(\frac{T}{300\,\mathrm{K}} \right)^{\eta} \tag{3.16}$$

with $\eta = 0.3$, which yields a reasonable increase of T_L. The global self-heating model in Sentaurus Device does not support a temperature dependent R_{TH}. In order to still produce Gummel plots with temperature dependent R_{TH}, successive simulations with stepwise increasing R_{TH} have to be performed. The quasi-stationary ramp-up of V_{BE} is interrupted and restarted with an updated R_{TH}, whenever the lattice temperature has risen by $10\,\mathrm{K}$. As a consequence of this simulation procedure, small steps can be observed in the simulated curves in Fig. 3.23. In case of the two larger transistors D2 and D3, the agreement between measurement and simulation is significantly improved by taking into account the temperature dependence of R_{TH}. However there is still a notable deviation at very high V_{BE}.

In case of the smallest HBT, the impact of the temperature dependence of R_{TH} is much smaller because of the lower lattice temperature T_L. Furthermore, the deviation between measured and simulated current at very high V_{BE} is most pronounced for D1. One probable cause of the poor agreement in case of D1 could be the non-uniformity of the SiGe base layer. The SiGe layer tends to be thicker at the edge of the collector window (cf. Fig. 3.1). In connection with current crowding, this could lead to an effective bandgap reduction at high current densities.
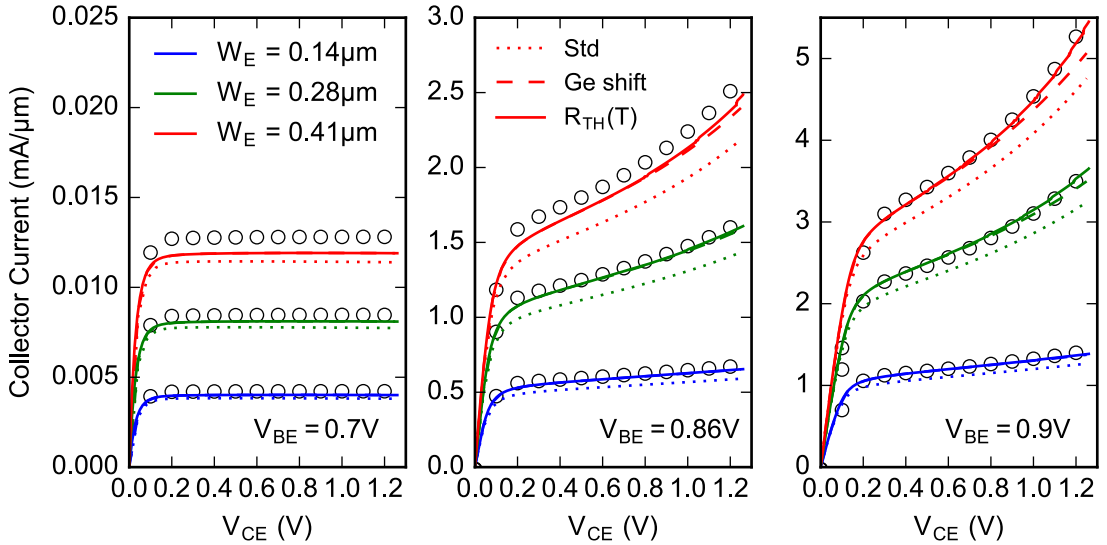

Figure 3.24.: Measured and simulated output characteristics at a fixed V_{BE} of $0.7\,\mathrm{V}$, $0.86\,\mathrm{V}$ and $0.9\,\mathrm{V}$.

Output characteristics for the three devices are shown in Fig. 3.24 for a fixed base-emitter voltage of $0.7\,\mathrm{V}$, $0.86\,\mathrm{V}$ and $0.9\,\mathrm{V}$. Simulation results are again shown for the standard setup, for the simulation with the shifted Ge profile and for the simulation that includes the shifted Ge profile and the temperature dependence of R_{TH}. As already seen

in the Gummel plot, the original transistor profile leads to slight underestimation of I_C. The overall agreement between measurement and simulation is very good. In case of the largest transistor, it is necessary to take into account the temperature dependence of R_{TH} to reproduce the measured output characteristics at $V_{BE} = 0.9\,\text{V}$. In summary it can be said that the effect of self-heating on the collector current is well reproduced by the simulation, at least up to a base-emitter voltage of $0.95\,\text{V}$.

3.4.4. Transit Frequency

The dependence of the transit frequency f_T on the emitter width W_E is illustrated in Fig. 3.25, which shows measured and simulated f_T for the 3 different emitter geometries. The highest f_T of $404\,\text{GHz}$ is reached for the intermediate device D2 with a width of $0.28\,\mu\text{m}$. At low and medium current densities, D2 and D3 nearly show the same f_T whereas D1 has a significantly smaller f_T due to the larger contribution of the device perimeter to the parasitic capacitance. The peak value of D3 is smaller than that of D2, due to various reasons: There is an additional delay as a result of the distributed base resistance. Self-heating is stronger in larger HBTs and the effect of R_C is increased, too. Apart from the fact that the f_T roll-off sets in too early, the dependence of f_T on the emitter width is well captured by the simulation. For D3, simulated peak f_T is $10\,\%$ smaller and for D1 it is $6\,\%$ smaller than in the measurement.
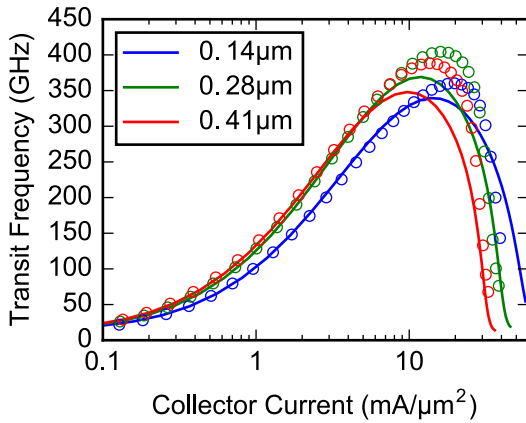
Figure 3.25.: Measured and simulated f_T for different emitter geometries. $V_{CB} = 0\,\text{V}$.
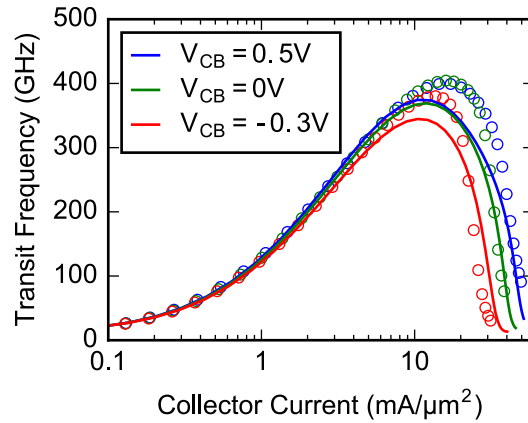
Figure 3.26.: Measured and simulated f_T for different V_{CB} for D2.

The dependence of f_T on the collector-base voltage V_{CB} is shown in Figure 3.26, where f_T is plotted as a function of I_C at $V_{CB} = 0.5\,\text{V}$, $0\,\text{V}$ and $-0.3\,\text{V}$. Two opposing effects govern the dependence of f_T on the collector-base voltage. If the BC junction is reverse biased ($V_{CB} > 0$) the additional electric field counteracts the widening of the base into the collector region and thus delays the onset of the Kirk effect. On the other hand, the additional electric field increases self-heating, which results in a degradation of f_T. For the transistor D2 one can see in Fig. 3.26 that peak f_T is merely affected by an increase of V_{CB} from $0\,\text{V}$ to $0.5\,\text{V}$. However the roll-off of f_T is shifted to a higher current density

for $V_{CB} = 0.5\,\mathrm{V}$. If V_{CB} equals $-0.3\,\mathrm{V}$, which means that the BC junction is forward biased, degradation of f_T due to high-current effects sets in much earlier. This results in a significant drop of peak f_T. Qualitatively, this behavior is well reproduced by the simulation. However, again the maximum values of the simulated f_T-curves are between 6 % to 9 % lower than the experimental values.
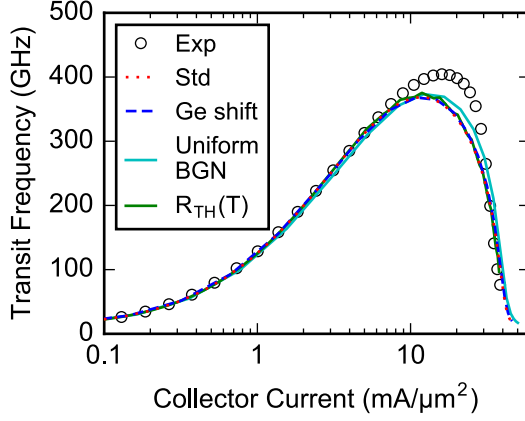

Figure 3.27.: Measured and simulated f_T for D2 at $V_{CB} = 0\,\mathrm{V}$.
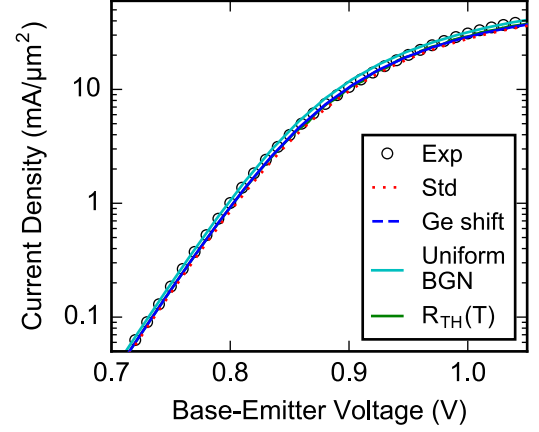
Figure 3.28.: Measured and simulated I_C as a function V_{BE} for D2 at $V_{CB} = 0\,\mathrm{V}$.

The impact of the profile and bandgap variations discussed in Section 3.4.2 on the transit frequency is shown in Fig. 3.27 for the example of D2. The corresponding collector currents are shown in Fig. 3.28. One can see that these variations have almost no impact on f_T. Only the uniform bandgap leads to a slight increase of f_T at high currents. Simulations including the temperature-dependent thermal resistance are also shown and the impact on f_T is also negligible.

In the following, the reason for the marked deviation between measured and simulated f_T is examined. Quasi-static transit time analysis is used to reveal the mechanism which is responsible for the premature f_T fall-off. In order to clarify how specific features of the hydrodynamic transport model (e.g. velocity overshoot) are responsible for this deviation, a comparison with drift-diffusion simulations is performed.

3.4.5. Transit Time Analysis and Comparison with Drift-Diffusion Simulation

The quasi-static transit time analysis is a convenient method for the investigation of the internal transistor operation. It is based on calculating the response of the charge density to a modulation of the applied voltage. The change of the electron charge stored in the device ΔQ_n is related to a time delay τ_{ec} which limits the current gain at high modulation frequencies. Within the quasi-static approximation, the following relation between f_T and ΔQ_n can be derived [91]:

$$\frac{1}{2\pi f_T} = \frac{\Delta Q_n}{\Delta I_C} = \tau_{ec} \tag{3.17}$$

The total transit time τ_{ec} can be divided into several regional transit times which account for the delay that is caused by charge storage in the specific region of the device:

$$\tau_{ec} = \tau_e + \tau_{eb} + \tau_b + \tau_{bc} + \tau_c \qquad (3.18)$$

Here, τ_e, τ_b and τ_c denote the emitter, base and collector transit time, respectively. The transit times which are related to the EB and BC depletion regions are denoted by τ_{eb} and τ_{bc}. A more detailed description of the quasi-static transit time analysis and the definition of the regional contributions can be found in Section 4.2.

In Figure 3.29, simulated f_T as well as the corresponding regional transit times are plotted as a function of I_C for both hydrodynamic and drift-diffusion simulations. The regional transit times are calculated from 1D cross sections of the quasi-static electron and hole density response Δn and Δp according to the definitions (4.19 – 4.23) given in Section 4.2.1. A vertical cross section through the center of the transistor has been used for the calculation of the transit times. In doing so, charge storage in the external transistor regions is neglected. Nevertheless, the behavior of the transistor is well represented by the transit times which are calculated from the 1D cross section through the center of the device. This can be seen by comparing the 1D and 2D quasi-static transit frequencies which are shown in Fig. 3.29a and Fig. 3.29b together with f_T from small-signal AC simulation. The quasi-static f_T is calculated according to (3.17). In the 1D case, ΔQ_n is given by the integral of the quasi-static electron density response Δn over the 1D cross section in the center of the device, whereas in the 2D case, Δn in integrated over the complete 2D domain. The 1D quasi-static f_T is higher than the corresponding 2D result, mainly because charge storage in the external transistor is omitted. However, both curves show the same dependence of f_T on I_C, which demonstrates that regional transit times of the inner vertical transistor represent the behavior of the whole device. Furthermore, it can be seen that quasi-static f_T agrees well with the results from small-signal AC simulation.

The regional transit times are shown in Fig. 3.29c and Fig. 3.29d as a function of I_C. Based on these figures, the operation of the transistor can be separated into three different regimes. At low injection, the total transit time is dominated by the depletion transit times τ_{eb} and τ_{bc}, which decrease rapidly with increasing I_C. Then, there is an intermediate region around peak f_T in which all transit times change comparatively little with increasing I_C. The dominating transit times in this region are τ_b and τ_{bc}. The following high injection regime is characterized by a strong increase of τ_b, τ_e and τ_c. In Figure 3.29 the transitions between the different regimes are indicated by dotted lines. Although these transitions are gradual, the following criteria are used here to indicate the transition points: The threshold from the low injection regime to the intermediate regime is defined as the current density at which τ_{eb} equals τ_b. The transition between intermediate and high injection regime is defined as the current density at which τ_b equals 80 % of the sum of the remaining transit times ($\tau_e + \tau_{eb} + \tau_{bc} + \tau_c$).

As expected, the drift-diffusion simulations yield a significantly smaller f_T than the hydrodynamic simulations. However, the shape of the simulated $f_T(I_C)$ curve corresponds more closely to the measurement. The difference between HD and DD simulations is primarily caused by the base transit time. The higher τ_b in the intermediate
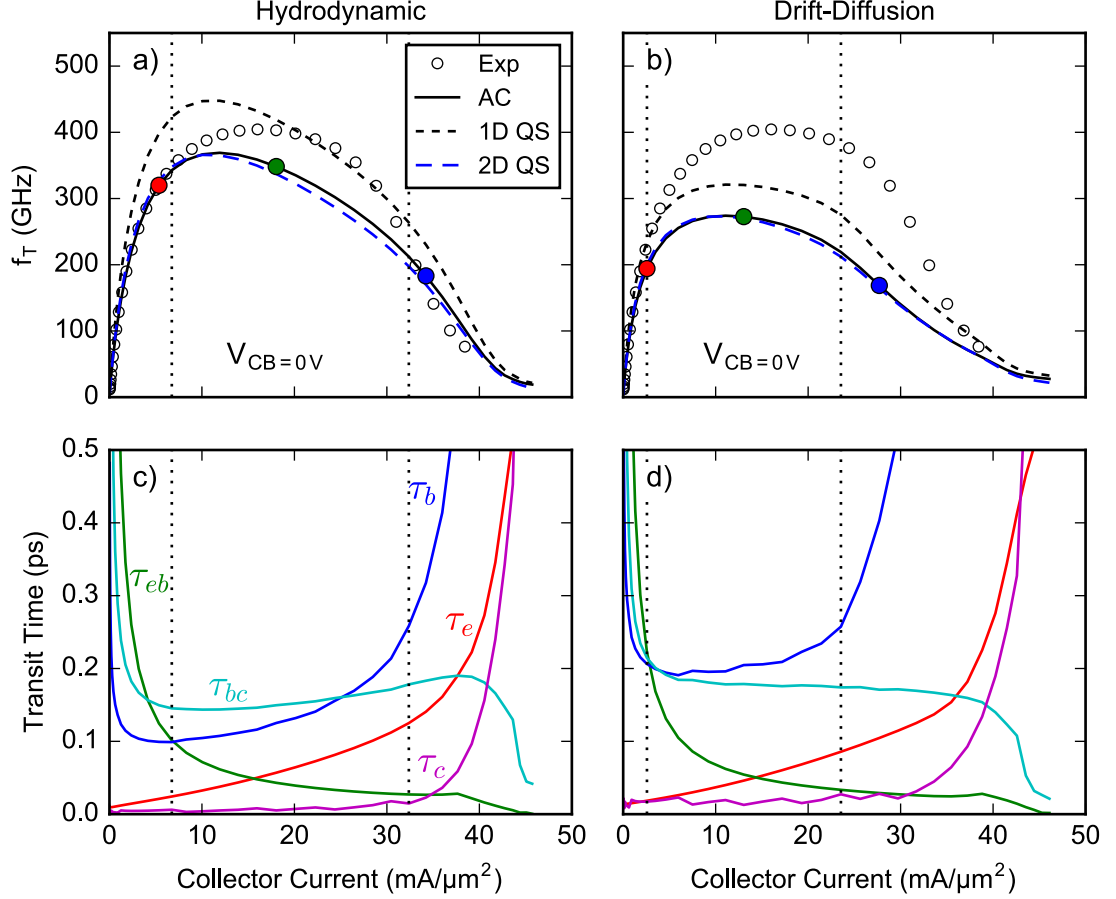
Figure 3.29.: Measured and simulated f_T as a function of I_C are plotted in a) and b). Solid lines show f_T from small-signal AC simulation whereas dashed lines show 1D and 2D quasi-static f_T. 1D f_T is calculated from the quasi-static Δn along a vertical cross section at the center of the transistor. In c) and d) the regional transit times are plotted as a function of I_C. Linearly scaled x-axis is used to highlight differences between measurement and simulation as well as between HD and DD.

regime is responsible for the smaller peak f_T of DD simulations. Furthermore, the high injection regime sets in earlier in case of DD simulation because the rapid increase of τ_b starts at a lower current density than for HD simulation. In addition to the delayed onset of the high injection regime, f_T shows a significant reduction within the intermediate region for HD simulations. The following numbers emphasize the different behavior of DD and HD simulation: At the transition to the high injection regime f_T is decreased by 20 % compared to its peak value in case of DD simulation, whereas it is decreased by 42 % in case of HD simulation. For a further analysis of the difference between HD and DD simulation, cross sections of internal variables are shown in Fig. 3.30 and Fig. 3.31.

Figure 3.30 shows DD simulations of the electron and hole densities (a), the quasi-static response of the electron density Δn relative to the collector current response ΔI_C (b), the electron velocity v_n (c) and the conduction band energy E_C (d) for the three

different operation regimes. The three cases shown in this figure correspond to the three colored markings in Figure 3.29b. The spatial distribution of Δn, which is plotted in Fig. 3.30b, illustrates which location in the device contributes most to the total delay. The integration of $q\Delta n/\Delta I_C$ over the whole device yields the total transit time τ_{ec}. In the low injection regime (red lines), a high peak at the BE junction can be seen, which is related to the change of the electron density at the BE depletion region boundary. This peak decreases with increasing I_C. In the intermediate regime (green lines), this peak has almost vanished. The transit time is now dominated by τ_b and τ_{bc}, which are related to electrons passing the base and the BC depletion region at saturation velocity. Δn in the base is higher than in the BC depletion region due to the lower saturation velocity in SiGe. At high injection, the Kirk effect leads to a strong increase of the hole density at the boundary of the BC depletion region. This results in a high peak in Δn, which is related to the rapid increase of the base transit time τ_b.
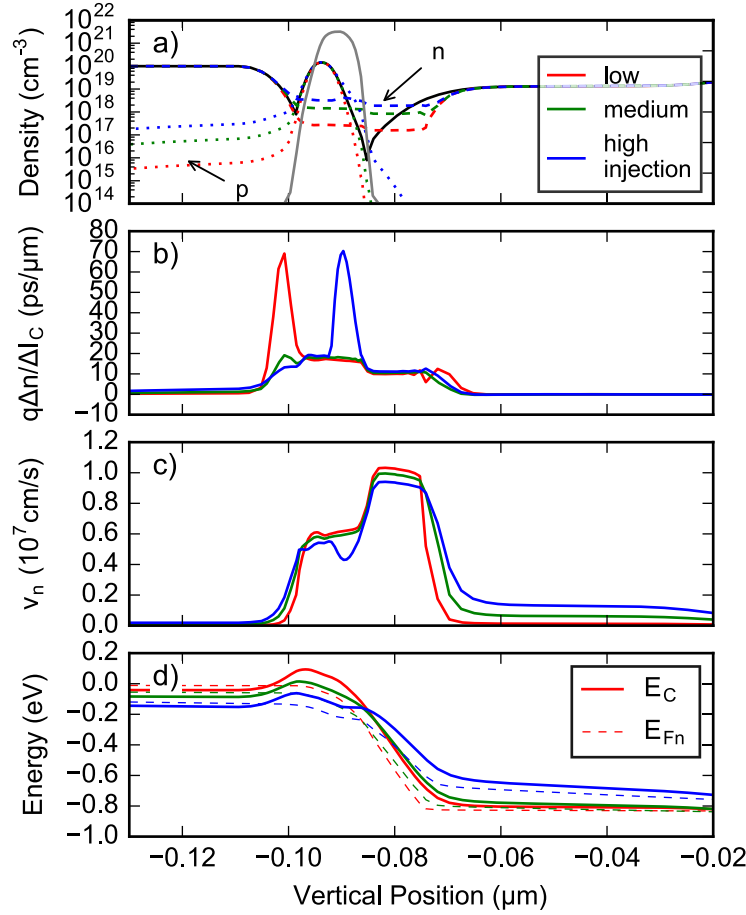
Figure 3.30.: Internal variables from DD simulation for three different operation regimes (marked in Fig. 3.29b by colored dots). a) Net doping as well as electron and hole density. b) Quasi-static change of the electron density. c) Electron velocity. d) Conduction band and quasi-Fermi energy.

The corresponding results from HD simulations are shown in Figure 3.31. Similar to the case of DD simulation, Δn shows a high peak at the BE junction in the low injection regime and another high peak at the BC junction in the high injection regime. Both peaks almost vanish in the intermediate regime. In contrast to the DD results, Δn also varies strongly at the collector side of the BC depletion region in the low injection regime. There is a positive peak on the base-facing side of the depletion region boundary and a negative peak on the collector-facing side. The impact of those peaks on the total transit time is relatively small because they compensate each other when integrating over the device. The reason for this behavior can be understood by looking at the electron density n in the BC depletion region. In case of DD simulation, n is approximately constant inside the space charge region with a sharp boundary to the neutral collector. In case of HD simulation, the electron density is not constant in the depletion region due to the strong variation of the electron velocity within this region (velocity overshoot). The transition between depletion region and neutral collector is much smoother. The change of sign in Δn results from a change of n at the depletion region boundary. With increasing injection, the boundary to the neutral collector gets flatter and reaches deeper into the collector, which is accompanied by a relocation of electrons from the right to the left side.

The relatively strong reduction of f_T within the intermediate regime in case of HD simulation can be explained by a comparison of the electron velocity. Within drift-diffusion transport, v_n cannot become higher than the saturation velocity v_{sat}. Consequently, the electron velocity in the base and the BC depletion region does not change significantly at the different current injection levels shown in Figure 3.30. In contrast, the hydrodynamic transport model accounts for velocity overshoot in the BC depletion region. This is accomplished by modeling velocity saturation as a function of the carrier energy, instead of the electric field or the gradient of the quasi-Fermi potential (cf. Sect. 2.2.2). Electrons that are accelerated in the BC space charge region can achieve high velocities ($> v_{sat}$) before they gain enough energy that saturation sets in. So, in HD simulation, the electron velocity in the base and the BC depletion region still depends on the shape of the conduction band, whereas in DD simulation, it is mainly determined by the saturation velocity. As a result, the electron velocity in the base and the BC depletion region changes significantly between the different current injection levels. The reduction of the electron velocity with increasing I_C, which can be observed in Fig. 3.31c, is responsible for the comparatively strong degradation of f_T in the intermediate regime in case of HD simulation.

3.4.6. Impact of the Parasitics

In this section, the contributions of various device parasitics to the high frequency performance of the HBT are investigated. In addition, the results of device simulation are compared to results that have been obtained by means of compact modeling.
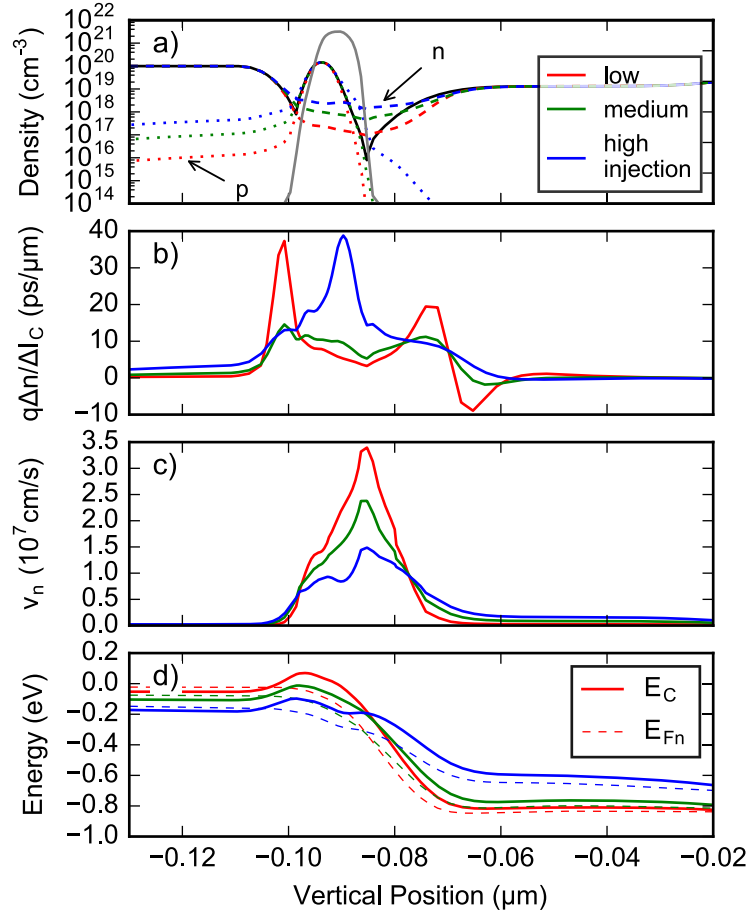
Figure 3.31.: Internal variables from HD simulation for three different operation regimes (marked in Fig. 3.29a by colored dots). a) Net doping as well as electron and hole density. b) Quasi-static change of the electron density. c) Electron velocity. d) Conduction band and quasi-Fermi energy.

Evaluation of the Parasitics by Means of Device Simulation

The impact of the different parasitic elements of the HBT is investigated by a stepwise reduction of the simulation structure towards a 1D transistor. Starting point is the full 2D simulation which includes self-heating, the capacitance of the contact wiring and the parasitic series resistances $R_{E,ex}$ and $R_{B,ex}$. As described above, the capacitances $C_{EB,met}$ and $C_{BC,met}$ and in part also the series resistances are included in the simulation by lumped elements connected to the contacts of the 2D simulation domain. The impact of these parasitic elements on the electrical characteristics is shown in Fig. 3.32 for the medium-size transistor D2.

As a first step, self-heating (SH) is omitted, which leads to an increase of maximum f_T by 58 GHz. The onset of the high injection regime is shifted to a higher current density. The simulated current density at peak f_T agrees much better with the measured value when self-heating is omitted. However, the shape of the $f_T(I_C)$ curve has not changed.
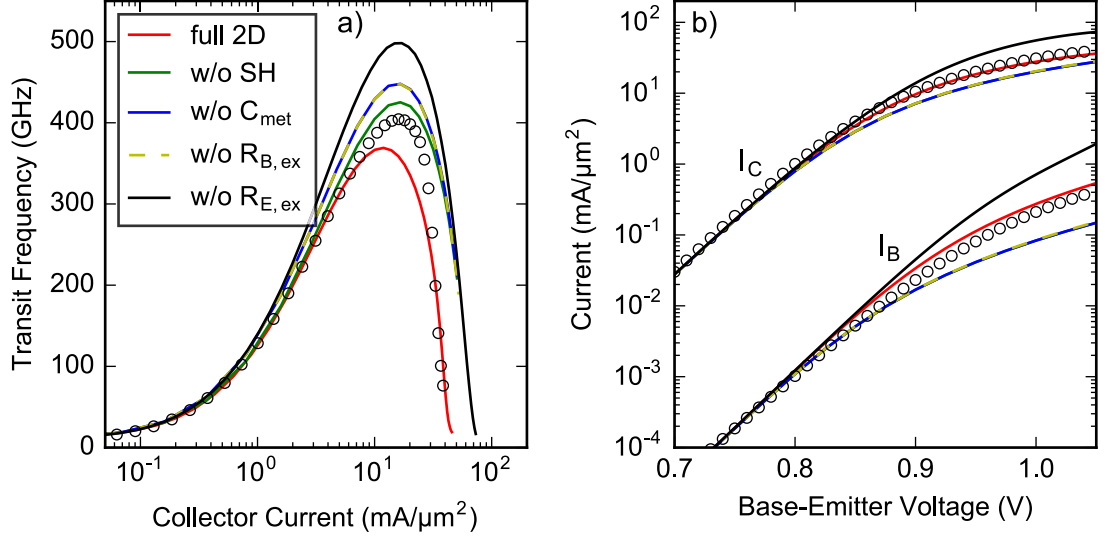
Figure 3.32.: Stepwise elimination of the following parasitic elements from the 2D simulation: Self-heating, contact capacitance, external base and emitter resistance. a) f_T as a function of I_C. b) Gummel characteristics. ("w/o C_{met}" coincides with "w/o $R_{B,ex}$").

Removing the capacitance of the contact wiring leads to an increase of maximum f_T by another 23 GHz. Whereas the elimination of self-heating mainly leads to an increase of f_T at high injection, the reduction of the parasitic capacitance also results in a higher f_T at low currents. Removing the external base resistance $R_{B,ex}$ has no impact on f_T, but a reduction of the emitter series resistance by removing $R_{E,ex}$ leads to a strong increase of f_T by 50 GHz. In Figure 3.33a the peak values of f_T of the different 2D simulation configurations are summarized for all three devices D1, D2 and D3. The impact of self-heating is lowest for the smallest transistor, with an increase of f_T by 8 % in comparison to 20 % for the largest device. On the other hand, the impact of the parasitic capacitance of the contact is most pronounced for the smallest transistor, where f_T increases by 10 % compared to only 5 % for D3. The impact of the additional series resistance is similar for all emitter geometries.

These simulation results show that f_T could be increased by 30 % to 40 %, if self-heating was eliminated and if the transistor terminals were connected ideally. In the next step, the impact of the parasitic capacitance and resistance of the transistor itself will be examined. For this purpose, quasi 1D simulations of the inner transistor have been performed. Using such a quasi 1D simulation domain eliminates all parasitic capacitance of the device perimeter as well as the base resistance. The resistance of the collector well between the contact and the center of the device can also be eliminated but series resistance of the inner collector and the emitter are still included in the 1D domain. Figure 3.34 shows a sketch which illustrates the 2D and 1D simulation configurations. A detailed description how to simulate the inner 1D transistor with Sentaurus TCAD is given in Chapter 4.

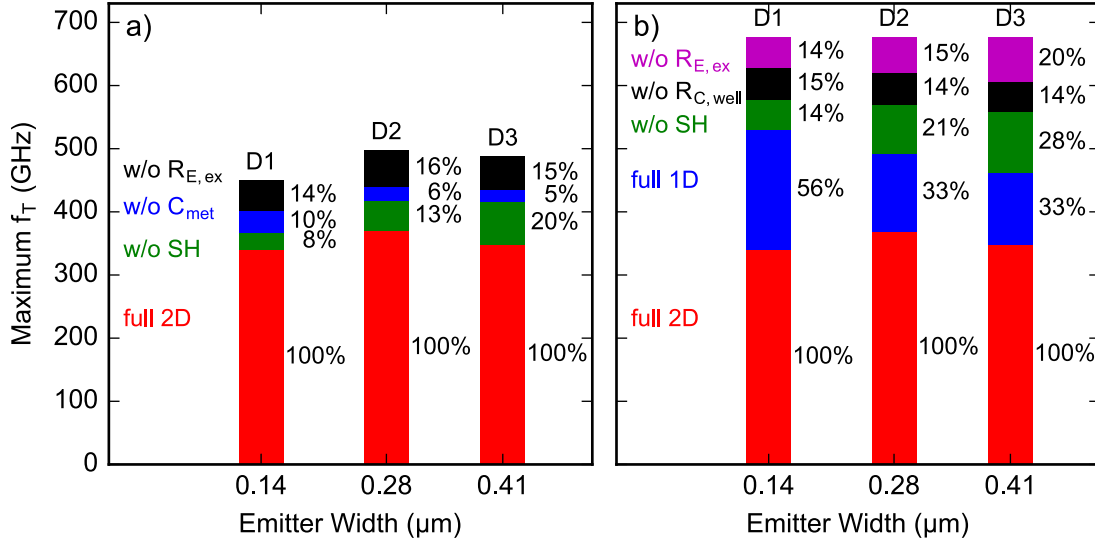Four different configurations of the 1D simulation are compared to the full 2D simu-

Figure 3.33.: Impact of the parasitics for devices D1, D2 and D3. a) Maximum f_T for the different 2D simulation configurations shown in Fig. 3.32. b) Peak f_T for the 1D/2D simulation configurations from Fig. 3.35.
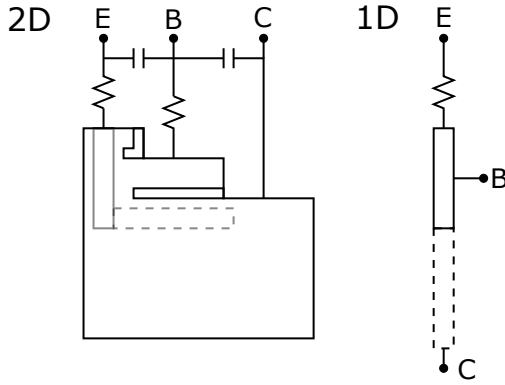

Figure 3.34: Sketch of the 2D and 1D simulation configuration. External parasitics are removed stepwise. Quasi 1D domain represents the inner transistor, where the dashed part accounts for the collector well.

lation in Figure 3.35 for D2. The first configuration, referred to as full 1D, includes self-heating, the additional emitter resistance $R_{E,ex}$ and the resistance of the collector well. The collector well is modeled in the 1D simulation by an extension of vertical profile with a constant doping concentration, as indicated in Figure 3.34. The I_C-V_{BE} characteristic of the full 1D simulation equals the 2D case (Fig. 3.35b). Due to the absence of the parasitic capacitance, the 1D transistor has a much higher f_T than the 2D transistor. Peak f_T increases by 31 % to 54 % depending on the size of the device (Fig. 3.33b). If self-heating is omitted, f_T increases by another 14 % to 27 %. As in the 2D simulation, this increase is due to a delay of the high injection regime. In a next step, the resistance of the collector well $R_{c,well}$ is eliminated by reducing the 1D domain to the inner transistor. This results in an increase of f_T by about 14 %. This increase is mainly caused by a delay of the Kirk effect, because $R_{c,well}$ leads to a reduction of the voltage at the BC depletion region. If also $R_{E,ex}$ is removed and only the inner transistor is

considered, a maximum f_T of 692 GHz is obtained. A comparison of the regional transit times of the different 1D and 2D configurations is given in Section 4.2.3.
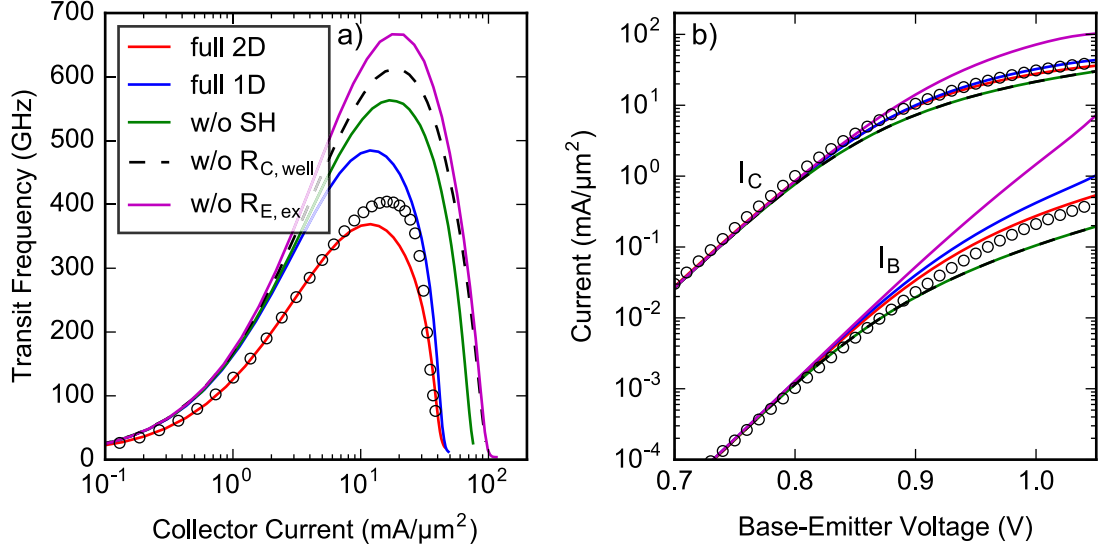
Figure 3.35.: Comparison between 1D and 2D simulation. 1D simulation is shown with and without self-heating and external emitter resistance.

Evaluation of the Parasitics by Means of Compact Modeling

Additionally, the impact of device parasitics on f_T of the transistors D1, D2 and D3 has been investigated in collaboration with TU Dresden using the HICUM compact model. The results of this study have been published in [76]. HICUM is a geometry scalable physics-based compact model for bipolar transistors. Its geometry scaling capabilities allow to investigate modifications of the transistor configuration [92]. Furthermore, it can be used to extract 1D transistor characteristics from actual devices. A set of scalable HICUM model parameters has been extracted from measured characteristics of transistors D1, D2 and D3. The impact of the different external device parasitics can be examined by successively switching them off in the compact model. The corresponding f_T-characteristics are shown in Fig. 3.36 for device D2 with $W_E = 0.28\,\mu$m.

A very good agreement between measured and simulated f_T can be observed, which demonstrates the high modeling capabilities of HICUM. If self-heating (SH) is turned off in the model, f_T increases by about 60 GHz. If additionally, the external contributions to the base-emitter and base-collector capacitances are removed, f_T increases by another 150 GHz. A similar increase by another 180 GHz is achieved if also the external contributions to the emitter, base and collector resistances as well as the internal base resistance are removed. HICUM also covers effects associated with the device perimeter such as collector current crowding and emitter perimeter injection. For this reason, important quantities of the internal transistor such as the transfer current and the transit times depend on the ratio between the area and the perimeter of the emitter [93].
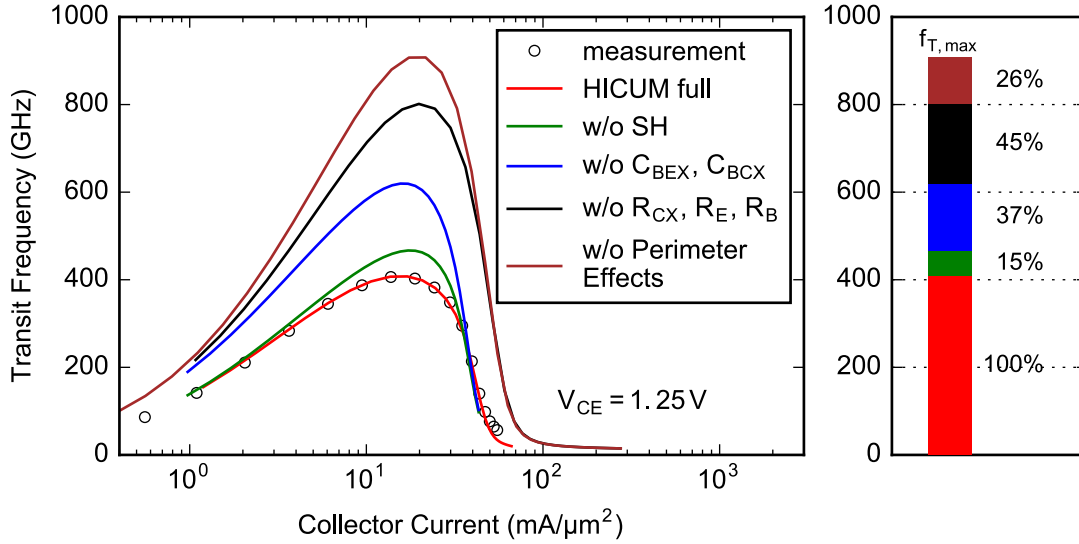
Figure 3.36.: f_T-characteristics of D2 simulated with the HICUM compact model. External parasitics are successively switched off.

The characteristics of a 1D transistor are thus obtained by elimination of the perimeter effects in the parameters of the internal transistor. This leads to a further increase of f_T by about $100\,\mathrm{GHz}$ (Fig. 3.36). De-embedding of the internal 1D transistor from devices D1, D2 and D3 by means of the HICUM compact model yields a maximum f_T of about $900\,\mathrm{GHz}$, which is significantly higher than the result of the 1D device simulation. A possible reason for this strong discrepancy might be a wrong partition between the internal transit time and the external parasitics in the compact model as suggested in [94]. In this publication, 1D characteristics are presented which have been extracted from similar transistors. Due to adaptions in the compact model a significantly smaller 1D transit frequency has been obtained.

3.5. Summary

In Chapter 3, measured electrical characteristics of SiGe HBTs with an advanced vertical doping profile and a measured peak f_T over $400\,\mathrm{GHz}$ have been compared to the results of 2D device simulation. The simulations are based on results of a comprehensive experimental characterization of the HBTs. The vertical doping profile has been determined by a combination of different experimental techniques in order to achieve the best available accuracy.

A good agreement between the measured and simulated Gummel characteristics could be achieved by a slight adjustment of the Ge-dependent bandgap parameters, that have been determined in the previous chapter. Such an adjustment is necessary for two reasons: First, even small errors in the measured Ge profile can influence the collector current noticeably. Second, as a result of process-inherent stress, the effective bandgap also depends on the layout of metal interconnects near the device.

3. Comparison of 2D Simulation and Experiment

A quite good agreement between measured and simulated output characteristics could be achieved by adjusting the heat flux parameter f^{hf}.

The simulated collector current ideality showed a considerable deviation from the measurement, which is caused by a bandgap peak in the base-emitter depletion region. This peak is a result of the doping induced bandgap narrowing. A much better agreement between measured and simulated ideality has been achieved by replacing the doping induced bandgap narrowing by a uniform bandgap reduction.

The simulated transit frequency deviates significantly from the measurement. In the simulation, f_T starts to fall-off at a lower current density than in the measurement, which results in an underestimation of the maximum f_T. The analysis of the regional transit times suggests that electron transport through the base is not described sufficiently well by the hydrodynamic simulation. It has been shown that the reduction of the electron velocity in the base and the BC depletion region with increasing current density is responsible for the premature fall-off of f_T.

The impact of the parasitic resistances and capacitances has been investigated by a stepwise reduction of the simulation domain. The ideal inner 1D transistor shows f_T of $692\,\mathrm{GHz}$, which is about $90\,\%$ higher than for the full device. About $40\,\%$ of the increase are related to the capacitance of the device perimeter and the metal contacts. $20\,\%$ are related to self-heating and about $15\,\%$ each are related to emitter and collector series resistance. This analysis shows that the advanced vertical profile of the investigated devices still offers a significant potential for further performance improvements. It also shows that the reduction of device parasitics is essential to reach the highest f_T values.

4. Impact of Vertical Profile Variations on the Transit Frequency

In this chapter, an experimental and numerical investigation of a scaled vertical transistor profile is given. Different variations of the vertical profile presented in Chapter 3 are examined regarding their RF-performance. One-dimensional quasi-static transit time analysis is used to understand the impact of profile changes and to identify performance limiting factors.

This chapter is organized as follows: The next section describes how quasi 1D simulations of HBTs are realized in Sentaurus TCAD. Then the method of quasi-static transit time analysis as well as the regional partitioning scheme are introduced in Section 4.2. After that, the different profile variations are examined, namely, the variation of the emitter-base junction width (Sect. 4.3.1), two different selectively implanted collectors and a variation of the position of the heterojunction (Sect. 4.3.2 and Sect. 4.3.3). Additionally, a comparison of the experimental profile with a hypothetical profile from the SiGe HBT scaling roadmap is made is Section 4.3.4. Finally, in Section 4.4, the question is discussed if the discrepancy between measured and simulated f_T, that has been observed in the Chapter 3, could be caused by an inaccurate determination of the vertical profile. Additionally, potential improvements of f_T by further optimizations of the vertical profile are examined.

4.1. Simulation of the inner 1D Transistor in Sentaurus Device

Sentaurus TCAD does not support one-dimensional simulation of bipolar transistors. Alternatively, a rectangular 2D simulation domain can be used to represent the vertical 1D transistor. An essential prerequisite for such simulations is the appropriate definition of the boundary conditions for the base contact. Usually, ohmic contacts are modeled by Dirichlet boundary conditions [15, 95]. This means that the carrier densities n and p, as well as the corresponding temperatures T_n and T_p are pinned to their equilibrium values at the contacts. The electron and hole densities are then defined by the following conditions:

$$n - p - N_D + N_A = 0 \tag{4.1}$$
$$np - n_{i,eff}^2 = 0 \tag{4.2}$$

The first equation is the charge neutrality condition, the second equation is the *pn*-product in thermal equilibrium. The carrier temperatures at the ohmic contact are

fixed at the value of the lattice temperature T_L,

$$T_n = T_p = T_L \tag{4.3}$$

The quasi-Fermi potentials of electrons and holes, Φ_{Fn} and Φ_{Fp} are set to the value of the applied voltage V_{app}. The electrostatic potential at the boundary is then given by

$$\Phi = V_{app} + V_{bi}, \tag{4.4}$$

with the built-in potential V_{bi} which is calculated from (2.46) or (2.50). In the case of Boltzmann statistics, an analytical solution for the boundary condition exists, which is given by

$$n = \frac{\sqrt{C^2 + 4n_{i,eff}^2} + C}{2}, \tag{4.5}$$

$$p = \frac{\sqrt{C^2 + 4n_{i,eff}^2} - C}{2}, \tag{4.6}$$

$$\Phi = V_{app} + \frac{kT}{q}\text{asinh}\left(\frac{C}{2n_{i,eff}}\right), \tag{4.7}$$

Here, C denotes the net doping $C = N_D - N_A$. When defining the simulation domain, one has to make sure that ohmic boundary conditions are satisfied at the contacts. Therefore the contacts have to be far enough away from the junctions of the transistor. The minority carrier density and the electron temperature in the intrinsic base are far away from equilibrium, which is why ohmic boundary conditions do not apply there. Thus, in a 1D simulation of the inner transistor, the base contact cannot be modeled as an ohmic contact. The usual way to simulate the base contact in 1D simulation is to set only the quasi-Fermi potential of holes to the value of the applied voltage. Sentaurus TCAD does not allow to pin the quasi-Fermi potential in a 1D simulation in such a way. However it is possible to emulate a 1D transistor by a 2D simulation domain. A sketch of this quasi 1D domain is shown in Fig. 4.1. The transistor is simply modeled by a rectangular region with two small extensions for the base contacts at both sides. Instead of using the equilibrium values n_0 and p_0 at the contacts, the device simulator allows to define a recombination velocity for electrons and holes at the contact so that n an p are defined by [37]

$$J_n \cdot \hat{n} = qv_n(n - n_0), \tag{4.8}$$

$$J_p \cdot \hat{n} = qv_p(p - p_0), \tag{4.9}$$

where v_n and v_p are the electron and hole recombination velocities and $\hat{n}$ is the normal vector to the contact surface. This option can be used to model the base contact in the quasi 1D domain by setting the electron recombination velocity to zero. Thereby, recombination of electrons at the contact is suppressed, while the quasi Fermi potential of the holes can still be modulated by the applied voltage. However, it is not possible to change the boundary condition for the electron temperature in such a way. So, to

prevent that the electron temperature in the base is distorted by inappropriate boundary conditions for T_n, the base contact is not directly placed at the rectangular region which represents the inner transistor. Instead, small base extensions are added at both sides and the contacts are placed at these extensions. Inside these regions, energy transport is switched off by setting the energy flux coefficient r to zero (cf. (2.45)). In this way, the inner transistor is thermally isolated from the base contact and the fact that T_n equals 300 K at the contact does not affect the electron temperature in the inner transistor.
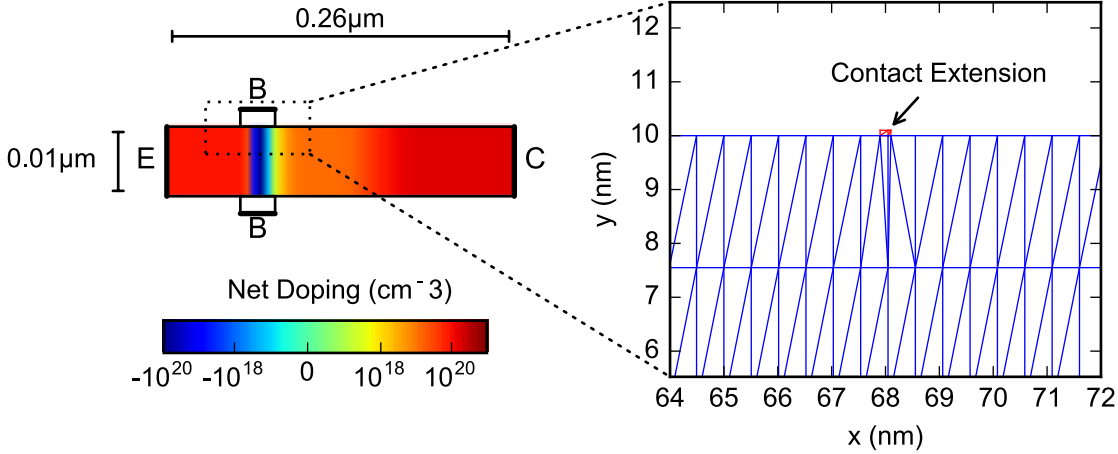
Figure 4.1.: Sketch of the 2D domain which is used to simulate the inner transistor (LHS). The base extensions are not in scale. A part of the actual simulation grid is shown on the RHS. The doping is extended in y-direction into the base extension.

The width of the quasi 1D domain should be small compared to its length, as otherwise the potential drop along the base becomes significant. Here, a width of $0.01\,\mu$m is used. In this case the difference of the quasi Fermi potentials between the base contacts and center of the base can be neglected and the current transport can be regarded as one-dimensional.

4.2. Quasi-Static Transit Time Analysis

Quasi-static transit time analysis is a useful method to investigate the performance of vertical transistor profiles. It is based on the concept of charge control which links the change of the terminal bias voltages to the change of the internal carrier densities [96]. Internally, the transistor responds to an AC input signal with a modulation of the charge which is stored in the device. This change of stored charge causes a time delay which limits the gain of the transistor at high frequencies. A measure of this delay is the transit time τ_{ec}, which is defined by

$$\tau_{ec} = \frac{1}{2\pi f_T}. \tag{4.10}$$

A relation between τ_{ec} and the charge in the transistor can be derived by the following consideration [97]: The small-signal current gain in common emitter configuration is

Figure 4.2: Comparison of f_T from quasi-static simulation (QS) and from small-signal AC simulation.
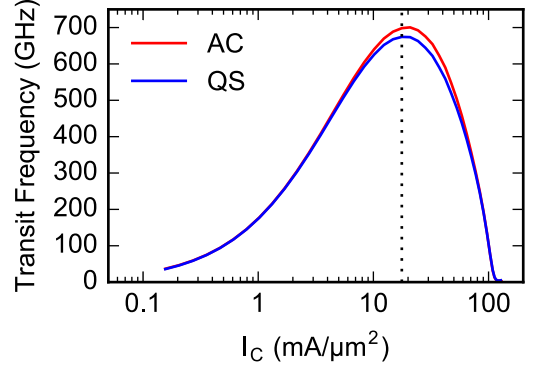
given by

$$h_{21} = \frac{\Delta I_C(\omega)}{\Delta I_B(\omega)}\bigg|_{V_{CE}=\text{const.}}. \tag{4.11}$$

Here, ΔI_C and ΔI_B denote the change of the collector and base current respectively, that is caused by a voltage change ΔV_{BE}. At high frequencies, one can assume that ΔI_B is determined solely by the modulation of holes stored in the transistor because recombination can be neglected at time scales of $1/\omega$:

$$\Delta I_B = \frac{d}{dt}\Delta Q_p = i\omega \Delta Q_p \tag{4.12}$$

Here, ΔQ_p is the small-signal hole charge. The time derivative is evaluated by writing ΔQ_p in the form $|\Delta Q_p|e^{i\omega t}$. From equations (4.10)-(4.12) and the definition of the transit frequency $|h_{21}(f_T)| = 1$, one obtains the desired relation between τ_{ec} and the stored charge in the transistor:

$$\tau_{ec} = \frac{\Delta Q_p}{\Delta I_C}\bigg|_{V_{CE}=\text{const.}} \tag{4.13}$$

Due to charge neutrality, ΔQ_p can be replaced in (4.13) by the electron charge ΔQ_n:

$$\tau_{ec} = \frac{\Delta Q_n}{\Delta I_C}\bigg|_{V_{CE}=\text{const.}} \tag{4.14}$$

Within the quasi-static approximation, one assumes that ΔQ_n and ΔI_C are independent of the frequency ω. This allows one to calculate the transit time of the transistor from steady-state device simulation by a small change of the DC voltage. A comparison of f_T calculated from (4.14) in the quasi-static approximation with the results of frequency dependent small-signal AC simulation is shown in Fig. 4.2. In case of the AC simulation, f_T is extrapolated from simulated h_{21} at a frequency of 40 GHz. One can see that the quasi-static approximation only leads to a small underestimation of peak f_T. Thus it can be regarded as a valid approximation for the investigation of the vertical profile.

4.2.1. Regional Partition of the Transit Time

With the aid of device simulation it is possible to calculate the spatial distribution of the stored charge in the device. Such simulations provide detailed information on the regional contributions to the total transit time τ_{ec} [15]. In a 1D simulation, the stored charge can easily be partitioned into different regional transit time components [98, 99]. Such a charge partitioning helps to interpret simulation results and to identify performance limiting factors in the vertical doping profile.
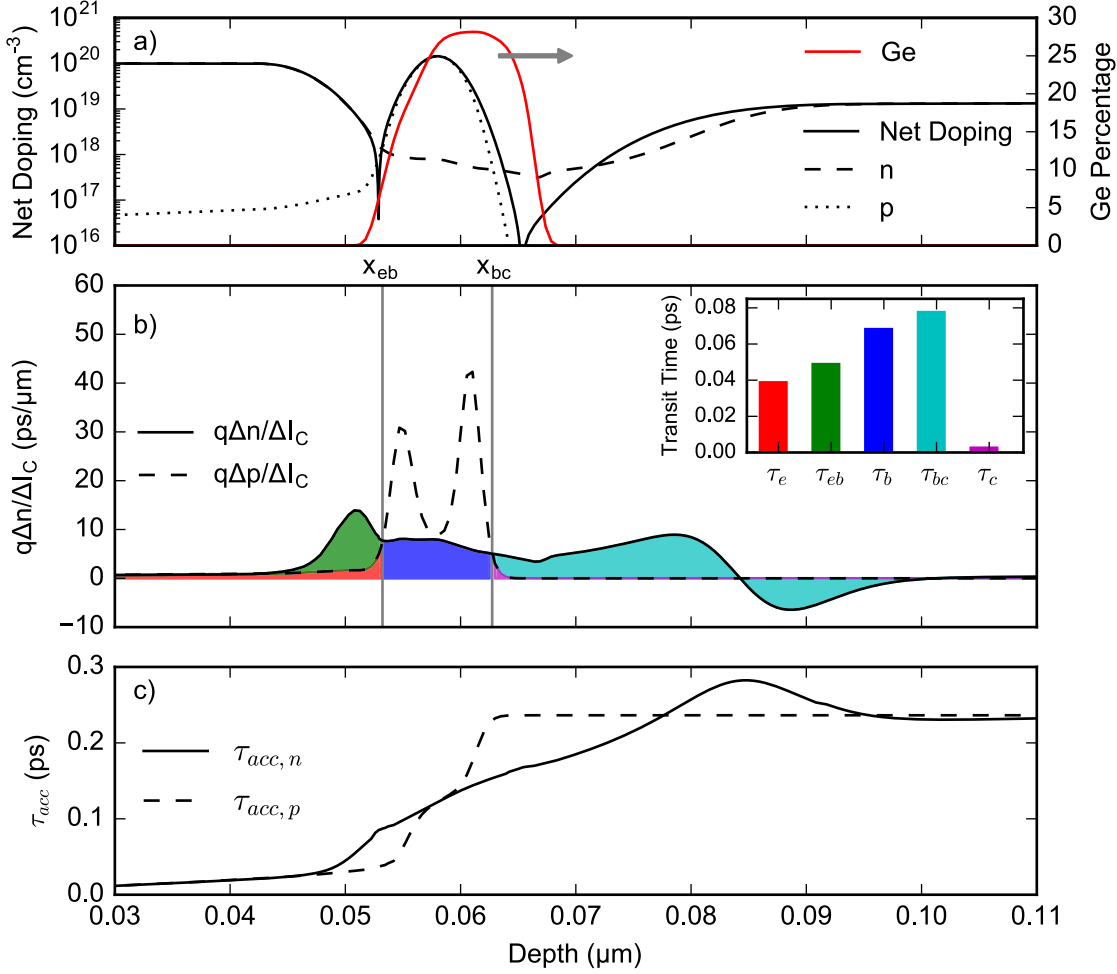
Figure 4.3.: Distribution of the transit time in the 1D transistor at peak f_T. a) Vertical transistor profile and carrier densities n and p. b) Quasi-static density response (divided by the collector current response). Inset shows the regional contributions as indicated by colored areas. c) Accumulated transit time of electrons $\tau_{acc,n}$ and holes $\tau_{acc,p}$.

The regional distribution of the transit time is illustrated in Figure 4.3. The total transit time τ_{ec} can be calculated from the small-signal change of the electron density

Δn as

$$\tau_{ec} = \frac{\Delta Q_n}{\Delta I_C} = \frac{q}{\Delta I_C} \int_0^l \Delta n(x)\, dx. \tag{4.15}$$

The integral runs over the whole device from the base contact at $x = 0$ to the collector contact at $x = l$. The equivalent relation for holes is given by

$$\tau_{ec} = \frac{\Delta Q_p}{\Delta I_C} = \frac{q}{\Delta I_C} \int_0^l \Delta p(x)\, dx, \tag{4.16}$$

with the small-signal change of the hole density Δp. Additionally, one can define the accumulated electron transit time $\tau_{acc,n}$ at position x by

$$\tau_{acc,n}(x) = \frac{q}{\Delta I_C} \int_0^x \Delta n\, dx \tag{4.17}$$

or, equivalently the accumulated hole transit time

$$\tau_{acc,p}(x) = \frac{q}{\Delta I_C} \int_0^x \Delta p\, dx. \tag{4.18}$$

In this work, the partitioning scheme of van den Biesen [98] is used to investigate the regional distribution of the time delay in the transistor. Within this approach the total transit time is divided into five contributions ($\tau_{ec} = \tau_e + \tau_{eb} + \tau_b + \tau_{bc} + \tau_c$). The advantage of the partition scheme of van den Biesen compared to others form the literature [99] is that only two boundaries are needed for the regional partition of the transit time. These boundaries, which are labeled by x_{eb} and x_{bc}, are defined by the intersections of Δn and Δp as shown in Figure 4.3b. The various transit times are defined as follows:

$$\tau_e = \frac{q}{\Delta I_C} \int_0^{x_{eb}} \Delta p\, dx \tag{4.19}$$

$$\tau_{eb} = \frac{q}{\Delta I_C} \int_0^{x_{eb}} (\Delta n - \Delta p)\, dx \tag{4.20}$$

$$\tau_b = \frac{q}{\Delta I_C} \int_{x_{eb}}^{x_{bc}} \Delta n\, dx \tag{4.21}$$

$$\tau_{bc} = \frac{q}{\Delta I_C} \int_{x_{cb}}^{l} (\Delta n - \Delta p)\, dx \tag{4.22}$$

$$\tau_c = \frac{q}{\Delta I_C} \int_{x_{bc}}^{l} \Delta p\, dx \tag{4.23}$$

The emitter, base and collector transit times τ_e, τ_b and τ_c are related to locally compensated minority charge in the respective region. They are also referred to as diffusion charging times. The base-emitter and base-collector transit times τ_{eb} and τ_{bc} are depletion charging times, which are related to locally uncompensated charge. The spatial distribution of Δn and Δp (at peak f_T) is shown in Fig. 4.3b. The integrals defining the regional transit times (4.19)-(4.23) are indicated by colored areas. Note that the negative Δn is a manifestation of the extension of the BC space-charge region into the collector [15]. This leads to a maximum of $\tau_{acc,n}$ at the position where Δn changes sign (see Fig. 4.3c).

Other methods for calculating the regional transit time contributions additionally distinguish between mobile charge in space-charge regions (SCR) and in neutral regions (NR) [100, 101]. However, finding an unambiguous and meaningful definition of the SCR boundaries at high injection is a cumbersome task [102, 99]. Within the partition scheme of van den Biesen, the delay $\tau_{eb,N}$ due to neutral charge storage in the base-emitter SCR is included in τ_e and τ_b. Thus, the emitter delay time τ_e consists of τ_e^*, the delay due to minority charge storage in the *neutral* emitter and a portion of $\tau_{eb,N}$. In a transparent emitter, where recombination at the contact predominates against recombination in the neutral region, τ_e^* can be estimated by [98]

$$\tau_e^* = \frac{1}{\beta} \left(\frac{d_e}{v_{rec,h}} + \frac{d_e^2}{2D_h} \right), \tag{4.24}$$

where d_e is the distance between the emitter contact and the base-emitter SCR, D_h is the hole diffusivity and $v_{rec,h}$ is the hole recombination velocity at the emitter contact.

Neutral charge storage in the base-collector depletion region, $\tau_{bc,N}$, is included in τ_b and τ_c. Similar to τ_e, the base transit time τ_b consists of the transit time of the *neutral* base τ_b^* and additional contributions of $\tau_{eb,N}$ and $\tau_{bc,N}$. Assuming a uniform base doping and a constant Ge concentration, the neutral base transit time can be estimated by [103]

$$\tau_b^* = \frac{w_b^2}{2D_n} + \frac{w_b}{v_{sat}} \tag{4.25}$$

with the base width w_b and the electron diffusivity D_n.

The depletion charging times τ_{eb} and τ_{bc} are due to uncompensated charge at the base-emitter and base-collector junction. With the corresponding depletion capacitances C_{jEB} and C_{jBC}, they can be expressed as [98]

$$\tau_{eb} = \frac{C_{jEB}}{g_m} \tag{4.26}$$

and

$$\tau_{bc} = \frac{C_{jBC}}{g_m} + \frac{x_d}{2v_{sat}}. \tag{4.27}$$

Here, x_d is the length of the base-collector depletion region and v_{sat} is the electron saturation velocity. The second term in (4.27) accounts for uncompensated mobile charge in base-collector SCR. Due to the high doping at the BE junction and the small size of the depletion region, the impact of uncompensated mobile charge in the BE depletion region is neglected in (4.26).

4.2.2. Small-Signal Equivalent Circuit

An expression for τ_{ec} can also be derived from the small-signal equivalent circuit model shown in Figure 4.4. The small-signal current gain h_{21} in common-emitter configuration under the condition of a short-circuit load is then given by [15]

$$h_{21} = \frac{i_c}{i_b} = \frac{g_m - j\omega C_{BC}}{1/r_\pi + j\omega(C_{EB} + C_{BC})}. \tag{4.28}$$

4. Impact of Vertical Profile Variations on the Transit Frequency

It is usually assumed that $g_m \gg \omega C_{BC}$ at frequencies of practical interest. The current gain then simplifies to

$$h_{21} = \frac{g_m}{1/r_\pi + j\omega(C_{EB} + C_{BC})}. \tag{4.29}$$

Further, one assumes that the denominator of equation (4.29) is dominated by the imaginary part at high frequencies so that h_{21} can be approximated by

$$h_{21} = \frac{g_m}{j\omega(C_{EB} + C_{BC})}. \tag{4.30}$$

The transit frequency is defined as the frequency at which $|h_{21}|$ becomes unity, which leads to

$$\frac{1}{2\pi f_T} = \frac{C_{EB} + C_{BC}}{g_m} \tag{4.31}$$

The capacitances C_{EB} and C_{BC} can be split into depletion and diffusion capacitances. The sum of the diffusion capacitances is usually expressed in terms of the forward transit time τ_f:

$$C_{dEB} + C_{dBC} = \tau_f g_m. \tag{4.32}$$

With this definition, equation (4.31) can be expressed in the commonly used form

$$\tau_{ec} = \frac{1}{2\pi f_T} = \tau_f + \frac{C_{jEB} + C_{jBC}}{g_m}, \tag{4.33}$$

with the depletion capacitances C_{jEB} and C_{jBC}. The forward transit time can now be related to the regional transit times defined in the previous section. By comparing (4.33) to (4.27) and (4.26) one obtains

$$\tau_{ec} = \frac{1}{2\pi f_T} = \tau_e + \tau_b + \tau_c + \tau_{bc,0} + \frac{C_{jEB} + C_{jBC}}{g_m}, \tag{4.34}$$

where $\tau_{bc,0} \approx x_d/v_{sat}$ is the current-independent part of τ_{bc}.
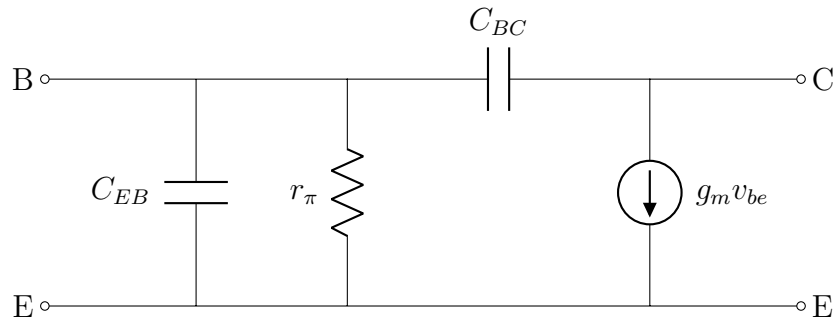
Figure 4.4.: Simple high-frequency small-signal equivalent circuit of a bipolar transistor.

The forward transit time τ_f includes all the delay components that are independent of the current density at low and moderate current injection. By plotting $(2\pi f_T)^{-1}$ as a
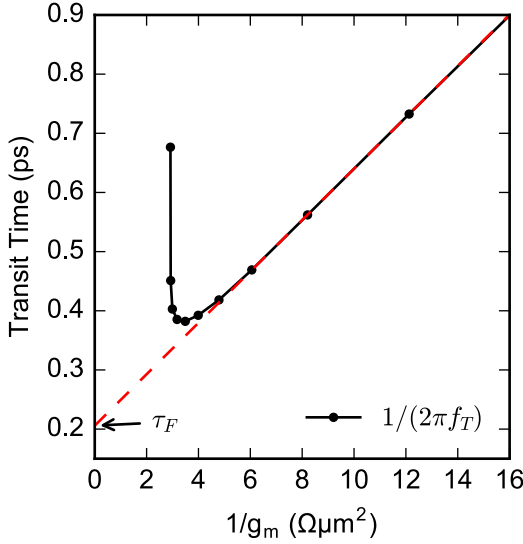
Figure 4.5.: Measured $(2\pi f_T)^{-1}$ as a function of inverse transconductance. Extrapolation from linear region yields τ_f.
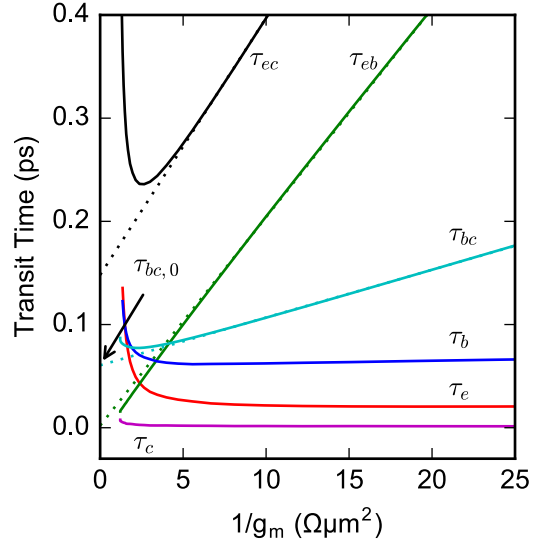
Figure 4.6.: Simulated transit time τ_{ec} and its regional contributions as a function of inverse g_m.

function of g_m^{-1} (see Fig. 4.5), both τ_f and the total junction capacitance can easily be extracted from f_T-measurements. The total capacitance $C_{jEB} + C_{jBC}$ is given by the slope of a linear regression line and τ_f is given by its y-intersection.

In Fig. 4.6 the simulated regional transit times are plotted as a function of inverse transconductance g_m^{-1}. At low injection, the total transit time is dominated by τ_{eb} due to the large depletion capacitance of the base-emitter junction. With increasing collector current, the depletion charging times decrease and the contributions of τ_e and τ_b become relevant. At high injection, both τ_e and τ_b increase rapidly leading to the strong decrease of f_T at high currents. The collector transit time τ_c has no significant impact on τ_{ec}.

4.2.3. Comparison of 1D and 2D Simulation

In this chapter, the impact of changes of the vertical doping profile on f_T is investigated by means of 1D simulation in order to highlight their impact on the internal transit times. These simulations are compared qualitatively with experimental results. It is important to know the differences between the transit times of a realistic 2D device and an idealized 1D transistor, to draw reasonable conclusions from such a comparison. In this section, these differences are investigated by looking at the transit times of the different 2D and 1D simulation configurations which have been investigated in Section 3.4.6. Figure 4.7 shows the regional transit times which correspond to the transit frequencies shown in Figure 3.35. The results from full 2D simulations including all parasitics are plotted in Fig. 4.7a and full 1D simulations are shown in Fig. 4.7b. Transit times from 1D simulation, where self-heating, collector well resistance and emitter resistance are successively removed, are plotted in Fig. 4.7c, Fig. 4.7d, and Fig. 4.7e, respectively. As

in Section 3.4.5, the regional transit times of the 2D simulations are calculated from a 1D cross section through the center of the transistor, so that charge storage in the external regions is not included.

The main differences between the transit times from full 2D and full 1D simulation are that τ_b and τ_{bc} are smaller and the increase of τ_c is delayed in the 1D case. The main reason is that the voltage at the BC junction is slightly smaller in the center of the 2D transistor due to the higher effective collector resistance.

Elimination of self-heating leads to a significant shift of the high injection regime where τ_b and τ_e increase rapidly. Furthermore, τ_{bc} is almost constant for $I_C > 10\,\mathrm{mA}/\mu\mathrm{m}^2$ in the isothermal simulation. This indicates that the increase of τ_{bc} is caused by self-heating due to an increase of the collector resistance with higher temperature. Removing the collector well resistance $R_{C,well}$ leads to a further shift of the high injection regime and to a small reduction of τ_{bc}. If also the external emitter resistance $R_{E,ex}$ is removed, a further reduction of τ_{bc} can be observed due to the higher g_m.

Simulations which only account for the internal 1D transistor, as shown in Fig. 4.7e, are used in the following to investigate the impact of variations of the vertical doping profile. By using 1D simulation, one avoids that the impact of the profile change on the transit time is concealed by the influence of device parasitics. The effect of the profile change is more pronounced in a 1D simulation which helps to interpret the experimental results. The f_T characteristics in the following section have been measured and simulated at a constant V_{CE} of $1.25\,\mathrm{V}$ because this setup usually yields the highest f_T. Corresponding regional transit times are shown in Fig. 4.7f. Compared to the simulations with $V_{BC} = 0\,\mathrm{V}$, a lower base transit time is achieved at medium and high current densities.
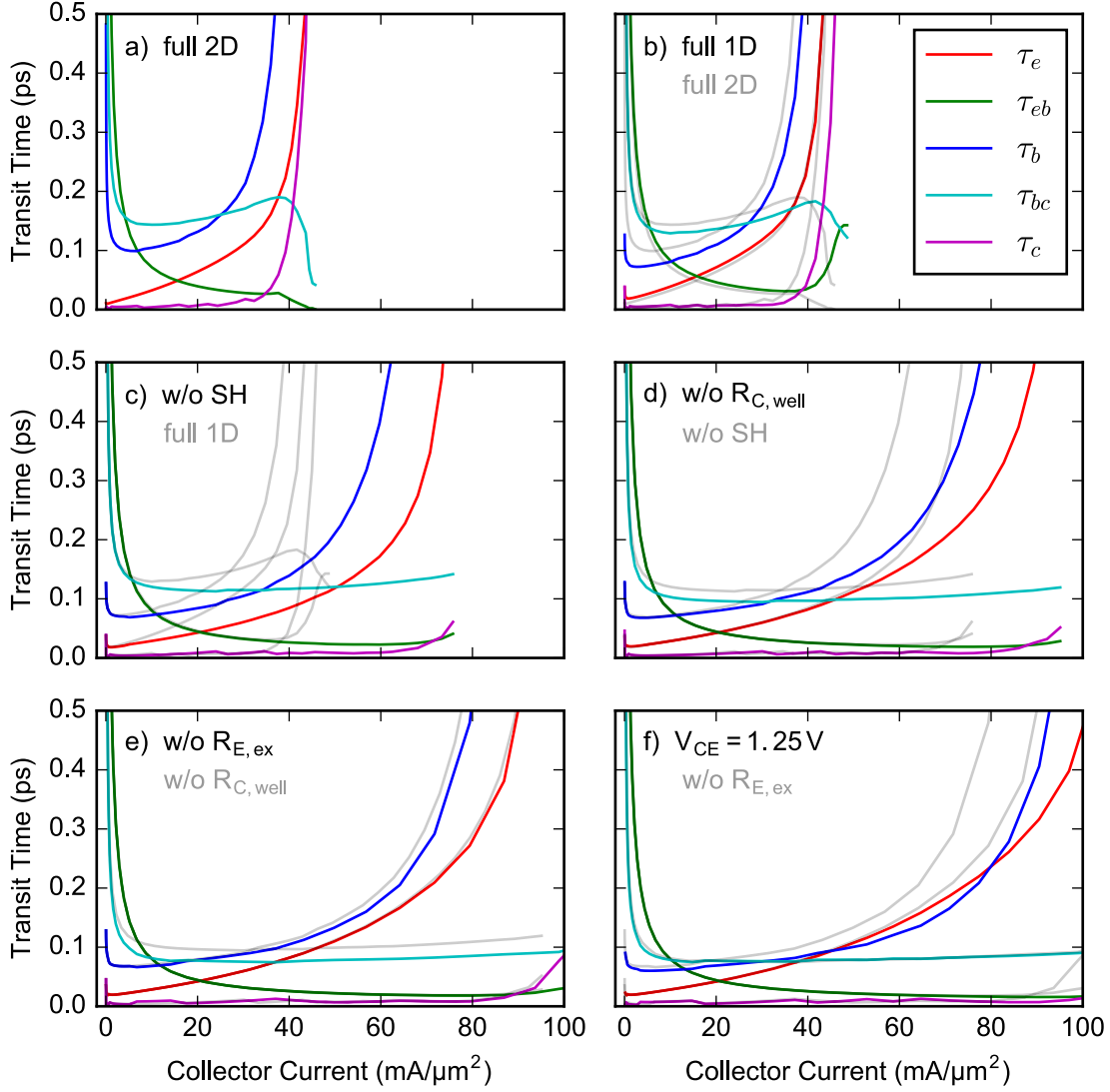
Figure 4.7.: Regional transit times for different configurations of 1D and 2D simulation. a) Full 2D simulation with all parasitics. b) 1D simulation including self-heating, the collector well and $R_{E,ex}$. These parasitics are successively removed in c, d and e. f) Same simulation setup as in e but with $V_{CE} = 1.25\,\text{V}$ instead of $V_{CB} = 0\,\text{V}$.

4.3. Examples of Vertical Profile Variations

In the following the impact of various profile variations on f_T is investigated both by simulation and experiment. For this purpose, different variations of the vertical profile from the transistors investigated in Chapter 3 have been been fabricated. In addition to the experimental results, 1D transit time simulation is used to elucidate how the profile variations affect the internal transistor operation.

4.3.1. Impact of the Base-Emitter Junction Width

Three wafer splits with a variation of the emitter-base junction width have been fabricated. This profile variation was obtained by changing the thickness of the undoped silicon cap layer which is deposited after the SiGe base. The wafer splits with a measured cap thickness of 13.4 nm (P0), 14.9 nm (P1) and 16.4 nm (P2) show a peak f_T of 416 GHz, 402 GHz and 381 GHz (see Fig. 4.8b). In Figure 4.8a, the measured transit time τ_{ec} is plotted as a function of inverse g_m. The extrapolated transit times are 0.205 ps, 0.209 ps and 0.232 ps and the corresponding capacitances which result from the slope of the curves are 44 fFμm^{-2}, 41 fFμm^{-2} and 37 fFμm^{-2}, respectively. The extrapolation of τ_{ec} to infinite g_m is done from the range where τ_{ec} lies between $1.25 \cdot \tau_{ec}^{min}$ and $2 \cdot \tau_{ec}^{min}$, with τ_{ec}^{min} being the minimum of the transit time. As expected, a reduction of the EB-junction width leads to an increase of the depletion capacitance C_{jEB} but also to a reduction of the forward transit time τ_f. Further insight into the distribution of the total delay can be gained by looking at the quasi-static transit time from a 1D simulation.
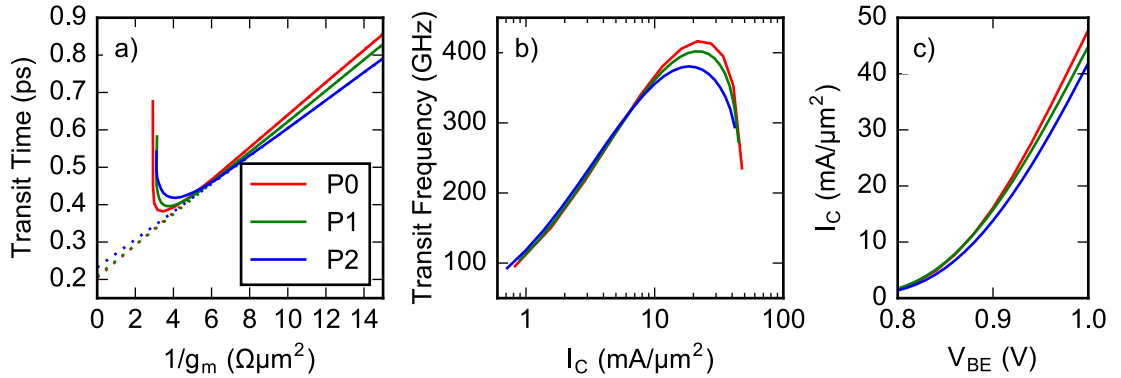
Figure 4.8.: a) Measured τ_{ec} as a function of g_m^{-1} for profile variations with increasing cap thickness from P0 to P2. b) f_T as a function of I_C. c) Collector current density as a function of base-emitter voltage. Measurements have been performed at fixed V_{CE} of 1.25 V.

The results of 1D simulations are shown in Figure 4.9. The different cap thicknesses are simply modeled by a shift of the emitter doping profile. The profile with the thinnest cap layer, labeled P0, corresponds to the standard profile described in Section 3.2.2. Profiles P1 and P2, have an emitter doping that is shifted away from the base by 1.5 nm and 3 nm, respectively. The vertical profiles are shown in Fig. 4.9a along with the simulated electron and hole density for a current density of 12 mA/μm^2 (which is close to peak

f_T). The transit frequency calculated in quasi-static approximation is shown in Fig. 4.9b and the collector current as a function of V_{BE} is plotted in Fig. 4.9c. The corresponding regional transit times are plotted in Figure 4.10. Here, it can be seen that a shift of the emitter doping away from the base leads to a strong increase of τ_e. In the zoomed inset in Fig. 4.10, one can also see the reduction of τ_{eb} due to the reduced depletion capacitance. Furthermore, an increase of τ_b can be observed for P1 and P2 which is caused by the an increase of the neutral base width.
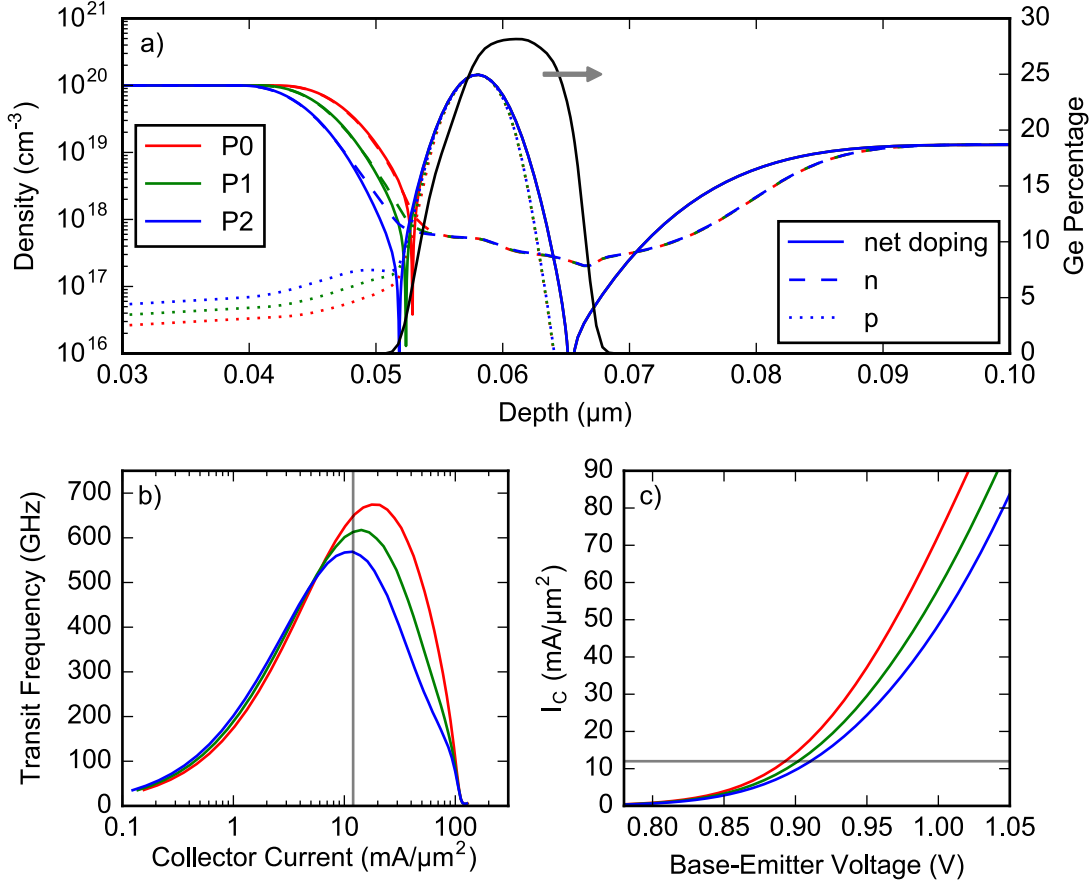
Figure 4.9.: 1D quasi-static simulations with different doping profiles at the EB junction. a) Doping profile as well as electron and hole density at $I_C = 12\,\text{mA}/\mu\text{m}^2$. b) f_T as a function of I_C. c) I_C as a function of V_{BE}. V_{CE} is fixed at 1.25 V.

Increased charge storage in the emitter is the main cause for the degradation of f_T with increasing BE junction width. Maximum f_T drops from 674 GHz for P0 to 569 GHz for P2 in the 1D simulation. The reason why a shift of the emitter doping away from the base leads to increased minority charge storage in the emitter is that a higher V_{BE} is needed to reach the same collector current density (see Fig. 4.9c). As a result, more holes are injected into the emitter at the same collector current density leading to an increased emitter transit time. Figure 4.11 shows the band structure for the different

doping profiles at a fixed base-emitter voltage of $V_{BE} = 0.91$ V. In case of profile P2 there is a marked barrier at the base-emitter junction. This barrier is a result of the larger bandgap in the EB space charge region due to a reduced bandgap narrowing as discussed in Section 3.4.2 (see also Fig. 3.22 for the shape of the bandgap in the EB depletion region). At low V_{BE} the shape of the bandgap in this particular region is masked by the high electric field. But at high V_{BE}, the EB space charge region collapses and reveals the shape of the band structure. This effect can be significantly reduced by a narrower EB-junction.
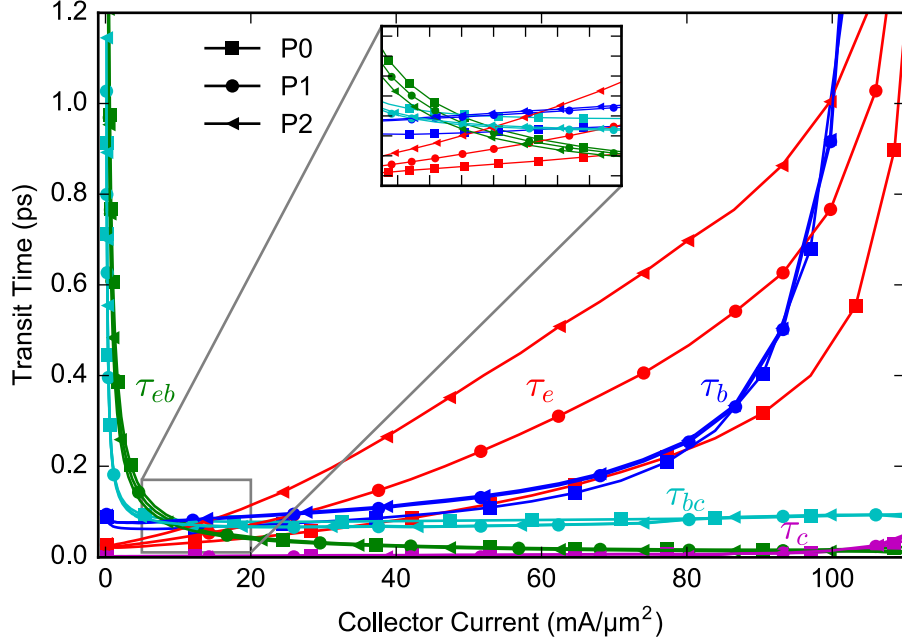
Figure 4.10.: Regional transit times as a function of the collector current.

The increase of τ_e also has a strong impact on the shape of the simulated f_T characteristics. Due to the domination of τ_e at high current densities, f_T drops significantly slower in case of a wider EB junction. However this effect is much more pronounced in the 1D simulation than in real transistor. As shown in the previous section, the Kirk effect sets in much earlier in a real device so that the increase of τ_e at high I_C is covered by a much higher τ_b. The measured f_T characteristics do also show a less abrupt decrease of f_T for profile P2, but the effect is much less pronounced.

Both simulation and experiment show that a reduction of the width of the base-emitter depletion region leads to a higher f_T because charge storage in the neutral emitter and in the EB depletion region is reduced. However, a smaller junction width also leads to increased tunneling at low and medium injection (cf. Fig. 3.3) and to a lowering of the base-emitter breakdown voltage BV_{EBO} which ultimately limits the maximum doping at the junction. Measured BV_{EBO} for P0, P1 and P2 are 1.0 V, 1.2 V and 1.6 V, respectively, where BV_{EBO} is defined by a reverse current of $I_E = 9\,\mu\text{A}/\mu\text{m}$. So even the device with the thinnest cap layer shows a BV_{EBO} which is still sufficiently high for many applications.
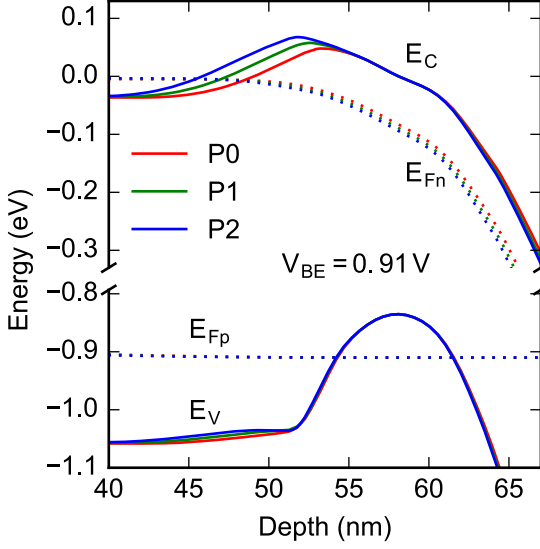
Figure 4.11: Band structure of the different profiles at $V_{BE} = 0.91\,\text{V}$. The y-axis is broken for sake of clarity.

4.3.2. Impact of the Selectively Implanted Collector

HBTs with two different collector doping profiles have been fabricated to investigate the impact of the selectively implanted collector (SIC) on the high-frequency performance. One of these variants has a highly doped SIC which is relatively close to the base. The other SIC-variant has a lower doping level and is further away from the base. The highly doped SIC receives twice the implantation dose of the lightly doped SIC, resulting in an increase of the arsenic concentration from about $8 \times 10^{18}\,\text{cm}^{-3}$ to $1.3 \times 10^{19}\,\text{cm}^{-3}$. Additionally, the thickness of the silicon buffer which is deposited prior to the SiGe base layer is decreased. This results in a narrower base-collector depletion region.
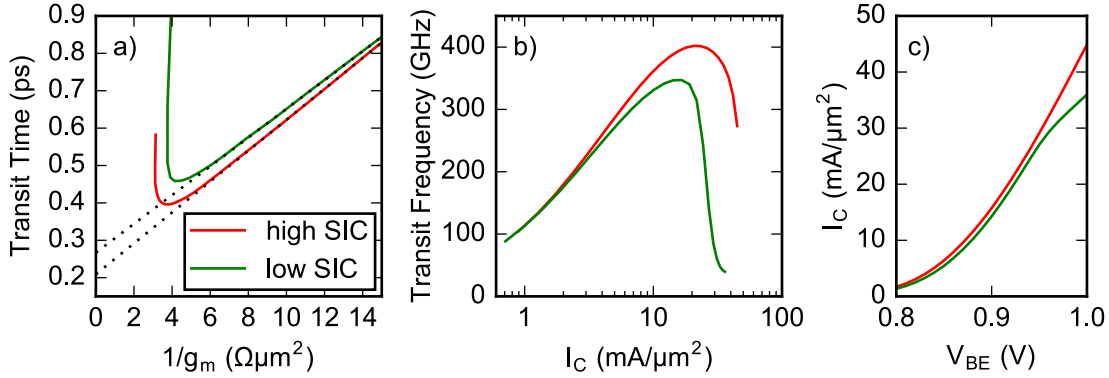
Figure 4.12.: a) Measured τ_{ec} as a function of g_m^{-1} for HBTs with different SIC doping. b) Corresponding $f_T(I_C)$ curves. c) Collector current density as a function of base-emitter voltage. Measurements have been performed at fixed V_{CE} of $1.25\,\text{V}$.

Measured τ_{ec}-curves of both transistor variants are shown in Fig 4.12. The device with the highly doped SIC has a maximum f_T of $402\,\text{GHz}$ (same sample as P1 in the previous section) while the one with the lightly doped SIC shows a maximum f_T of only $347\,\text{GHz}$.

The forward transit time τ_f extrapolated from the measured τ_{ec}-curve is $0.209\,\text{ps}$ for the transistor with the highly doped SIC and $0.267\,\text{ps}$ for the lightly doped SIC. The corresponding total capacitances from the slope of τ_{ec} are $41\,\text{fF}/\mu\text{m}^{-2}$ and $38\,\text{fF}/\mu\text{m}^{-2}$. Comparing the f_T characteristics in Fig 4.12b, one can see that f_T degradation sets in at a significantly lower current density in the case the lightly doped SIC. A higher SIC doping shifts the onset of the Kirk effect to a higher current density, which is shown in the following by means of simulation.
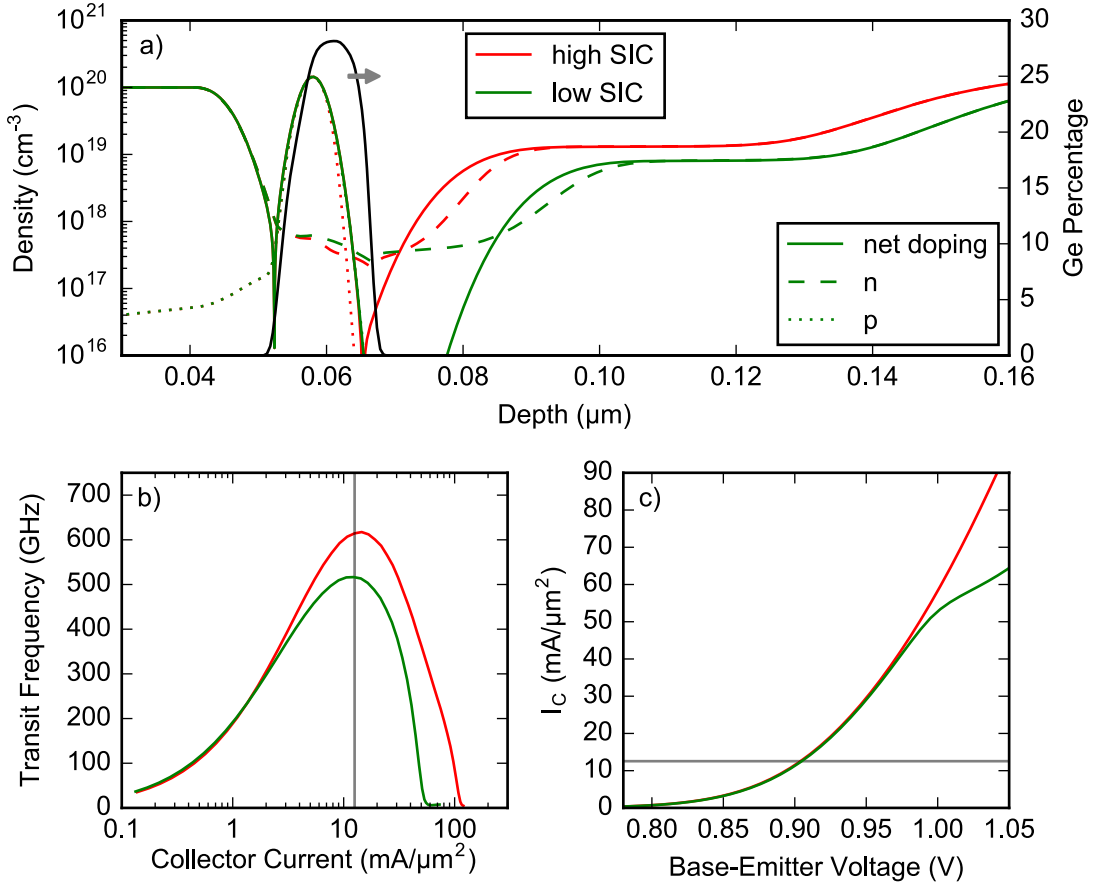
Figure 4.13.: a) Vertical doping profile and carrier densities at $I_C = 12\,\text{mA}/\mu\text{m}^2$. b) Simulated f_T as a function of I_C. c) Simulated I_C as a function of V_{BE}. V_{CE} is fixed at $1.25\,\text{V}$.

The vertical doping profiles used in the simulation are plotted in Fig. 4.13a. Calculated f_T values resulting from quasi-static 1D simulations are shown in Fig. 4.13b. As in the experiment, the lightly doped SIC variant has a significantly lower peak f_T than the HBT with the highly doped SIC. The corresponding regional transit times are plotted in Figure 4.14 as a function of the collector current density. This plot shows that the reduction of f_T is caused by an increase of τ_{bc} and τ_b. The higher τ_{bc}, which is constant at high injection, is related to the larger width of the base-collector space charge region.

The increase of τ_b is mainly caused by an enhancement of the Kirk effect. The Kirk effect refers to the push-out of the base into the collector at high injection, which leads to a rapid increase of the base transit time. It is caused by mobile electrons which are passing the base-collector region. If the charge density of these electrons exceeds the charge in the depletion region, majority carriers from the base spread into the depletion region. This results in a widening of the neutral base. A higher doping of the BC depletion region suppresses this effect. From the simulated base transit time, one can see that the Kirk effect sets in at a current density of about $35\,\mathrm{mA}/\mu\mathrm{m}^2$ in case of the low SIC doping, whereas for the high SIC doping, it sets in at about $80\,\mathrm{mA}/\mu\mathrm{m}^2$.
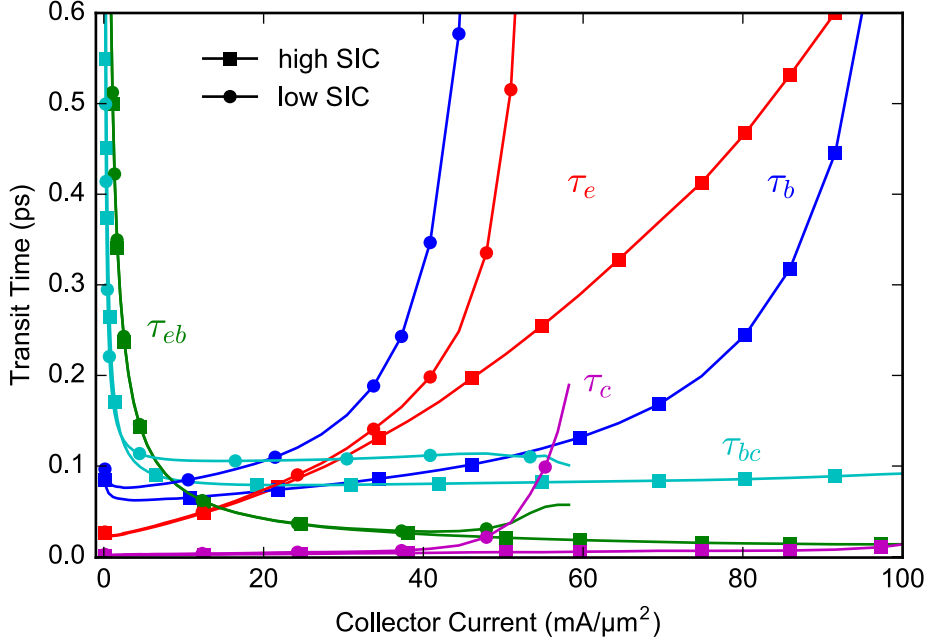
Figure 4.14.: Regional transit times for different variants of the selectively implanted collector.

Despite the significant delay of the Kirk effect, the current density at which f_T reaches its maximum only changed slightly in the simulation. The reason for this is that the impact of the emitter transit time is overrepresented in the 1D simulation. In case of the high SIC, τ_e leads to a premature degradation of f_T. As a consequence, 1D simulation might underestimate the potential performance improvements of an optimized collector profile.

4.3.3. Impact of the Position of the Heterojunction

The transistor characteristics are highly sensitive to the position of the heterojunction relative to the base-emitter depletion region. If the leading edge of the Ge profile extends considerably into the emitter, neutral charge storage in the BE space charge region will degrade f_T [104]. If, on the other hand, the leading edge of the Ge profile is located inside the neutral base, f_T is degraded by a reduction of the collector current. Hence, there

is an optimum range for the position of the heterojunction relative to the base-emitter depletion region [100].
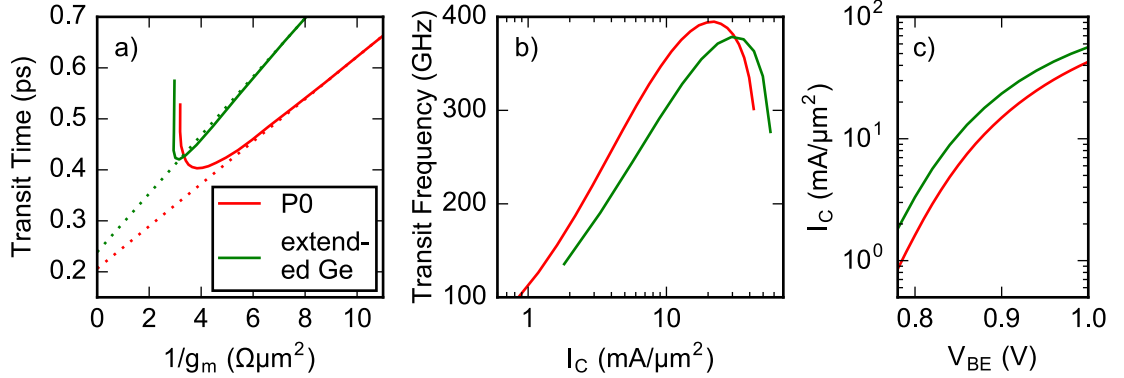
Figure 4.15.: a) Measured τ_{ec} as a function of g_m^{-1} for HBTs with different Ge profiles. b) Corresponding f_T vs. I_C curves. c) Collector current density as a function of base-emitter voltage. Measurements have been performed at fixed V_{CE} of 1.25 V.

Figure 4.15 shows measured f_T values as a function of the current density as well as the corresponding Gummel characteristics of two HBTs with different germanium profiles. One device has the standard profile described in Chapter 3, the other device has a Ge profile that extends about 3 nm further into the emitter. The doping is similar for both HBTs. The extension of the germanium into the emitter leads to a shift of the f_T curve to higher current densities and to a slight decrease of maximum f_T. For a better interpretation of this result, simulations of three different Ge-profiles have been performed. Besides the standard profile (labeled by P0) there are two variations where the leading edge of the germanium is shifted towards the emitter by 2 nm (P1) and 4 nm (P2), respectively (see Fig. 4.16a).

Simulated f_T characteristics and Gummel plots are shown in Fig. 4.16b and 4.16c, respectively. Highest f_T is achieved by P1, the profile which is shifted by 2 nm. Both the standard profile and the profile which extends further into the emitter show a significant degradation of f_T. The shift of the Ge profile also leads to a significant change in the collector current between P0 and P1 because the Ge percentage at the BE junction is increased. Between P1 and P2 there is also a small increase in I_C. At low currents, P0 and P1 show a similar transit frequency but the roll-off of f_T sets in much earlier in case of P0. For P2, the f_T curve is shifted to higher current densities. At high injection, where f_T drops heavily with increasing I_C, P2 shows the same behavior as P1.

To identify the different delay contributions that limit f_T for P0 and P2, the spatial distribution of the quasi-static electron and hole density response is shown in Fig. 4.17a and 4.17b at a current density of $5\,\text{mA}/\mu m^2$ and $50\,\text{mA}/\mu m^2$, respectively. At low current, P2 shows a much lower f_T than P0 and P1. This is a result of enhanced charge storage in the base leading to a higher τ_b as can be seen in the inset of Fig. 4.17a. The strong degradation of f_T at high current densities in case of P0 is due to enhanced charge storage in emitter. This is similar to the effect of a wider BE junction discussed above. The higher V_{BE} at a fixed current density leads to an enhanced injection of minority
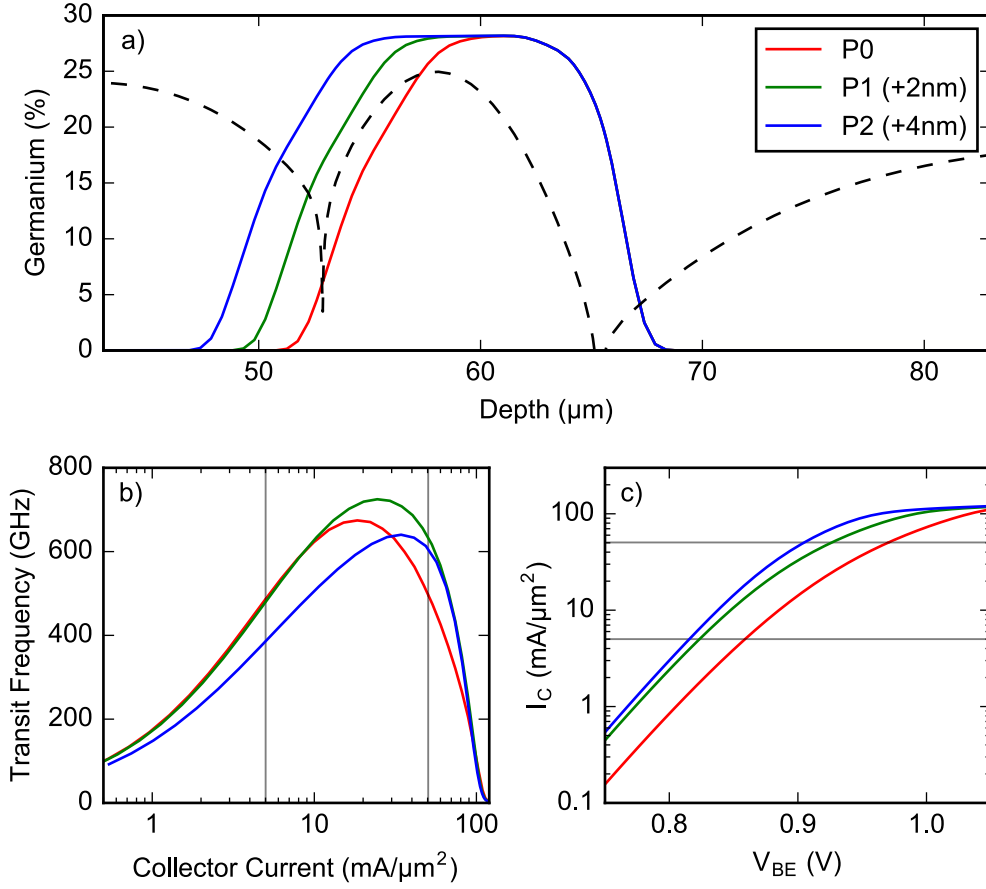
Figure 4.16.: a) Germanium profiles. b) 1D quasi-static simulation of $f_T(I_C)$ for the different Ge profiles. c) Corresponding simulated collector current as a function base-emitter voltage. $V_{CE} = 1.25\,\mathrm{V}$.

carriers into the emitter which results in a strong increase of τ_e. Profile P1 represents the optimal position of the heterojunction which minimizes both charge storage in the emitter and in the base.

The results of the 1D simulation suggest that the standard profile P0 is still considerably far away from the optimum and that one could gain performance improvements by an extension of the germanium profile into the emitter. However, the experimental results shown in Fig. 4.15, indicate that an extended Ge profile degrades maximum f_T. This discrepancy is again a result of the strong impact of τ_e in case of 1D simulation. A profile which minimizes charge storage in the emitter might perform better in a 1D simulation, but in reality, a profile which minimizes the base transit time might deliver a higher f_T. To demonstrate this effect, full 2D simulations including self heating have been performed for the Ge profiles P0, P1 and P2. The results are shown in Figure 4.18. The lateral geometry of device D2 from Chapter 3 has been used with an emitter width of $0.28\,\mu\mathrm{m}$. In contrast to the 1D simulation, there is practically no improvement of f_T
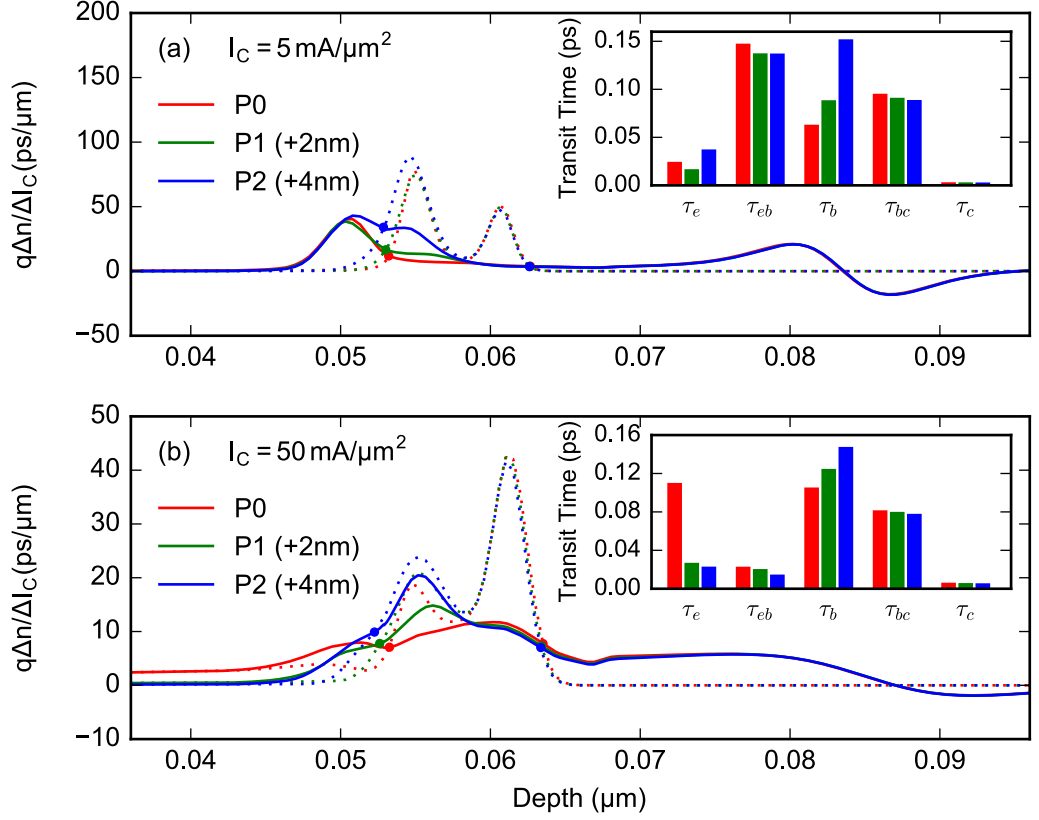
Figure 4.17.: Spatial distribution of the quasi-static electron (solid lines) and hole (dashed lines) density response at $I_C = 5\text{mA}/\mu\text{m}$ (a) and $I_C = 50\text{mA}/\mu\text{m}$ (b). The insets show the regional transit times.

from P0 to P1. They nearly show the same $f_T(I_C)$-characteristics. One the other hand, a strong reduction of f_T is shown for P2.

The results so far show that enhanced charge storage in the neutral emitter can be mitigated both by a reduction of the EB junction and by an appropriate positioning of the heterojunction. This leads to the question if the impact of the EB junction width, that was investigated in Section 4.3.1, is still considerable if the heterojunction is always at the optimum position. Corresponding simulations are shown in Fig. 4.19. Here, f_T is simulated for the three different EB junction widths and the heterojunction is shifted to the optimum position in each case. The optimum position is met by a shift of the heterojunction by $-1.5\,\text{nm}$, $-2\,\text{nm}$ and $-2.5\,\text{nm}$ for the standard emitter, the emitter shifted by $1.5\,\text{nm}$ and the emitter shifted by $3\,\text{nm}$, respectively. One can see that the impact of the EB junction width is much weaker if the the germanium profile is optimized accordingly. However, there is still a marked difference in maximum f_T, which is mainly caused by enhanced charge storage in the depletion region in case of a larger junction width.
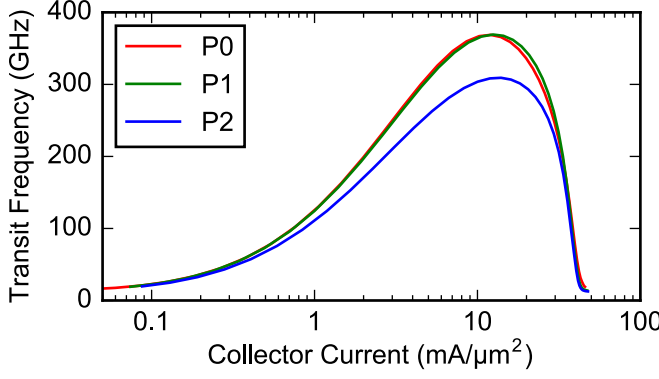
Figure 4.18: Full 2D simulation of f_T as a function of I_C for the different Ge profiles. Lateral geometry of D2 from Chapter 3 has been used ($W_E = 0.28\,\mu$m). $V_{CE} = 1.25\,$V.
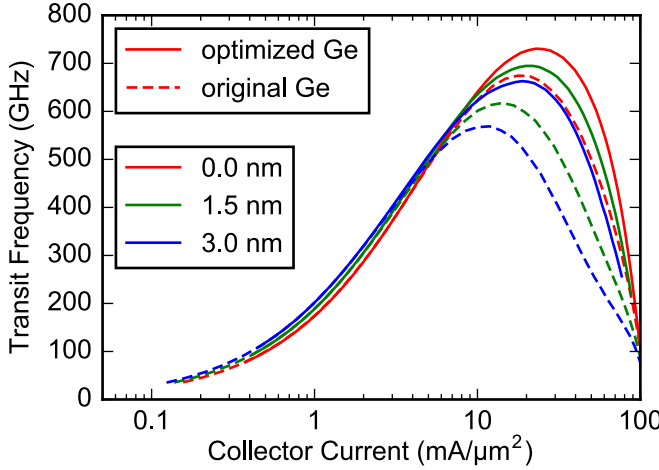
Figure 4.19: f_T from 1D quasi-static simulation for different EB junction widths (cf. Fig. 4.9c). Emitter is shifted by 0 nm, 1.5 nm and 3 nm away from the base. Dashed lines correspond to the results from Fig. 4.9, while solid lines are simulations with an optimized position of the heterojunction. $V_{CE} = 1.25\,$V.

4.3.4. Comparison with Profile N3 from Scaling Roadmap

In [73], a technology roadmap for SiGe HBTs based on TCAD simulation was suggested. This roadmap defines five technology nodes with an increasing targeted maximum oscillation frequency f_{max}. Starting with node N1 which represents the currently available technology with a f_{max} of 500 GHz to node N5 with a f_{max} of 2 THz which is considered as the physical performance limit [75]. Vertical doping profiles are derived for each node which yield the targeted performance under certain assumptions upon the lateral device structure. The vertical profile of node N3, which targets a f_{max} value of 1 THz, represents the theoretical profile that comes closest to the present experimental profile P0. In this section, the experimental profile P0 is compared to the hypothetical profile of node N3 by 1D quasi static simulation.

The vertical doping profiles P0 and N3 are plotted in Fig. 4.20a. The emitter profile of N3 has a lightly doped region at the BE junction to reduce the base-emitter tunneling current. This results in a relatively low base-emitter junction capacitance of 9.35 fFμm^{-2} compared to 11.7 fFμm^{-2} of P0. Further, the maximum doping concentration is higher in case of N3. An experimental realization of the lightly doped emitter region is difficult due the diffusion of dopants. The p-doping of the base is quite similar in both profiles. The base of N3 is slightly thinner and has a smaller peak concentration. The hypothetical
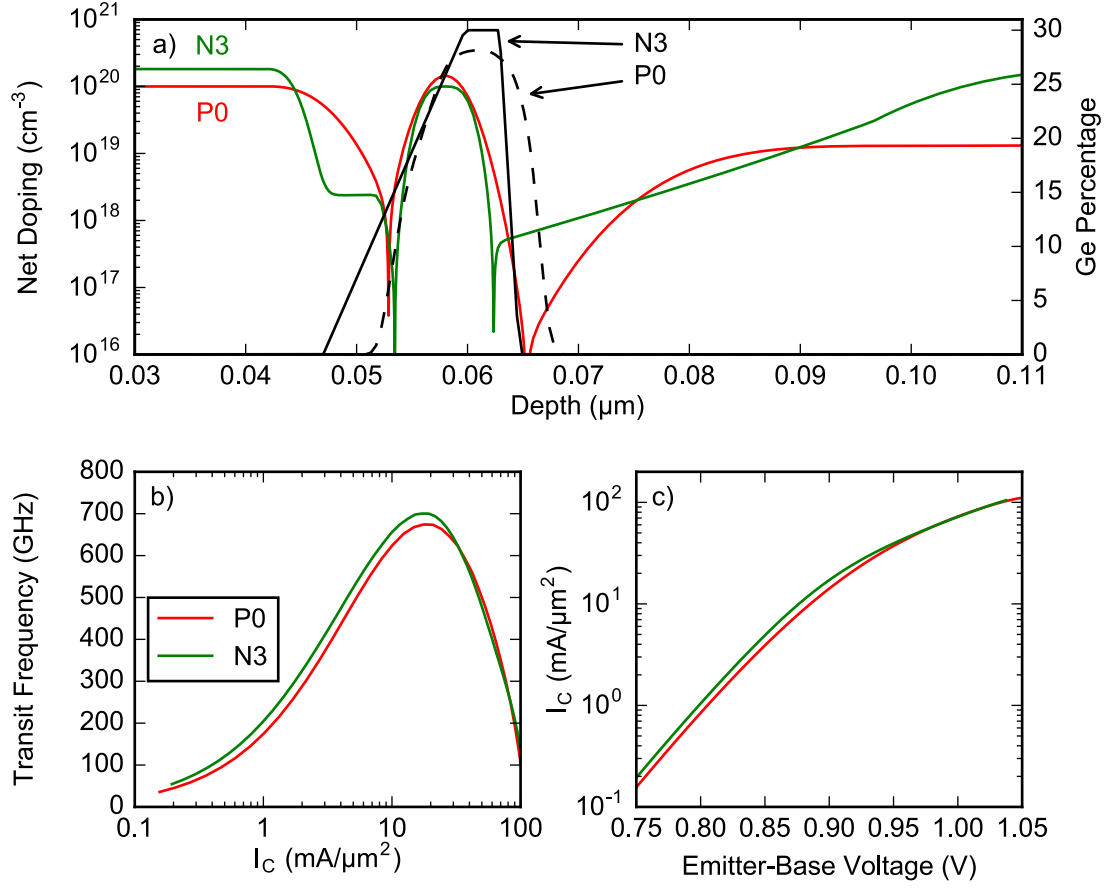
Figure 4.20.: Comparison of experimental profile P0 with hypothetical profile N3 from [73]. a) Vertical doping profiles. b) f_T as a function of collector current density from 1D quasi-static simulation. c) Gummel characteristics. $V_{CE} = 1.25\,\text{V}$.

Ge profile N3 has an extended gradient that reaches into the lightly doped emitter as well as a higher Ge peak concentration. The collector doping of N3 differs significantly from the experimental profile. It increases exponentially from about $5 \times 10^{17}\,\text{cm}^{-3}$ at the BC junction to a maximum concentration of $2 \times 10^{20}\,\text{cm}^{-3}$ within a range of 50 nm. The selectively implanted collector of P0 has an extension of about 70 nm and the peak concentration is reached 100 nm after the BC junction. It is hardly possible to reproduce the collector profile of N3 with the conventional process flow.

The simulated f_T characteristics of profiles P0 and N3 are shown in Fig. 4.20b and the corresponding Gummel plots are shown in Fig. 4.20c. N3 shows a higher f_T than the experimental profile. Its maximum value is 701 GHz[1] which is 4 % higher than

---

[1]In [73] a 1D peak f_T of 822 GHz was obtained for N3 by device simulations based on the BTE. However, in [105], the authors show that the BTE simulator predicts abnormal $f_T(I_C)$ characteristics for N3 due to an artificial barrier in the conduction band which is caused by the specific doping profile of N3.
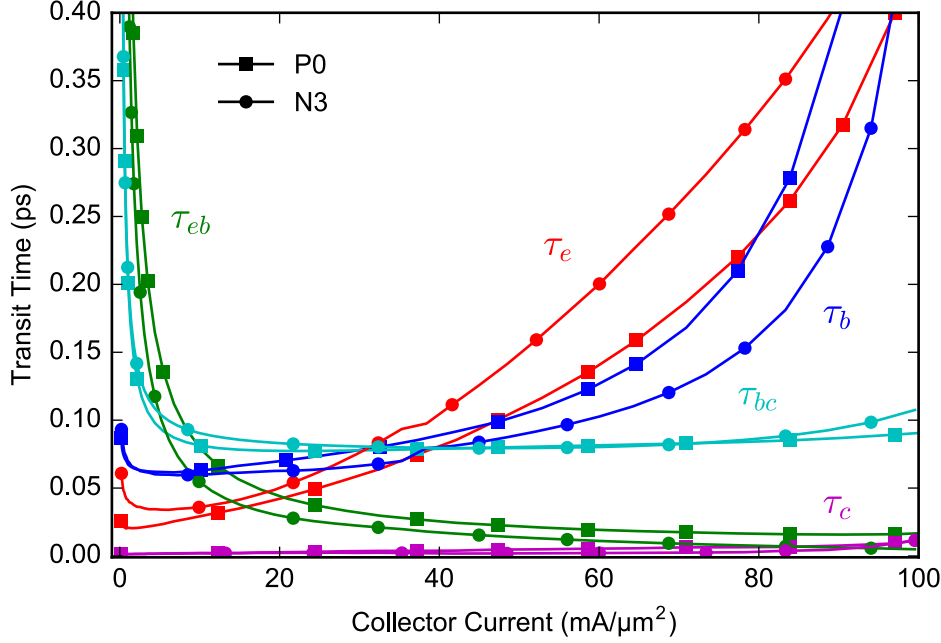
Figure 4.21.: Regional transit times as a function of the collector current for profiles P0 and N3.

maximum f_T of P0. Figure 4.21 shows the regional transit times for P0 and N3. The main contribution to the higher f_T comes from a reduced emitter-base transit time τ_{eb}. It is significantly smaller for N3 due to the smaller C_{jEB} and the extended Ge gradient. On the other hand, the base-collector transit time τ_{bc} is a bit higher at peak f_T. The emitter transit time τ_e is also higher for N3 due to increased minority charge storage in the low doped emitter. The base transit time of profile N3 is reduced compared to P0 due to the Ge gradient in the neutral base, which induces an additional drift field.

The results of the previous sections showed that small variations of the experimental profile, such as a shift of the heterojunction or the reduction of the EB depletion region, can lead to 1D f_T values above 700 GHz. It is therefore interesting to examine if the performance of the experimental profile can further be increased by an extended Ge gradient or a low doped emitter region as employed in the N3-profile.

To answer this question, four profile variations are compared in the following, which represent combinations of two different emitter profiles with two different kinds of the Ge profile. All combinations have the same base and collector doping. The first profile (shown in Fig. 4.22a) represents the experimental profile. The doping concentration corresponds to P0 from Fig. 4.20a. Here, as for all other profiles too, the leading edge of the Ge has been shifted to the position that yields maximum f_T. The second profile (Fig. 4.22b) is a combination of the experimental doping with a graded Ge profile as in N3. The third and fourth profiles combine the N3 emitter with the experimental germanium and N3 germanium, respectively (see Fig. 4.22c and Fig. 4.22d). In order to find the optimum position of the heterojunction for each profile, the leading edge of the germanium was varied in 0.5 nm steps. For the profiles with experimental emitter,
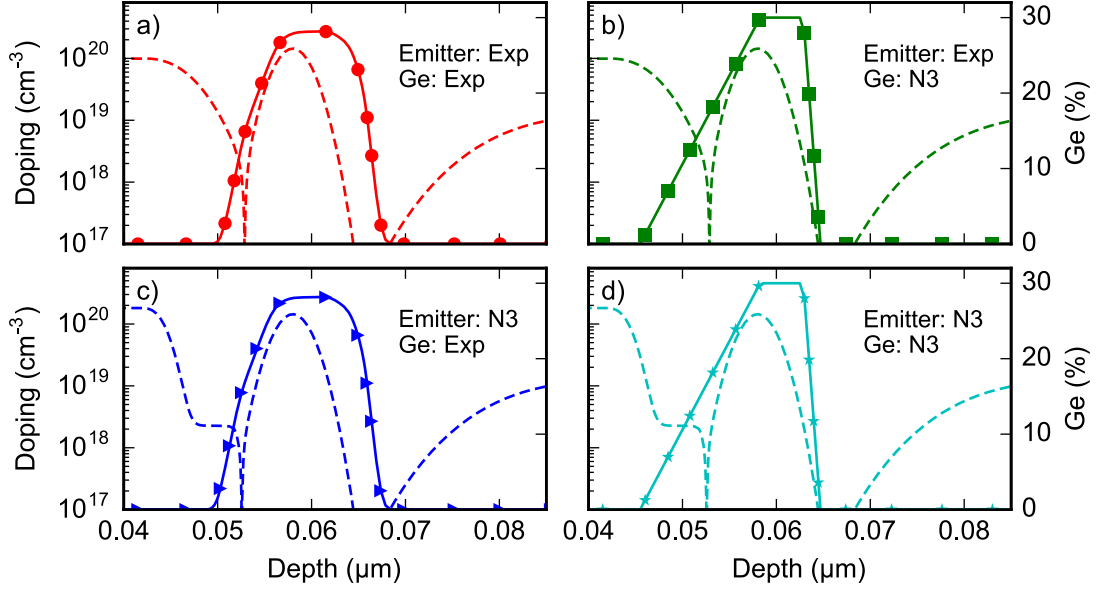
Figure 4.22.: Variations of the experimental profile using the emitter and Ge profile of N3.

maximum f_T was achieved with a shift of $-1.5\,\text{nm}$. In case of the N3-emitter, the experimental Ge edge was shifted by $-2\,\text{nm}$ and the Ge edge of N3 was shifted by $-1.5\,\text{nm}$.
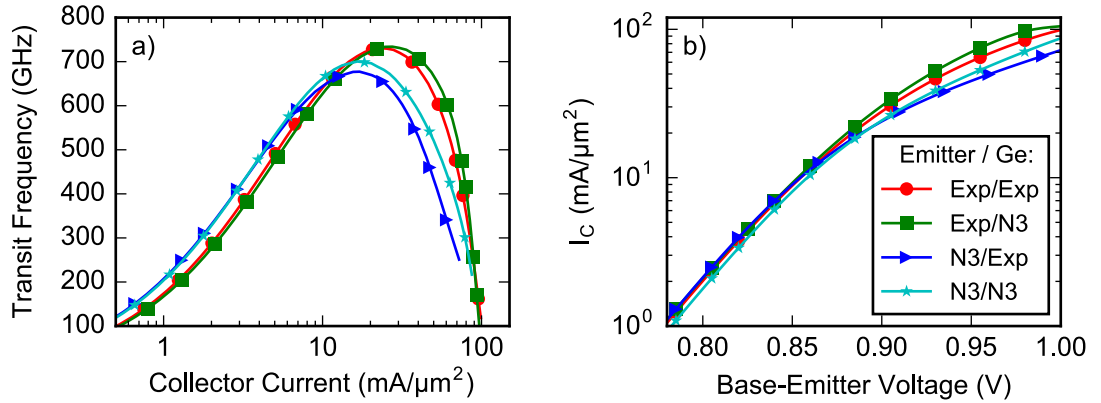

Figure 4.23.: f_T as a function of I_C from 1D quasi static simulation. f) Gummel plot.

Simulated f_T as a function of I_C is shown in Fig. 4.23a and the corresponding Gummel plots are shown in Fig. 4.23b. One can see that the different emitter profiles have a relatively large impact on f_T, whereas the impact of the Ge profile is smaller. The profiles with the N3 emitter have a higher f_T at low and medium current densities due to the smaller junction capacitance. However, the f_T roll-off sets in earlier due to enhanced charge storage in the low-doped emitter region at high injection. The extended Ge gradient of N3 leads to a small increase of the maximum f_T for both emitter variants. This effect is more pronounced in case of the N3-emitter.

Further insight into the impact of the Ge gradient can be gained from Fig. 4.24a and Fig. 4.24c where the regional transit times are compared for the different Ge profiles. In Fig. 4.24a, the two variants with the experimental emitter are compared. The extended Ge gradient of N3 leads to an increase of τ_e, particularly at low injection. On the other hand, it reduces τ_{eb} and τ_b at high current densities which results in a higher peak f_T. Basically the same behavior can be observed when the Ge variants with the N3-emitter are compared (see Fig. 4.24c). The only difference is that the impact on τ_{eb} and on τ_b is much more pronounced in this case, because the extended Ge gradient reduces charge storage in the low doped emitter region.
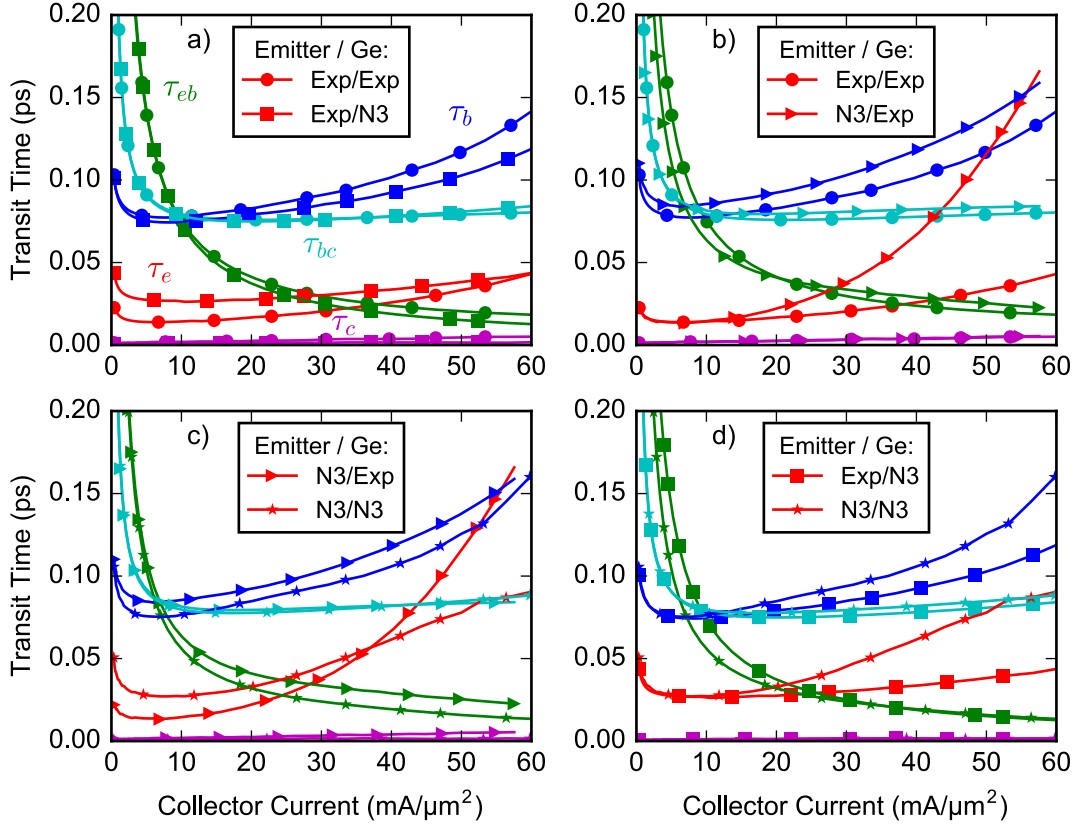
Figure 4.24.: Impact of the emitter and Ge profile on the regional transit times. Comparisons of the Ge variants are shown in a) and c). Comparisons of the emitter variants are shown in b) and d).

The impact of the different emitter variants on the regional transit times is demonstrated in Fig. 4.24b and Fig. 4.24d. The lower junction capacitance of the N3-emitter results in a reduction of τ_{eb} at low current. On the other hand, the storage of mobile charge in the low doped emitter region leads to a strong increase of τ_e and τ_b. Comparison of Fig. 4.24b and Fig. 4.24d shows again that the extended Ge gradient can partially compensate the negative effect of the low doped emitter. However this compensation is limited, even though the profiles shown here already represent the ideal position of the heterojunction.

4. Impact of Vertical Profile Variations on the Transit Frequency

The next question that will be examined is whether the experimental profile would benefit from the collector doping suggested by N3. Therefore, simulated f_T-curves with the original collector as well as with the collector of N3 are compared in Fig. 4.25a. These simulations have been performed both with standard Ge profile and with a shift of the leading Ge edge by $-2\,$nm. In the both cases, maximum f_T is increased by the N3 collector due to a mitigation of the Kirk effect. The impact of the collector is weaker for the standard Ge profile, because in this case, f_T is dominated by charge storage in the emitter. Hence, a mitigation of the Kirk effect has less influence on the peak value of f_T. In the real transistor, where f_T is determined by τ_b, one could expect a significant impact of such a collector profile.
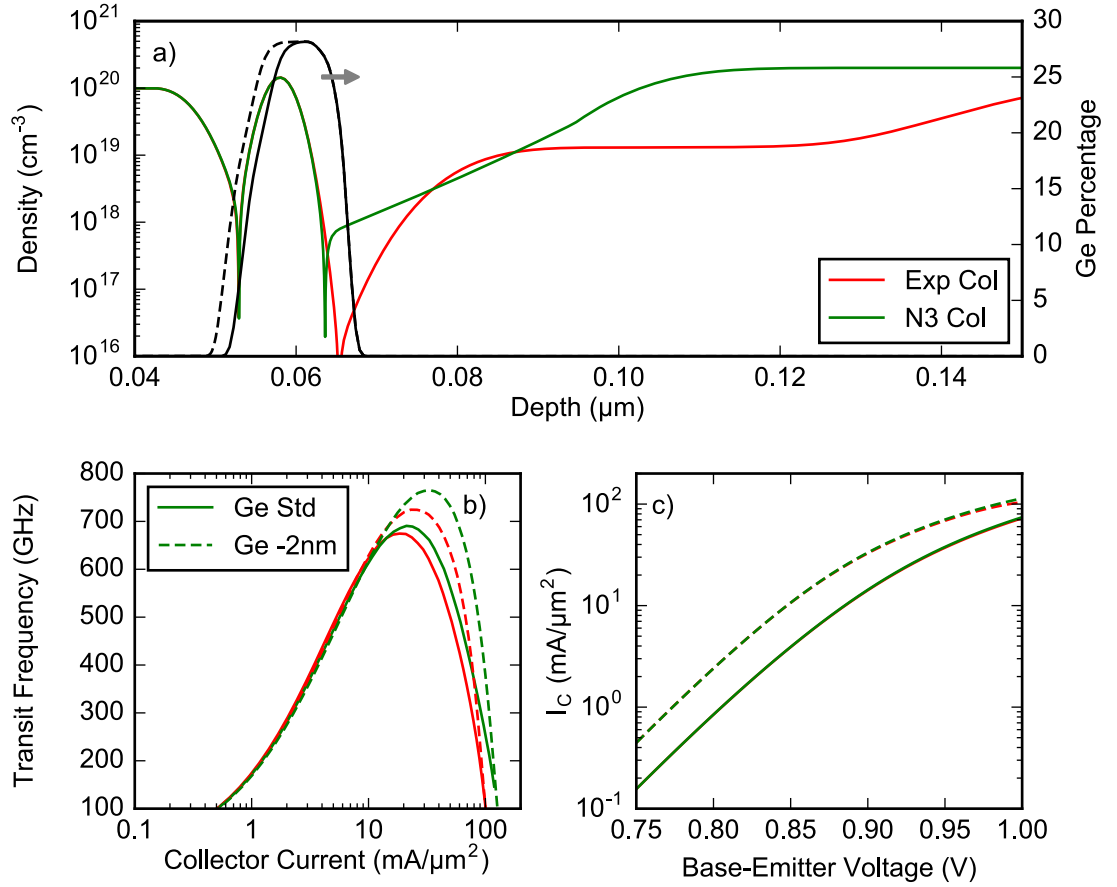
Figure 4.25.: Comparison of collector profiles. a) Vertical profiles. b) f_T from 1D quasi-static simulation. c) Collector current as a function of V_{BE}.

4.4. Sensitivity of the Simulated Transit Frequency to the Vertical Profile

The previous sections have shown that the transit frequency is very sensitive to certain changes of the vertical profile. Small variations of the doping at the base-emitter junction or the position of the heterojunction, for example, can change f_T significantly. This raises two questions: Can the discrepancy between measured and simulated f_T, that was found in Section 3.4, be caused by errors in the vertical profile? And to what extend can the transit frequency of the HBTs, which are investigated here, be increased by further optimizations of the vertical profile? In order to answer these questions, full 2D simulations with the lateral geometry of device D2 and different variations of the vertical doping profile have been performed.
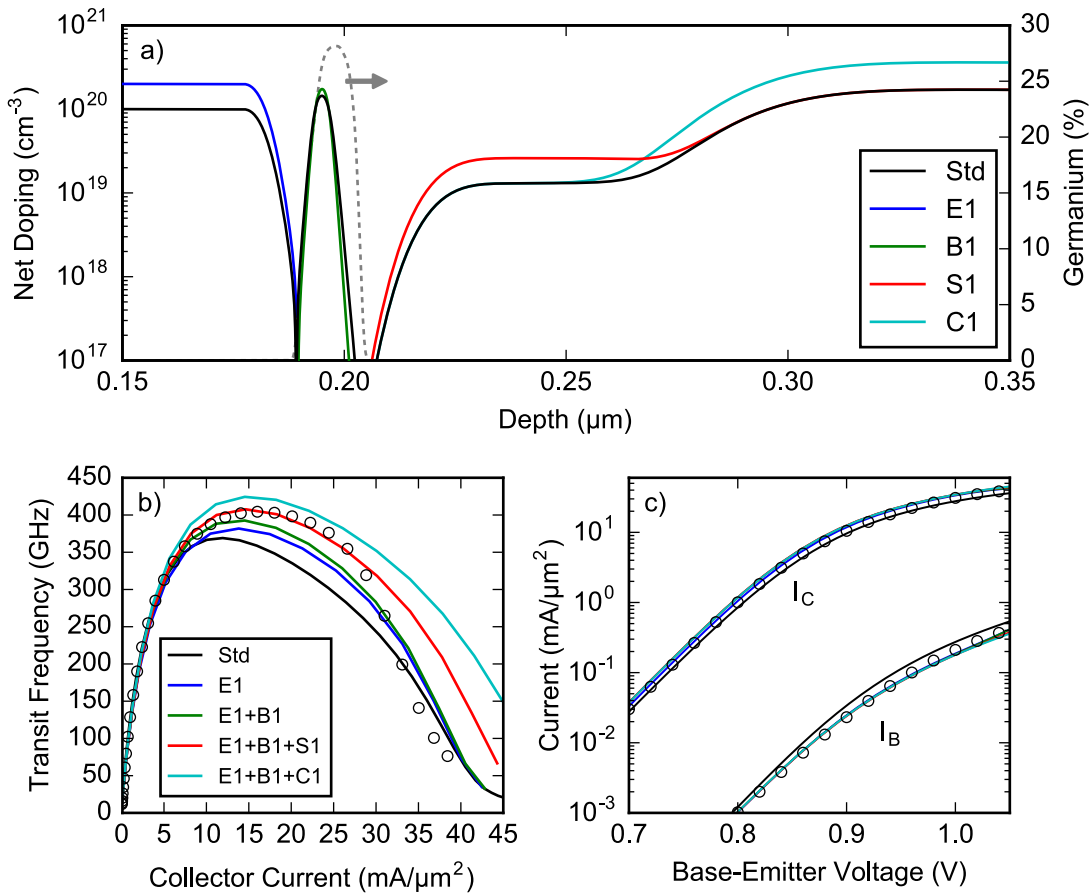
Figure 4.26.: a) Variations of the vertical profile. b) Measured f_T (symbols) and f_T from 2D simulation (lines) at $V_{CB} = 0\,\mathrm{V}$. c) Measured and simulated Gummel characteristics.

The first question is addressed by checking if a better agreement between measurement and simulation can be achieved by reasonable changes of the vertical profile. 2D simulations of several profile variations, which all lead to an increase of f_T, are shown

in Figure 4.26. The first profile variation concerns the emitter doping. The emitter profile (E1) is shifted towards the base by $1\,\text{nm}$ and the maximum doping concentration is increased from $1 \times 10^{20}\,\text{cm}^{-3}$ to $2 \times 10^{20}\,\text{cm}^{-3}$. This leads to an increase of f_T, which is mainly caused by a reduction of the emitter transit time. The higher doping concentration leads to a reduction of the minority carrier density in the emitter[2], which reduces τ_e additionally to the effect of a smaller BE depletion width (cf. Section 4.3.1) and the smaller R_E. The higher doping also leads to a small reduction of the base current. A further enhancement of f_T by shifting the Ge profile as in Section 4.3.3, could not be achieved for the profile with the optimized emitter doping. The agreement between measured and simulated f_T in the intermediate injection regime is improved by the higher emitter doping, but the premature f_T fall-off remains. Furthermore, the higher doping leads to an increase of the junction capacitance to $14.3\,\text{fF}/\mu\text{m}^2$ which disagrees with the measured value of $11.7\,\text{fF}/\mu\text{m}^2$ and leads to a reduction of f_T at low injection (not visible in Fig. 4.26b due to linear x-axis).

The next profile variation that is shown in Fig. 4.26 is a reduction of the base width (B1), which accounts for a possible error of the measured boron profile. The ion bombardment during SIMS leads to a relocation of the atoms at the sample surface. This effect, which is known as ion-induced mixing, can lead to a broadening of the depth profile. Thus, it might be possible that the width of the real boron profile is smaller than the width of the measured profile. To evaluate the possible impact of such a measurement error, the base width (at $N_A = 10^{18}\,\text{cm}^{-3}$) is reduced from $11\,\text{nm}$ to $9\,\text{nm}$ while keeping the total dose constant. The green curve in Fig. 4.26b corresponds to a vertical profile which includes both the higher emitter doping and the thinner base. One can see that the reduction of the base width only leads to a small increase of f_T. The impact of a lower boron dose on f_T has also been checked and turned out to be even smaller.

Variations of the collector profile have also been tested. Simulated f_T characteristics are shown for two variations. One includes a higher SIC doping (S1) and the other includes a higher doping of the collector well (C1). Both simulations also include the higher emitter doping and the thinner base. One can see that a higher doping both in the SIC as well as in the collector well leads to a significant increase of f_T. This is caused by a reduction of the collector resistance and a shift of the onset of the Kirk effect. However, the overall agreement between measured and simulated f_T cannot be improved by a higher collector doping. One can improve the agreement in the intermediate regime, but this leads to a strong overestimation of f_T in the high injection regime. Thus, the higher emitter doping is the only profile variation that improves the overall agreement between measurement and simulation. However, the higher emitter doping leads to an increase of the BE junction capacitance which contradicts with the measured values. In conclusion, the sensitivity analysis has shown that the deviation between measurement and simulation cannot be attributed to errors in the measured doping profile.

The profile variations of the emitter, the SIC and the collector well, (E1, S1, and C1),

---

[2]Within the depletion approximation, the minority hole density at the edge of the BE depletion region can be expressed as $p_n = \frac{n_i^2}{N_D} \exp\left(\frac{V_{BE}}{kT}\right)$, which illustrates the reduction of the hole injection into the emitter with increasing donor concentration N_D.

are suitable candidates for a further enhancement of the transit frequency. In contrast to the emitter and the collector profile of node N3 from the scaling roadmap, they can be realized with the conventional process flow. The reduction of the base width (B1), on the other hand, only led to a small increase of f_T, although τ_b strongly contributes to the total transit time. Therefore, a second base profile variation (B2) with a strongly reduced base width of 5 nm (but with the same total dose) has also been tested. To fabricate such a thin base profile, a process with a low thermal budget would be required to reduce dopant diffusion. The different base profiles are plotted in Figure 4.27.
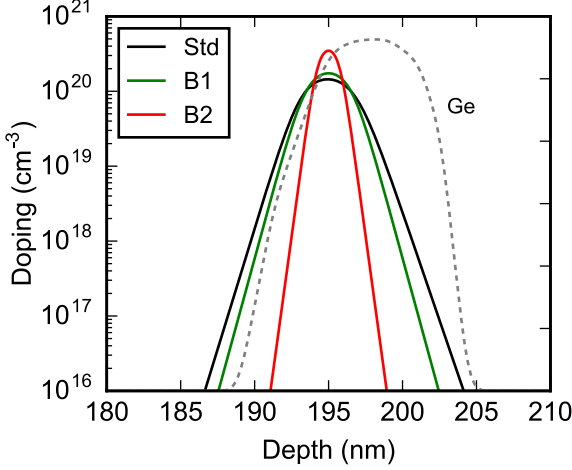
Figure 4.27.: Base profiles with a width of 11 nm (Std), 9 nm (B1) and 5 nm (B2) at a density of 10^{18} cm^{-3}. Total dose is the same.
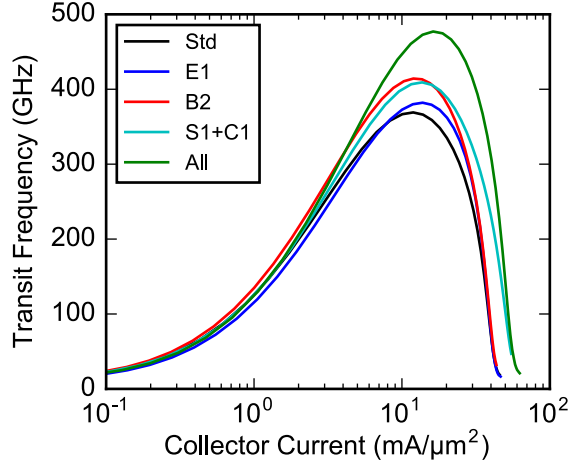
Figure 4.28.: 2D simulation of f_T as a function of I_C for different vertical doping profiles at $V_{CB} = 0$ V.

The impact of the individual profile variations on f_T is shown in Figure 4.28. The optimized emitter doping (E1) leads to an increase of maximum f_T by 4 % (as already shown in Fig. 4.26b). The reduction of the base width from 11 nm (Std) to 5 nm (B2) results in an increase of f_T by 12 %. In contrast to the higher emitter doping, which reduces f_T at low injection, the thinner base leads to an enhancement at both low and medium injection. A higher doping of the collector (by a combination of S1 and C1) leads to an increase of maximum f_T by 11 %. Combining all these variations results in an enhancement of f_T by 29 %.

These results show that there is still a considerable potential to increase the transit frequency of the HBTs by further optimizing the vertical doping profile. A reduction of the base width and an optimized collector profile are particularly important because f_T is determined by τ_b and τ_{bc}. A realization of the optimized emitter and collector profiles should be possible with the current process technology. However, the realization of a thin and highly doped base profile such as B2 would require a substantial reduction of the thermal budget.

4.5. Summary

In Chapter 4, the impact of the vertical doping profile on the transit frequency f_T has been investigated by experiment and simulation. Experimental variations of the base-emitter junction width, the Ge profile and the collector doping have been compared to 1D quasi-static simulations. The impact of the profile variations on f_T is qualitatively well captured by the simulation. Using regional transit-time analysis, it has been possible to assign the performance changes due to profile variations to individual transit times, which helps to identify the critical features of the vertical profile. Two important features are the width of the BE-depletion region and the position of the heterojunction. Experiment and simulation have shown that a narrow BE-depletion region helps to minimize charge storage in the emitter. Further it has been shown, that the position of the heterojunction relative to the BE-depletion region has to be chosen carefully to minimize both charge storage in the base and in the emitter.

It should also be pointed out that an idealized 1D simulation of the intrinsic device alone is not sufficient to optimize the vertical profile for a real transistor. The Kirk effect is shifted to significantly higher currents in the 1D simulation. This can lead to an overestimation of the impact of τ_e, because a reduction of charge storage in the emitter is much more pronounced when the Kirk effect is delayed. Thus, 1D simulation prefers profiles which minimize τ_e over profiles which lead to a reduction of τ_b. To obtain a realistic representation of the regional transit times, self-heating as well as the emitter and collector series resistance have to be included in the simulation.

Considering these effects, the regional analysis has shown that even for such advanced vertical profiles, f_T is still dominated by the transit times of the base and the BC depletion region. To increase f_T significantly in the future, it is therefore necessary to realize transistors with thin base layers and steep collector profiles as suggested in the ITRS roadmap for SiGe HBTs. Here, the simulation has shown that fairly moderate changes of the present profile can increase f_T by up to 30 %.

Another important result of this chapter is that it is not possible to improve the agreement between measured and simulated $f_T(I_C)$ significantly by reasonable changes of the vertical profile. This shows that the deviation, that has been found in the previous chapter, cannot be attributed to errors of experimental measurement of the vertical profile.

5. Impact of Variations of the Lateral Architecture on the RF Performance

In this chapter, the influence of variations of the lateral device architecture on the RF figures of merit is investigated. The starting point for the investigations is the device geometry shown in Fig. 5.1 (which is similar to the lateral geometry of IHP's SG13G2 technology [3]). The vertical doping profile from Chapter 3 is used here too. The aim of this chapter is to examine potential performance improvements from modifications of the lateral architecture by 2D simulation. Since no quantitative comparisons with measurements are made, an additional resistance or capacitance at the contacts of the simulation domain and an adjustment of the hole mobility in the external base region as in Chapter 3 is not required here.
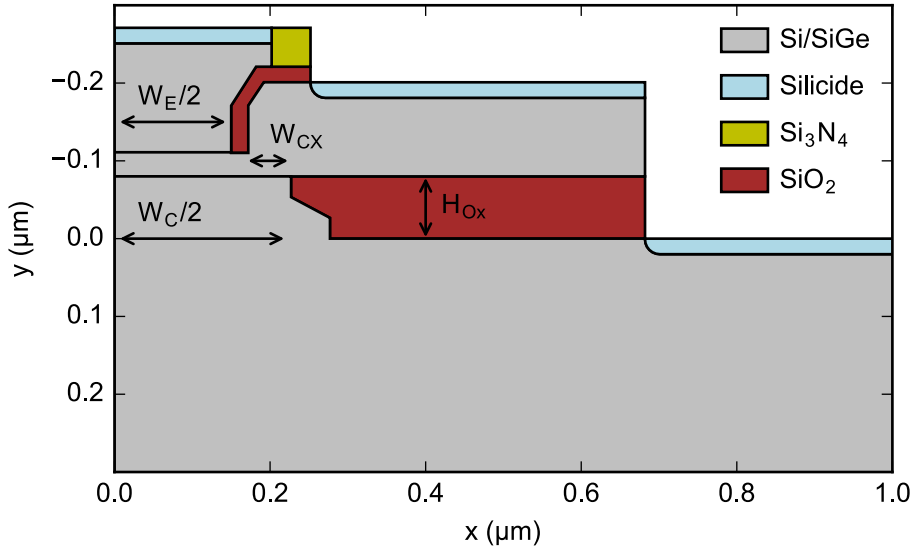
Figure 5.1.: Geometry of the 2D simulation domain which represents one half of the HBT. The complete domain extends to $x = 1.3\,\mu$m in lateral direction and to $y = 2\,\mu$m in vertical direction.

This chapter is organized as follows: In Section 5.1, the impact of the width of the emitter window W_E is discussed. In Section 5.2, the influence of boron diffusion from the external base into the inner transistor is investigated. Section 5.3 and Section 5.4 deal with the impact of the collector window width W_C and the oxide thickness H_{Ox}, respectively. In Section 5.5, the reference geometry shown in Fig. 5.1 is compared to the scaled geometry of the 700 GHz HBT that resulted from the DOTSEVEN project [21].

5.1. Impact of the Width of the Emitter

In order to increase the maximum oscillation frequency f_{max}, a reduction of the base resistance R_B is essential, as can be seen from the well-known relation between f_T and f_{max} which is given by [15]

$$f_{max} = \sqrt{\frac{f_T}{8\pi C_{BC} R_B}}. \tag{5.1}$$

For a given vertical transistor profile, the base resistance can be reduced by an optimization of the external base link or by a reduction of the emitter width W_E. However, a smaller W_E increases the relative contribution of the device perimeter to the total capacitances and hence degrades f_T. Thus, when designing the layout of an HBT there is an inevitable trade-off between maximum f_T or maximum f_{max}. In practical applications, a balanced device design is usually desired in which f_{max} assumes a value between $1 \times f_T$ and $2 \times f_T$.

Simulated values of maximum f_T and f_{max} as a function of W_E are plotted in Fig. 5.2a for W_E ranging from $0.03\,\mu$m to $0.3\,\mu$m. Simulations have been performed both with and without taking into account self-heating. For the simulations that take into account self-heating, a linear dependence of the thermal resistance R_{th} on the emitter width is assumed in accordance with the measured values of devices D1, D2 and D3 from Section 3.1. The following relation is used to model R_{th} as a function W_E:

$$R_{th} = 3310\frac{\mathrm{K}}{\mathrm{W}} - 1480\frac{\mathrm{K}}{\mathrm{W}\mu\mathrm{m}} \cdot W_E \tag{5.2}$$

This assumption is obviously not valid for very wide or extremely narrow devices. However, for the values of W_E which are considered here, this should be a reasonable assumption.

Without self-heating, the HBT shows a maximum f_T of $500\,$GHz for W_E larger than $0.22\,\mu$m. Scaling down W_E leads to a degradation of f_T due to an increasing impact of the resistance and capacitance of the external transistor regions. On the other hand, f_{max} increases with decreasing W_E until it reaches a maximum of almost $800\,$GHz at $W_E = 0.05\,\mu$m. At this point the reduction of R_B and C_{BC} due to a smaller W_E is compensated by the degradation of f_T. Taking into account self-heating leads to a reduction of both f_T and f_{max}. This effect is more pronounced for wider transistors, because the higher power dissipation of larger devices leads to a stronger self-heating, even if the thermal resistance is smaller. As a result, f_T reaches its maximum at smaller W_E than in the isothermal case. The peak value of $435\,$GHz is attained at a W_E of $0.2\,\mu$m.

Besides f_T and f_{max}, additional parameters are shown in Fig. 5.2 to characterize the simulated device structures. The parameters considered here are the same that are usually used to characterize the device experimentally. These are the zero-bias capacitances C_{EB} and C_{BC}, the base resistance R_B, the emitter resistance $R_{E,flb}$ and the collector resistance $R_{C,sat}$. Experimentally, the base resistance can be determined from measured s-parameters using the circle impedance method [106], which yields the sum of the base
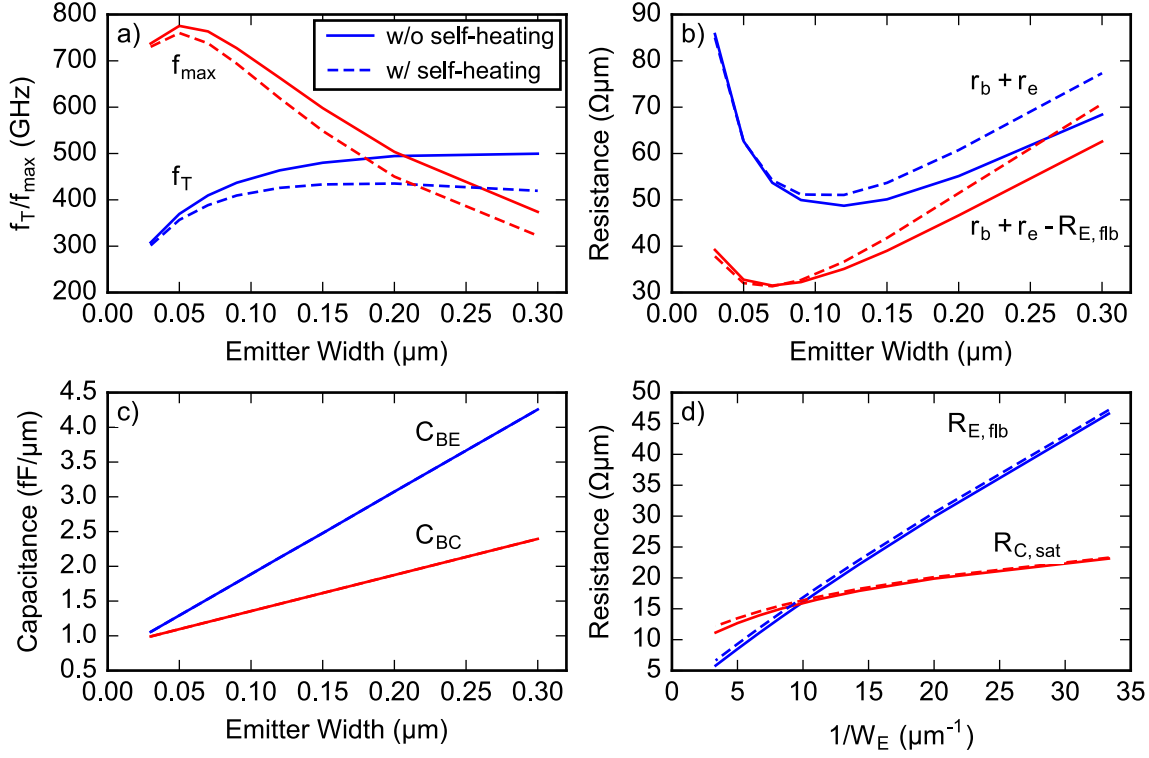
Figure 5.2.: Influence of the emitter width on the characteristics of the HBT. a) f_T and f_{max}. b) Sum of R_B and R_E extracted from s_{11} using the circle fit method. c) C_{EB} and C_{BC} at zero bias. d) R_E extracted by the fly-back method and $R_{C,sat}$ as a function of inverse W_E

and emitter resistance $r_b + r_e$. The symbols in lower case indicate the results of small signal AC measurements. To determine the base resistance by the circle impedance method, the emitter resistance must be known. A common method to determine the emitter resistance is the flyback method [107], which works as follows: A current is fed into the base while the floating voltage V_{CE} is measured at the open collector. The emitter resistance can then be calculated by $R_{E,flb} = \Delta V_{CE}/\Delta I_B$. $R_{E,flb}$ is also used to extract the base resistance from the results of the circle impedance method, even though $R_{E,flb}$ and r_e are not strictly equal. Here, the base resistance R_B is calculated by $R_B = r_b + r_e - R_{E,flb}$.

The collector resistance $R_{C,sat}$ is determined in saturation operation. At a predefined current gain β, typically between 1 and 2, the collector voltage V_{CE} is measured. $R_{C,sat}$ can then be calculated by $R_{C,sat} = \Delta V_{CE}/\Delta I_C - (1 + 1/\beta)R_E$ [107]. The emitter resistance R_E must also be known to calculate $R_{C,sat}$. Again, $R_{E,flb}$ is used for this purpose.

C_{EB} and C_{BC} are plotted in Fig. 5.2c as a function of the emitter width. They show an ideal scaling behavior resulting from a constant external capacitance and an internal junction capacitance that scales linearly with W_E. $R_{E,flb}$ and $R_{C,sat}$ are plotted in Fig. 5.2d as a function of the inverse emitter width. $R_{E,flb}$ scales nearly linear with

W_E^{-1}. Both, $R_{E,flb}$ and $R_{C,sat}$ are determined at relatively low current densities[1], hence the influence of self-heating is weak. Fig. 5.2b shows the base resistance R_B as well as $r_b + r_e$, which results from the extrapolation of s_{11} to infinite frequency. R_B shows a nearly linear decrease with W_E down to $0.07\,\mu$m. For smaller W_E, the base resistance R_B increases again. However, this increase is only an artefact of the extraction method of R_B. The emitter resistance is much larger than the base resistance at small W_E. For this reason, the difference between r_e and $R_{E,flb}$ can become larger than r_b for narrow emitters, which leads to a significant error in the extracted base resistance R_B. However, such narrow emitters are not considered in the subsequent sections.

In the following, several changes of the lateral device structure are investigated. In addition to f_T and f_{max}, the series resistances $R_{E,flb}$, R_B and $R_{C,sat}$ as well as the capacitances at zero bias C_{EB} and C_{BC} are calculated as described above in order to characterize the changes of the device structure. Self-heating is not taken into account in the following. If not stated otherwise, an emitter width of $0.12\,\mu$m is used.

5.2. Base Link Doping

The doping profile at the transition from the internal base to the external base region cannot be determined as simply and accurately as the vertical profile (which can be measured by SIMS with high accuracy). However, the doping profile in this region can have a significant impact on the characteristics of the HBT.

In the SG13G2 technology, the external base region is formed by in-situ B-doped Si epitaxy after the fabrication of the emitter and after wet etching a sacrificial nitride layer [108]. The complex fabrication of the external base region makes it difficult to make realistic assumptions on the doping profile. It is particularly unclear how far the p-doping extends into the inner transistor. For this reason, it is important to understand the impact of the doping at the external base link on the characteristics of the HBT, in order to avoid that wrong assumptions on the doping profile lead to improper simulation results.

In order to investigate the impact of the base link doping, four different profile variations are compared in this section. The extension of the boron doping from the external base is varied in both vertical and lateral direction. Fig. 5.3a shows the 2D profile which represents the maximum spreading of the external boron doping in both lateral and vertical direction. On the other hand, Fig. 5.3b shows the profile with a reduced spreading in both directions. One dimensional cross sections of the doping profiles along the lateral and vertical directions are shown in Fig. 5.3c and Fig. 5.3d. The profiles that represent the stronger out-diffusion are labeled A and the profiles corresponding to less out-diffusion are labeled B. The shape of the boron profile is not changed, it is simply shifted by 10 nm in lateral direction and 5 nm in vertical direction. In addition to the two profiles shown in Fig. 5.3a and Fig. 5.3b, two mixed variants with an extended spread of B doping in the lateral direction and a reduced spread in vertical direction and vice versa

---

[1]$R_{E,flb}$ is determined at $I_B = 2\text{mA}/\mu\text{m}^2$ and $R_{C,sat}$ is determined at $I_B = I_C = 2\text{mA}/\mu\text{m}^2$

are also considered. The simulated electrical parameters of the four different variations are summarized in Table 5.1.
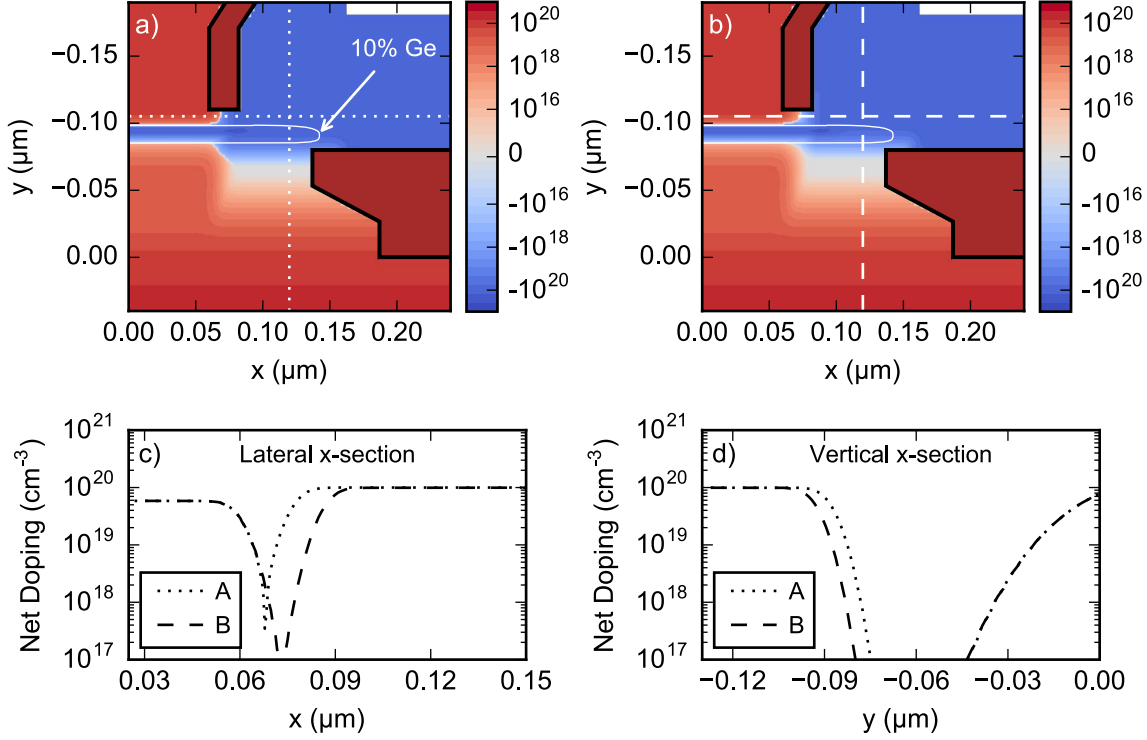

Figure 5.3.: Doping of the base link region. a) Base link doping with strong boron out-diffusion. b) Base link doping with reduced boron out-diffusion. The solid white line is the contour of 10 % germanium. c) Cross section of the doping profile in lateral direction at $y = -0.105$. d) Cross section in vertical direction at $x = 0.12$.

Simulated f_T and f_{max} of the four profile variations are plotted in Figure 5.4. A comparison of Profile 1 and Profile 2 shows a significant increase of both f_T and f_{max} due to the reduced spread of p doping in vertical direction. This is accompanied by a small reduction of C_{BC} (see Tab. 5.1). A similar behavior can be observed by comparing profiles 3 and 4, which, in contrast to 1 and 2, have a reduced spread of p doping in lateral direction. However, the difference in both f_T and f_{max} is less pronounced in this case.

The impact of a change of the p doping in lateral direction can be seen by comparing profiles 1 and 3 as well as 2 and 4. The reduction of boron doping under the EB spacer leads to a decrease of C_{EB} in both cases. A small reduction of C_{BC} can also be observed which is more pronounced in the former case where the p doping is closer to the collector. The base resistance increases significantly with less doping under the EB spacer. The transit frequency f_T also increases with reduced spreading of the p doping in both cases, however the difference is much larger between Profile 1 and Profile 3 than between Profile 2 and Profile 4. In contrast, f_{max} is decreasing in both cases, but the decrease is much stronger from Profile 2 to Profile 4.

5. Impact of Variations of the Lateral Architecture on the RF Performance

Table 5.1.: Simulated electrical parameters. The second and third columns contain the profile variant in lateral and vertical direction.

No.	x-Dop lateral	y-Dop vertical	f_T GHz	f_{max} GHz	C_{EB} fF/μm	C_{BC} fF/μm	R_B $\Omega\mu$m	R_C $\Omega\mu$m	R_E $\Omega\mu$m
1	A	A	464	663	2.125	1.459	33.2	15.0	13.6
2	A	B	507	730	2.125	1.428	32.5	14.9	13.7
3	B	A	485	656	2.115	1.452	36.7	15.0	13.6
4	B	B	509	680	2.115	1.426	37.4	14.9	13.7
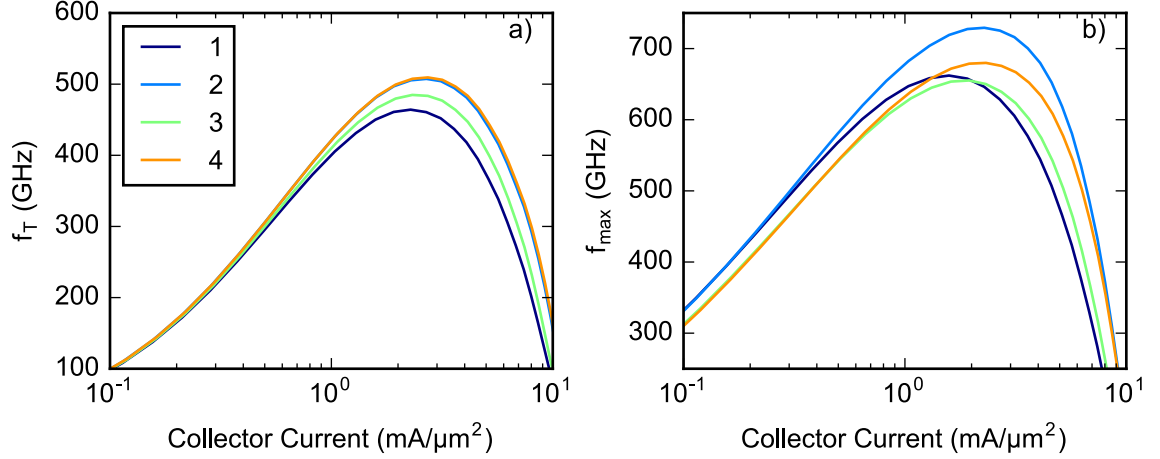
Figure 5.4.: Impact of boron out-diffusion from the external base on f_T (a) and f_{max} (b).

Overall, it can be stated that a relatively small change of the p doping at the conjunction of the internal and external base can lead to a significant change of f_T and f_{max}. In case of Profile 1 and Profile 2 for example, a 5 nm shift of the external base doping away from the collector leads to an increase of f_T and f_{max} by 9 % and 10 %, respectively. In order to understand the cause of this behavior one has to examine the charge storage in the device. Figure 5.5 shows the spatial distribution of the quasi-static change of the electron density Δn for Profile 1 and Profile 4 (cf. Sect. 4.2.1 for the definition of Δn). One can see that a significant amount of charge is stored in the area where the SiGe layer extends into the external base region. Laterally, the SiGe base reaches up to the edge of base-collector oxide. (In fact, the silicon or SiGe that grows on the oxide is polycrystalline. In the simulation, however, it is simply modeled as crystalline Si.) In case of Profile 4 the charge storage in this area is significantly smaller than for Profile 1. The cause of this difference lies in the position of the base doping edge relative to the SiGe layer. Cross sections of both profiles in vertical direction at $x = 0.12\,\mu$m are shown in Fig. 5.6, along with the conduction band energy E_C, the electron density n and the quasi-static change Δn. Here, one can see that the enhanced charge storage in the external base region is caused by an additional barrier at the p-side of the space charge region between the external base and the collector. This barrier is caused by the lower bandgap of SiGe compared to Si. The height of this barrier varies with the
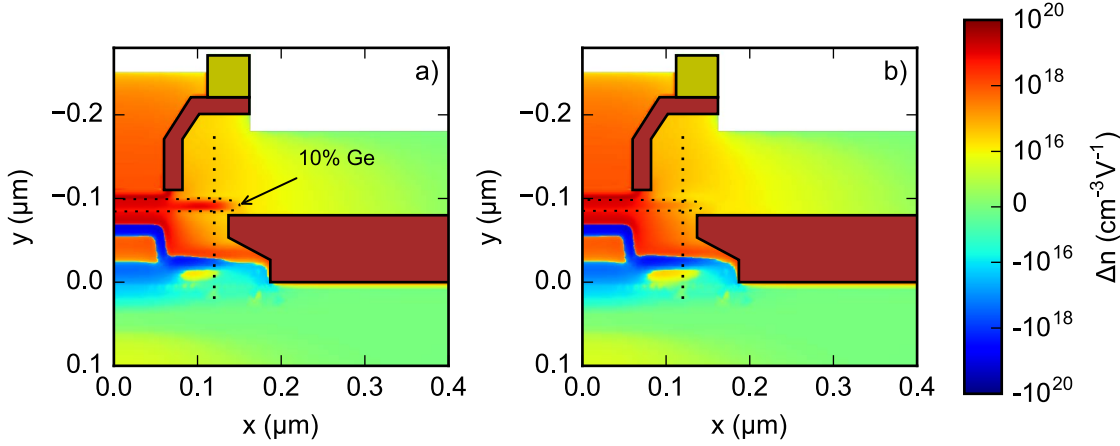
Figure 5.5.: Quasi-static change of the electron density Δn for the two different degrees of boron out-diffusion shown in Fig. 5.3 at $V_{BE} = 0.9\,\text{V}$. a) Profile 1. b) Profile 4.

position of the space charge region relative to the SiGe layer. Ideally, the trailing edge of the Ge profile should lie inside the space charge region so that the bandgap difference is masked by the electric field, as it is the case in the internal base region. However, as the p-doping extends beyond the SiGe base, an additional barrier in the conduction band is formed which leads to enhanced charge storage. This is basically the same effect as the well-known formation of a parasitic barrier due to boron out-diffusion from the internal base [109].

The occurrence of parasitic barriers in the internal base can be detected experimentally by an unexpectedly strong dependence of I_C on V_{CB}. This method is not applicable to parasitic barriers in the external base, because the bigger part of the electrons moving from emitter to collector do not pass this barrier. However, it can be assumed that an increased concentration of minority carriers in the base link region leads to enhanced recombination, and thus to a higher base current. For this reason, it should be possible to identify parasitic barriers in the base link region by increased neutral base recombination.

Usually, the base current is dominated by hole injection into the emitter. However, a significant contribution of neutral base recombination has been observed for some SiGe HBTs, which can lead to a degradation of the current gain and to a lower output resistance in forced-I_B operation [110, 111]. Several reasons for the observed neutral base recombination have been discussed in the literature. In [110], an extremely low electron lifetime was assumed in the vicinity of the CB heterojunction to model the observed base current, although no defects have been found by TEM investigations. In [112], it was shown that such an enhancement of the recombination can also be caused by parasitic barriers in the CB depletion region due to dopant outdiffusion from the SiGe base into the Si collector. In [113], the presence of traps in the CB depletion region has been assumed to reproduce the observed dependence of the base current on the collector-base bias.

A reduction of the base current with increasing V_{CB} is a clear indication of neutral base recombination. A higher reverse bias voltage at the base-collector junction leads
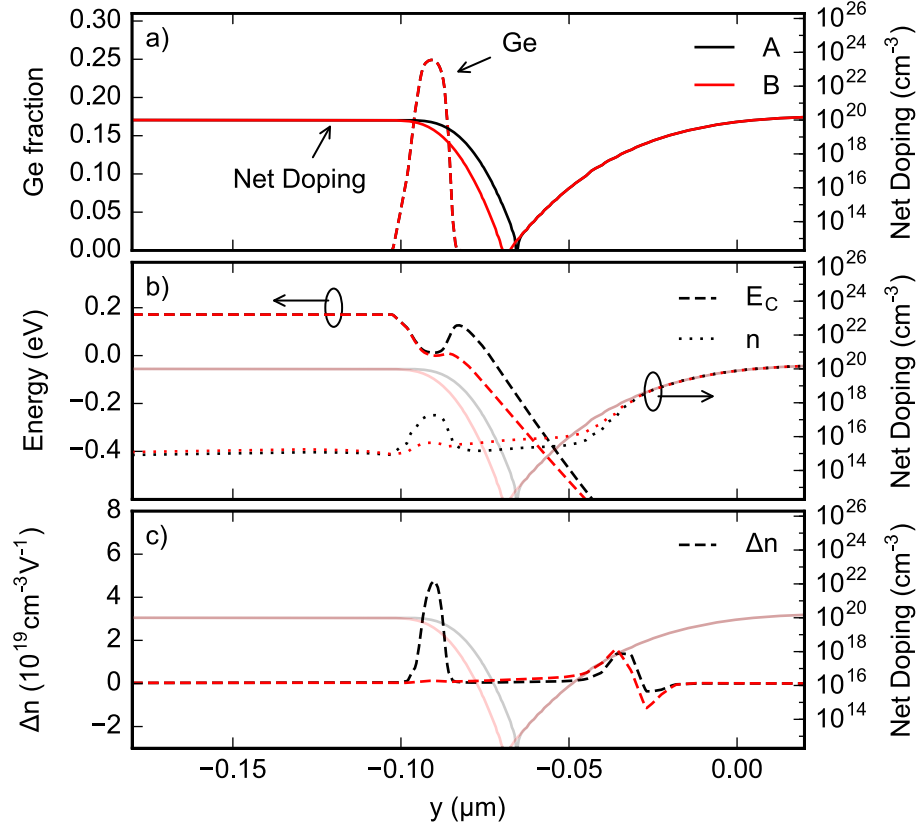
Figure 5.6.: 1D plot in vertical direction at $x = 0.12\,\mu$m (cf. dotted lines in Fig. 5.5) to illustrate charge storage in the external base. a) Doping and Ge profile. b) Conduction band energy E_C and electron density n. c) Quasi-static change of n.

to a spreading of the depletion region and to a reduction of the neutral base width. As the total recombination in the neutral base is proportional to its width, the base current decreases with increasing V_{CB}. Such an effect can also occur in the base link region, when excess minority charge due to a parasitic potential barrier leads to enhanced recombination. A higher V_{CB} can lead to a reduction of the barrier height, which results in less accumulation of minority carriers and thus in less neutral base recombination. Simulation results which demonstrate this effect are shown in Figure 5.7. Here, the base current (normalized to its value at zero V_{CB}) is plotted as a function of collector-base voltage for the different variants of the base link doping. These simulations have been performed using drift-diffusion transport because hydrodynamic simulations led to an unrealistic increase of I_B, even if recombination in the base was completely switched off. Both Auger and SRH recombination (with doping-dependent SRH lifetimes) has been included in the drift-diffusion simulations [37]. Impact ionization is not included in the simulation because the local models do not adequately capture the threshold of avalanche generation (cf. [113, 36]). The simulation shows a significant decrease of the normalized base current for Profile 1 which is caused by a modulation of the recombination in the base link region. For Profile 2 and Profile 4, the dependence on V_{CB} is considerably less

pronounced due to the absence of the parasitic potential barrier. For Profile 3, which also exhibits a parasitic barrier, the normalized base current lies between the two former cases because the barrier is further away from the internal region.
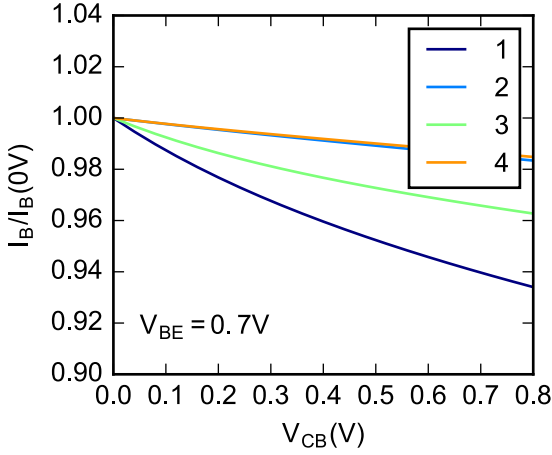
Figure 5.7.: Simulated base current as a function of V_{CB}. The base current is normalized to I_B at $V_{CB} = 0$ V. Impact Ionization is omitted in the simulation.
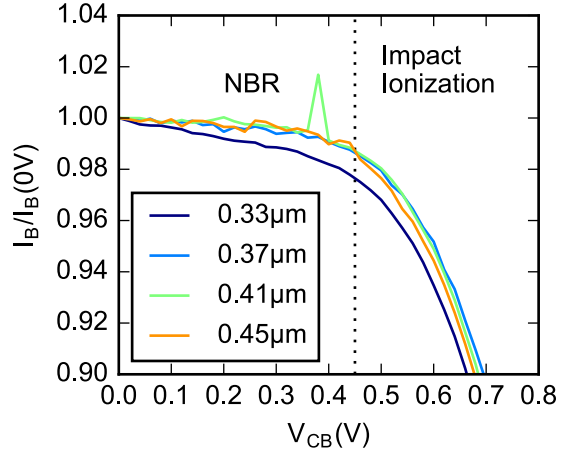
Figure 5.8.: Measured normalized base current as a function of V_{CB} for HBTs with different widths of the polysilicon emitter. $V_{BE} = 0.7$ V.

In order to find experimental evidence for a parasitic potential barrier at the base-collector perimeter, transistors with different emitter layouts have been examined. These devices have been fabricated in an experimental process during the DOTSEVEN project. They correspond to the wafer split D7a described in [21]. Four HBTs with different sizes of the emitter polysilicon window have been compared with regard to neutral base recombination. A smaller overlap between the polysilicon emitter and the external base has the effect that the p-doping of the external base can reach further towards the inner region of the transistor. So, if such barriers occur at all, it can be expected that they are most pronounced in devices with a small overlap between emitter and base.

Figure 5.8 shows the measured normalized base current for the four different HBTs, which have a drawn width of the polysilicon emitter of $0.33\,\mu\mathrm{m}$, $0.37\,\mu\mathrm{m}$, $0.41\,\mu\mathrm{m}$ and $0.45\,\mu\mathrm{m}$, respectively. The three devices with the largest emitter poly widths show nearly the same behavior. The slight decrease of the base current that can be observed up to a collector-base voltage of 0.45 V can be attributed to neutral base recombination (NBR). At higher V_{CB}, the decrease of the base current is determined by impact ionization. The device with the smallest base-emitter overlap shows a significantly stronger NBR. In accordance with the simulation results shown in Fig. 5.7, this can be explained by an enhanced accumulation of minority electrons in the base link region due to a parasitic potential barrier. However, it might also be possible that the stronger doping in the base link region introduces additional traps, which increase neutral base recombination. So, increased NBR can be regarded as an indication of parasitic potential barriers in the base link region but not as an evidence. Further investigations are needed to clarify this issue.

In summary, it can be stated that out-diffusion of boron from the external base in vertical direction should be prevented because it might lead to the formation of a parasitic potential barrier in the external base region and thus to a strong reduction of f_T and f_{max}. The diffusion of boron in the lateral direction under the EB-spacer also leads to a small increase of charge storage. On the other hand, it leads to a significant reduction of the base resistance and thus enhances f_{max}.

5.3. Impact of the Collector Window Width

In this section, the impact of the width of the collector window W_C is investigated. The parameter W_{CX} is used to change W_C. W_{CX} is the lateral distance from the base-emitter spacer to the base-collector oxide. With W_{SP} being the width of the spacer, W_C is given by $W_E + 2(W_{SP} + W_{CX})$, as illustrated in Figure 5.1. Simulations with W_{CX} of 0 nm, 25 nm and 55 nm have been performed, each with Profile 1 and Profile 4 in the external base. The electrical parameters are summarized in Table 5.2 and the corresponding f_T and f_{max} plots are shown in Figure 5.9.

In the case of doping profile 1, the reduction of W_{CX} leads to a strong enhancement of f_T and f_{max}. The main reason for this enhancement is a reduced charge storage in the external base. As the width of the SiGe layer is determined by the width of the collector window, the SiGe layer extends less into the external base when W_{CX} is reduced. This effect is illustrated in Fig. 5.10, which shows the quasi-static change of the electron density for a W_{CX} of 0 nm and 55 nm. In the case of Profile 4, the impact of W_{CX} on f_T and f_{max} is considerably less pronounced, because charge storage in the external base is strongly reduced compared to Profile 1, as discussed in the previous section. Assessing the electrical parameters in Tab. 5.2, it can be seen that a smaller W_{CX} leads to a considerable increase of R_C, despite the strong increase in f_T. The base resistance also increases slightly with decreasing W_{CX}, whereas C_{BC} is reduced. It should be noted that in a real transistor, the base resistance is heavily influenced by the crystalline structure and the conductivity of the material at the edge of the collector window. These properties strongly depend on the specific fabrication process and cannot be predicted by TCAD.

Table 5.2.: Simulated device parameters for different width of the collector window W_C and two variations of the p-doping profile at the external base link.

Doping	W_{CX} nm	f_T GHz	f_{max} GHz	C_{EB} fF/μm	C_{BC} fF/μm	R_B $\Omega\mu$m	R_C $\Omega\mu$m	R_E $\Omega\mu$m
1	0	509	786	2.124	1.271	32.5	19.6	13.5
	25	490	734	2.125	1.361	31.8	16.9	13.5
	55	464	663	2.125	1.459	33.2	15.0	13.6
4	0	513	706	2.114	1.259	39.5	19.8	13.6
	25	514	691	2.115	1.339	38.4	16.9	13.6
	55	509	680	2.115	1.426	37.4	14.9	13.7
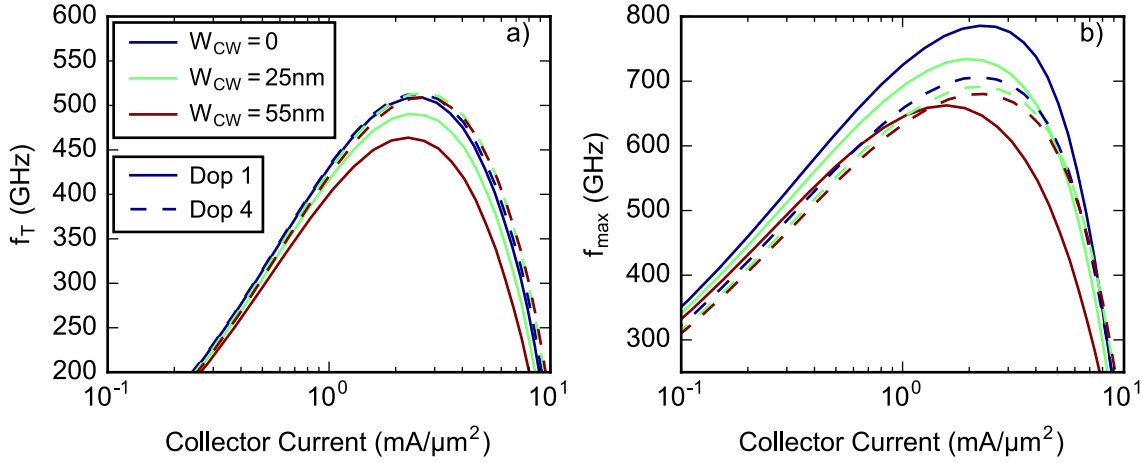
Figure 5.9.: Impact of the collector window width W_{CW} on f_T (a) and f_{max} (b) for two different variants of boron diffusion from the external base.

These results show that parasitic barriers in the base link region lead to a strong dependence of f_T and f_{max} on the width of the collector window (with respect to the emitter window). If a variation of the collector window geometry results in an unexpectedly strong change of the high-frequency performance of the HBT, one can assume that this device suffers from degradation due to boron out-diffusion from the external base. On the other hand, if f_T is not affected significantly, the presence of parasitic barriers is unlikely. In the data that has been available for this work, such a significant dependence on the collector window width has not been observed. However, no data was available from HBTs with the smallest overlap of the emitter polysilicon and the external base, in which such barriers most likely occur. A comparison of devices with different collector window geometries at smallest possible overlap the emitter polysilicon and the external base is still to be made.
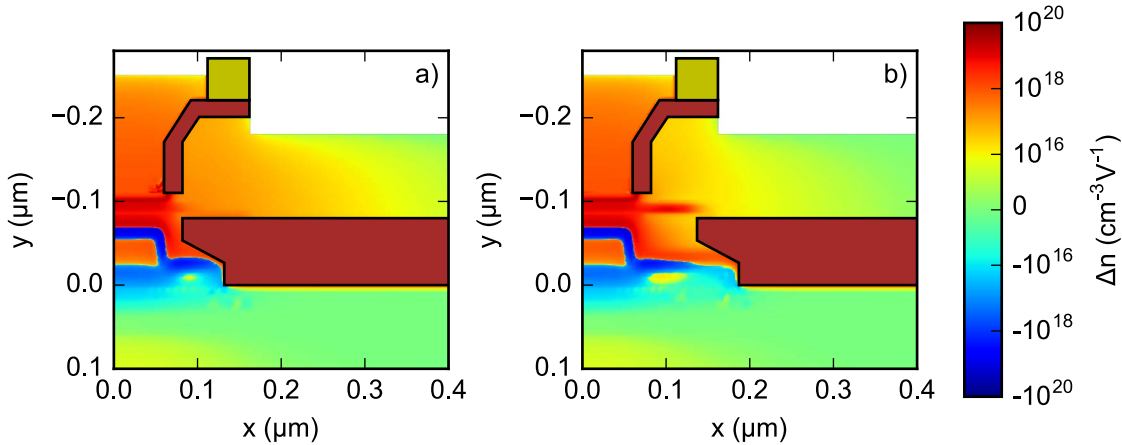
Figure 5.10.: Quasi-static change of the electron density Δn at $V_{BE} = 0.9\,\text{V}$ for $W_{CX} = 0\,\text{nm}$ (a) and $W_{CX} = 55\,\text{nm}$ (b), each with doping profile 4.

5.4. Impact of the Oxide Thickness

In Section 4.3.4 it was shown that an advanced vertical collector profile, as suggested in [73], can increase f_T by a reduction of the Kirk effect. However, such a steep collector profile cannot be realized with the standard device architecture of SG13G2 because the vertical collector profile is determined by the fabrication process flow, and in particular by the thickness of the base-collector oxide. The relevant part of the process flow starts with the formation of the collector well by ion implantation, followed by the deposition of the oxide. Then the collector window is opened and filled by selective Si epitaxy. After that, the SIC is formed by a masked ion implantation, followed by the SiGe base epitaxy. Hence, the vertical distance between the collector well and the base is defined by the oxide thickness. Reducing the thickness of the oxide thus allows to produce a steeper collector profile.

In this section the impact of a reduced oxide thickness on the RF performance of the HBT is investigated. A comparison of the doping profile of the standard device with an oxide thickness of 80 nm and the profile of a device with a reduced thickness of 40 nm is shown in Fig. 5.11. It is assumed that the same SIC implantation is used for both versions of the oxide. The new vertical profile that results from the shift of emitter, base and SIC towards the collector well is shown in Fig. 5.11c.
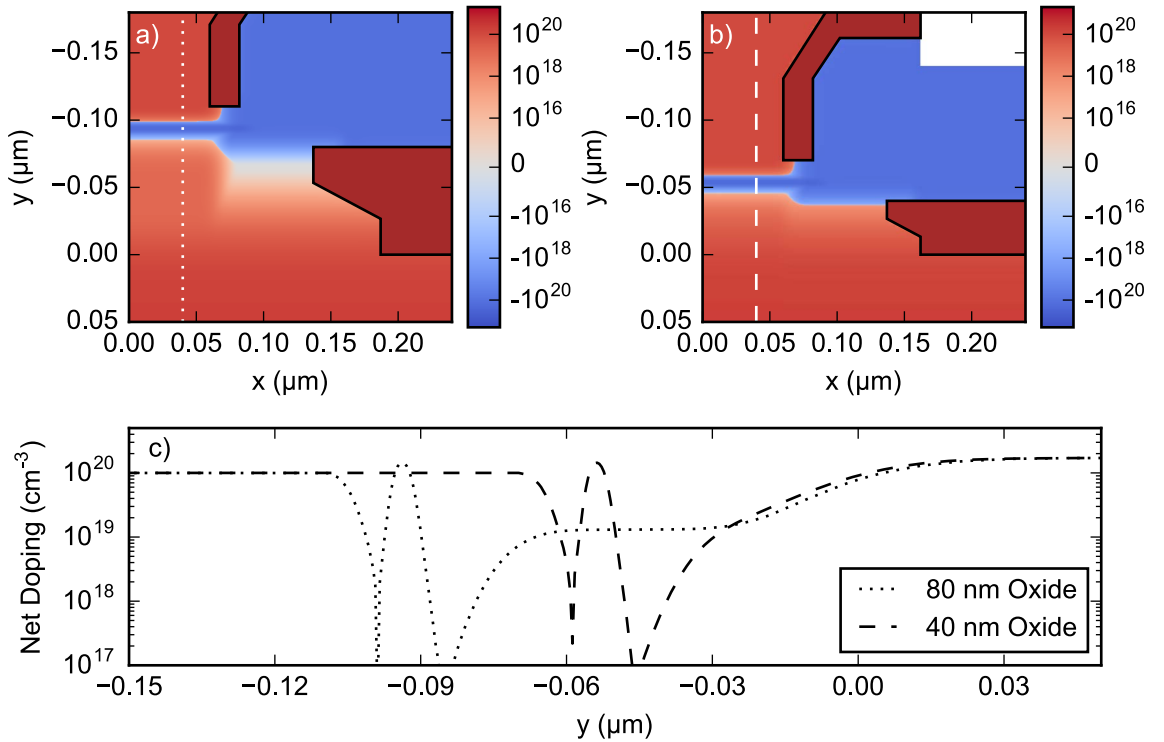
Figure 5.11.: 2D doping profile for an oxide thickness of 80 nm (a) and 40 nm (b). The doping profiles in vertical direction at $x = 0.04 \, \mu m$ are plotted in (c).

Simulated f_T and f_{max} curves of the devices with a H_{OX} of 40 nm and 80 nm are

shown in Fig. 5.12. Again, the simulations have been performed both with Version 1 and Version 4 of the external p-doping. The corresponding electrical parameters are given in Table 5.3. The reduction of the oxide thickness leads to a significant increase of f_T at high current densities. At low and medium currents, f_T becomes smaller with a thinner H_{OX}. R_C and also R_E are reduced by a reduction of H_{OX}. The reduction of R_E, however, is an artefact of the determination by the flyback method which assumes a constant reverse current gain [107]. The emitter itself is not affected by a thinner oxide, so the true R_E should not be affected either. The strong reduction of R_B is also not real. The circle impedance method assumes a negligible external BC capacitance C_{BCX}. C_{BCX} and R_C form a capacitive path between the base and collector terminals which reduces the input resistance at high frequencies. Thus the resulting R_B decreases with increasing C_{BCX} [107]. The base-collector capacitance C_{BC} increases significantly when the oxide becomes thinner and the distance between collector and base shrinks. The increased capacitance is the reason for the lower f_T at low and medium current densities. The reason for the strong enhancement of f_T at high I_C is that the onset of the Kirk effect is delayed due to the higher p-doping at the base-collector junction. In contrast to f_T, the peak value of f_{max} decreases with a thinner oxide. At low currents, f_{max} is significantly smaller due to the high C_{BC}. Its peak value is shifted to higher I_C due to the suppression of the Kirk effect. However, the increase in f_T is not sufficient to compensate for the increase of C_{BC}.
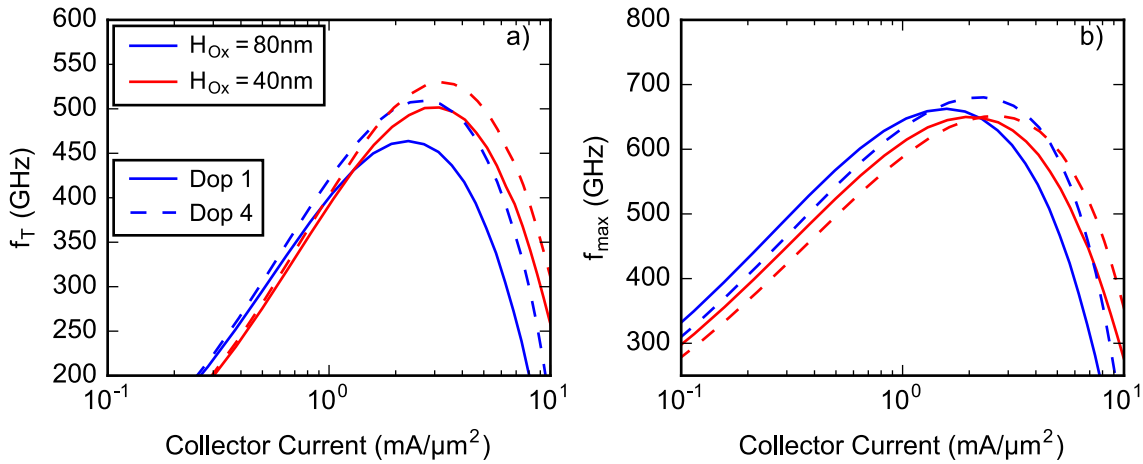
Figure 5.12.: Impact of the oxide thickness H_{OX} on f_T (a) and f_{max} (b) for two different variants of boron diffusion from the external base and $W_{CX} = 55\,\text{nm}$.

A comparison of the solid and dashed lines in Fig. 5.12 shows, that the doping of the external base has only a small effect on the impact of the oxide thickness. The increase of peak f_T is slightly higher in case of Profile 1, but mainly, the changes between 40 nm and 80 nm oxide are similar for Profile 1 and Profile 4.

Furthermore, simulations with a collector window width of $W_{CX} = 0$ are shown in Fig. 5.13. Here, the impact of the oxide thickness on f_T and f_{max} is basically the same as for $W_{CX} = 55\,\text{nm}$, even though one could expect a smaller impact due to the reduction of the external base-collector resistance. The fact that neither the external base doping,
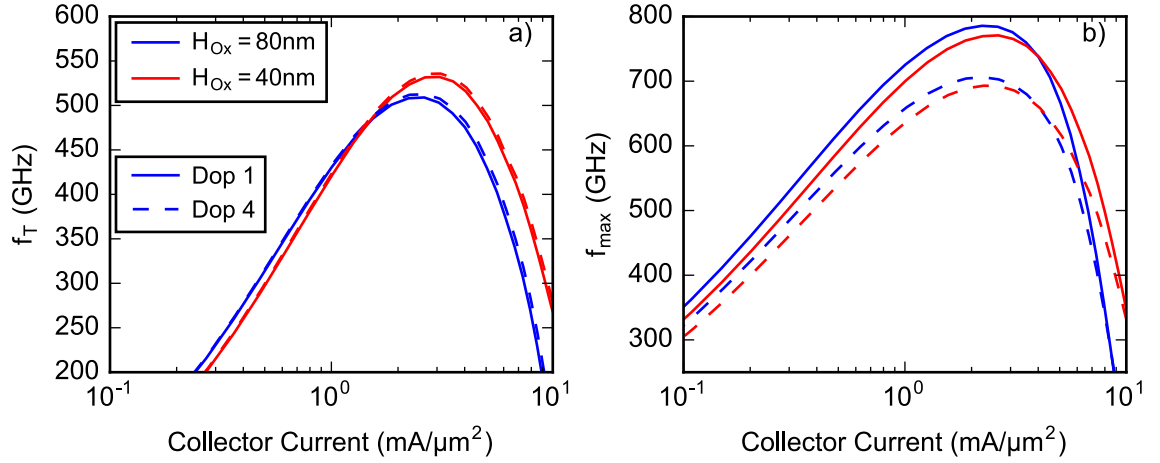
Figure 5.13.: Impact of the oxide thickness H_{Ox} on f_T (a) and f_{max} (b) for two different variants of boron diffusion from the external base with $W_{CX} = 0\,\text{nm}$.

nor the width of the collector window influence the impact of the oxide thickness on f_T and f_{max}, leads to the conclusion that these changes are mainly determined by the Kirk effect and the internal base-collector capacitance.

Table 5.3.: Simulated Device Parameters

$\mathbf{W_{CX}}$ nm	**Dop**	$\mathbf{H_{Ox}}$ nm	$\mathbf{f_T}$ GHz	$\mathbf{f_{max}}$ GHz	$\mathbf{C_{EB}}$ fF/μm	$\mathbf{C_{BC}}$ fF/μm	$\mathbf{R_B}$ $\Omega\mu$m	$\mathbf{R_C}$ $\Omega\mu$m	$\mathbf{R_E}$ $\Omega\mu$m
55	1	80	464	663	2.125	1.459	33.2	15.0	13.6
		40	501	650	2.122	2.226	29.6	12.5	12.9
	4	80	509	680	2.115	1.426	37.4	14.9	13.7
		40	530	652	2.111	2.173	34.3	12.7	13.0
0	1	80	509	786	2.124	1.271	32.5	19.6	13.5
		40	532	771	2.121	1.795	30.9	15.4	12.4
	4	80	513	706	2.114	1.259	39.5	19.8	13.6
		40	536	693	2.110	1.772	36.9	15.8	12.5

Scaling of the vertical doping profile usually results in a trade-off between performance and breakdown voltage. A higher doping of the collector can lead to a higher electric field across the BC junction which might decrease the breakdown voltage. In order to asses if a degradation of the BC breakdown voltage has to be expected for the thinner oxide, simulations of output characteristics in common-base and common-emitter configuration have been performed. For this purpose, the following recombination models have been activated in Sentaurus Device: Auger recombination, Shockley-Read-Hall recombination with doping-dependent lifetimes and avalanche generation with the impact-ionization model from [114] (referred to as New University of Bologna Model).

Simulated output characteristics in common-base configuration for fixed I_E are shown

in Figure 5.14. The simulations predict that for V_{CB} above $4\,\text{V}$, avalanche generation becomes significant and leads to a strong increase of I_C. This result is in accordance with the measured BV_{CBO} of $4\,\text{V}$ for this doping profile (see Tab. 3.1). The current due to impact ionization is also plotted in Fig. 5.14 (dashed lines). Comparing the simulated output characteristics of the different oxide thicknesses shows that there is practically no impact on the breakdown voltage in common-base configuration.
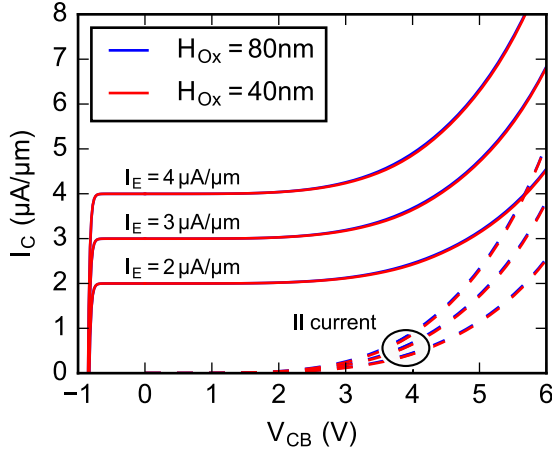
Figure 5.14.: Output characteristics in common-base configuration with fixed emitter current. Dashed lines show the current due to impact-ionization (II).
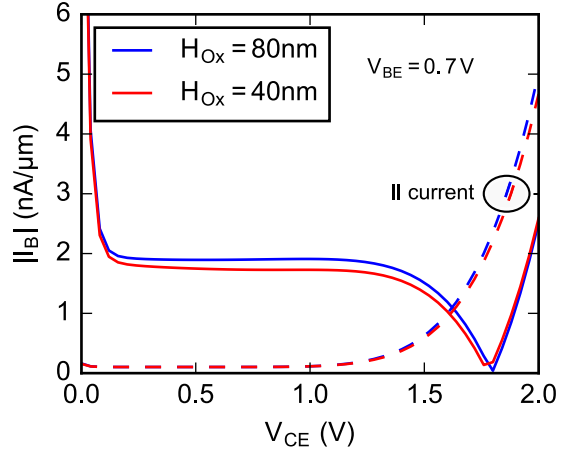
Figure 5.15.: Absolute value of the base current as a function of V_{CE} for fixed V_{BE} of $0.7\,\text{V}$. BV_{CEO} is defined by the change in sign of I_B.

The breakdown voltage in common-emitter configuration BV_{CEO} can be determined from the base current reversal with increasing V_{CE} [115]. When avalanche generation sets in, holes which are created in the BC depletion region cancel out the forward base current, which leads to a reversal of the terminal base current. Figure 5.15 shows the absolute value of the simulated base current at $V_{BE} = 0.7\,\text{V}$ as a function of V_{CE}. The base current becomes zero at about $1.8\,\text{V}$ for both variations of the oxide thickness. This suggests, that the proposed variation of the collector profile does not lead to a degradation of the breakdown voltage.

5.5. Analysis of Lateral Scaling of the DOTSEVEN HBT

As a result of the DOTSEVEN (D7) project, a new generation of SiGe HBTs has been developed, which shows a maximum f_T/f_{max} of $505\,\text{GHz}/720\,\text{GHz}$. The main technological measures which have led to such a strong enhancement of the high-frequency performance compared to the previous transistor generation are an optimized vertical profile, a decrease of base and emitter resistance due to the application of millisecond annealing in combination with a low temperature back-end as well as lateral device scaling. A summary of the technological progress can be found in [21].

5. Impact of Variations of the Lateral Architecture on the RF Performance

Scaling of the lateral device dimensions has significantly contributed to achieve f_{max} of more than $0.7\,\text{THz}$. Its impact is demonstrated in [21] by a comparison of wafer splits with the reference device architecture of SG13G2 and the scaled device architecture of the D7 transistor. However, these wafer splits also include process optimizations which lead to changes of the vertical doping profile and the conductivity of the external base regions. For this reason, it is difficult to estimate the impact of the scaled device dimensions independently of the doping profile from the experimental results only.

In this section, device simulation is used to get a more detailed understanding of how scaling of the lateral device dimensions contributed to the performance enhancement of the new HBT generation. Therefore, simulations of a HBT with the lateral geometry of the reference technology (G2) are compared to simulations of the optimized geometry (D7). The same doping profile is used for all simulations, so that the differences in the electrical characteristics are only determined by the device geometry. The doping profile with minimal extension of the external p-doping (No. 4 in Tab. 5.1) is used for all device modifications.
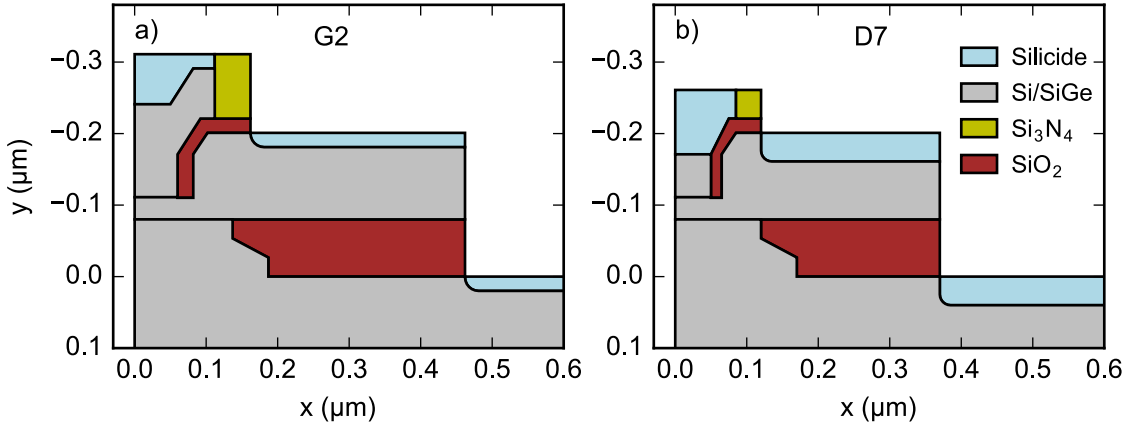
Figure 5.16.: Simulated device structures. a) Reference structure with G2-like geometry. b) Scaled structure with the lateral geometry of the D7 HBT.

Figure 5.16 shows the 2D device structures that are used in the simulation to represent the reference and the scaled HBT design. Several intermediate stages of the device scaling are simulated to determine their individual impact on the transistor performance. Starting with G2, the scaled D7 device is obtained by a stepwise application of the following modifications:

1. The width of the emitter window W_E was reduced from $0.12\,\mu\text{m}$ to $0.1\,\mu\text{m}$.

2. The thickness of the emitter has been reduced by $50\,\text{nm}$.

3. The thickness of the base-emitter spacer has been reduced.

4. The silicide thickness has been increased from $20\,\text{nm}$ to $40\,\text{nm}$.

5. The width of the base poly W_{BPLY} has been reduced by $0.1\,\mu\text{m}$.

Simulated f_T and f_{max} of the subsequent device modifications are plotted in Fig. 5.17. The corresponding electrical parameters are summarized in Table 5.4. The reduction of W_E from $0.12\,\mu$m to $0.1\,\mu$m results in a small reduction of f_T from $523\,$GHz to $512\,$GHz due to the stronger impact of the capacitance from the device perimeter. At the same time, f_{max} increases by $11\,\%$ from $693\,$GHz to $766\,$GHz due to a reduction of the base resistance. This behavior has already been discussed in Section 5.1. The next modification is the reduction of the emitter thickness H_E, which results in a decrease of the emitter resistance. The smaller R_E leads to a slight increase of f_T. A greater impact is achieved by the reduction of the width of the EB-spacer. The width of the oxide between emitter and external base is reduced from $22\,$nm to $15\,$nm and the width of the nitride is reduced from $50\,$nm to $35\,$nm. Additionally the width of the emitter poly is reduced by $10\,$nm. The optimized spacer leads to a significant reduction of R_B which results in an increase of f_{max} by $10\,\%$ from $766\,$GHz to $843\,$GHz. The transit frequency, however, decreases a bit due to a higher C_{EB}. Increasing the thickness of the silicide leads to a small increase of f_{max}. A scaled base poly width results in to a reduction of the base-collector capacitance from which both f_T and f_{max} benefit.
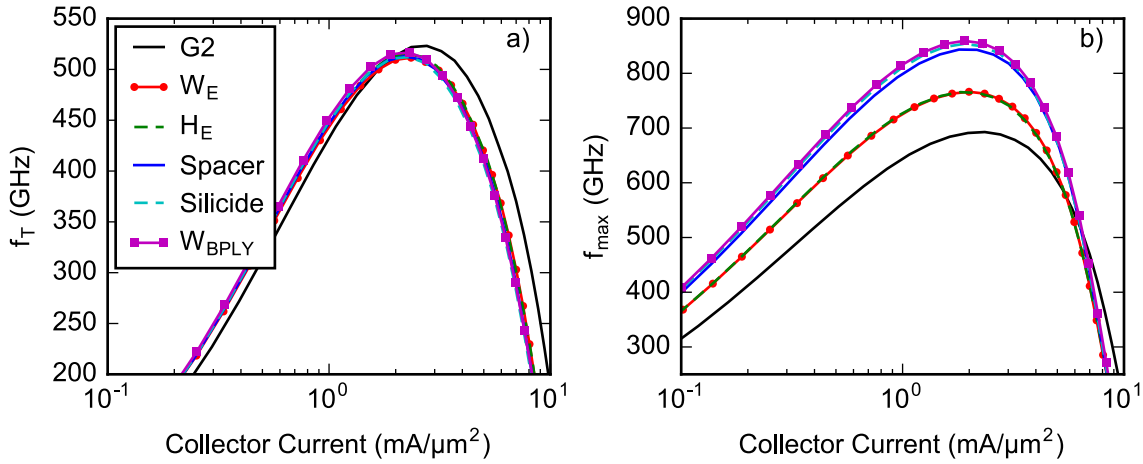
Figure 5.17.: a) Transit frequency vs. collector current density. b) Maximum oscillation frequency vs. collector current density.

Table 5.4.: Simulated device darameters for various modifications of the HBT structure.

#	Modification	f_T GHz	f_{max} GHz	C_{EB} fF/μm	C_{BC} fF/μm	R_B $\Omega\mu$m	R_C $\Omega\mu$m	R_E $\Omega\mu$m
Reference		523	693	2.112	1.314	38.1	13.2	13.2
1	W_E reduced	512	766	1.875	1.211	34.7	14.0	15.8
2	H_E reduced	515	766	1.879	1.212	38.2	14.0	13.4
3	Thinner Spacer	513	843	1.938	1.180	32.7	14.4	13.4
4	Thicker Silicide	512	854	1.939	1.180	33.1	14.4	12.7
5	W_{BPLY} reduced	517	859	1.939	1.137	33.7	13.6	12.6

Summarizing the results above, both scaling of the emitter window and the EB spacer width contribute significantly to the strong increase in f_{max} of the D7 HBT compared to the G2 HBT. However, these modifications also lead to a degradation of f_T which has to be compensated by an optimized vertical profile. A reduction of the base poly width is beneficial for both f_T and f_{max}.

The wafer split D7a presented in [21] contains modifications 1 to 4 from Table 5.4 as well as additional process changes which alter the vertical profile and increase the conductivity of the external base region. Compared to a standard SG13G2 wafer, these changes lead to an increase of f_T by 7 % from 314 GHz to 337 GHz and to an increase of f_{max} by 43 % from 414 GHz to 591 GHz. The simulation yields a decrease of f_T by 2 % and an increase of f_{max} by 23 % if modifications 1 to 4 are applied. The absolute values of the simulation results are significantly higher than the measured values because a more advanced vertical profile is used and because self heating is omitted. Comparing the relative changes from simulation and experiment leads to the conclusion that roughly half of the f_{max} enhancement can be attributed to the scaled lateral dimensions. The other half can be attributed to the additional process changes, namely the higher emitter doping, the reduced temperature of the final RTP which reduces broadening of the base doping, the enhanced conductivity in the external base and the SIC implantation trough a hard mask instead of a resist mask which helps to reduce C_{BCx}.

5.6. Summary

In Chapter 5, several possible modifications of the lateral structure of the SiGe HBT have been investigated by means of simulation, regarding their impact on f_T and f_{max}. The simulation has shown, that a reduction of the collector window width as well as a thinner oxide between collector well and external base can improve the RF-performance.

The simulations have shown that out-diffusion of boron can lead to parasitic barriers in the transitional region between internal and external base. Such barriers lead to a strong degradation of f_T and f_{max}. This effect is similar to the well known effect of parasitic barriers in the internal base [109]. The difference is, that it is not caused by boron out-diffusion from the internal base doping but from diffusion of boron from the external base into the inner transistor. However, further experimental investigations are necessary to clarify if this effect plays a role in real transistors.

Furthermore, the comparison between the lateral geometry of the DOTSEVEN HBT and the reference geometry of SG13G2 has shown that geometry scaling alone leads to an increase of f_{max} by almost 25 %.

6. Conclusions and Outlook

The primary objective of this thesis is to evaluate the ability of state-of-the-art hydrodynamic simulation to predict the RF-performance of advanced SiGe HBTs. For this purpose, measured and simulated DC characteristics and RF figures of merit from SiGe HBTs with a vertical doping profile tailored for high transit frequencies have been compared. In addition, the simulations provided valuable insight on performance-limiting mechanisms in the HBTs.

The commercial TCAD tool Sentaurus Device has been used for the hydrodynamic simulations. Calibrated parameter models, which were developed by Sasso et al. during the DOTFIVE project, are applied in this work. The effective bandgap in the SiGeC base layer has been determined experimentally for different Ge mole fractions. The obtained bandgaps are in good agreement with published data extracted from SiGe HBTs. No significant impact of carbon on the effective bandgap was found for concentrations up to $1.5 \times 10^{20}\,\mathrm{cm}^{-3}$. Thus, the literature values for the effective bandgap in SiGe HBTs can also be used to simulate SiGeC HBTs with typical carbon concentrations of about $10^{20}\,\mathrm{cm}^{-3}$.

For a meaningful comparison between measurement and simulation, it is necessary that the doping profile and the device geometry, which define the simulation domain, closely correspond with the real device. Based on a comprehensive experimental characterization, a simulation structure has been created which comprises all relevant features of the reference transistors. Long devices have been used for this comparison, so that charge transport can be described sufficiently accurate by a 2D simulation domain. The precise knowledge of the vertical doping profile is particularly important to assess the predictive power of the simulation. Here, the vertical profile has been determined by a combination of different experimental techniques in order to achieve the best available accuracy.

A good agreement between measured and simulated Gummel and output characteristics has been achieved by an adjustment of the bandgap and the heat flux parameter f^{hf}. However, the simulated collector current ideality showed a considerable deviation from the measurement. It turned out that this deviation is related to the doping-induced bandgap narrowing model, which produces a bandgap peak in the base-emitter depletion region. Future work should address the question if a better agreement can be achieved by an improved bandgap narrowing model. A model which takes into account the dependence of the effective bandgap narrowing on the carrier density (as in [116]) might produce a more realistic bandgap in the depletion regions.

The comparison of experimental profile variations with simulation has shown that the impact of these variations on f_T is qualitatively well captured by the simulation. This implies that hydrodynamic device simulation is a useful tool for the optimization of the

vertical doping profile.

However, it was found that the dependence of f_T on the collector current I_C is not reproduced properly by the simulation. The degradation of f_T sets in at lower currents than in the experiment. Additionally, the transition to the high injection regime, where the Kirk effect leads to a rapid decrease of f_T, is less abrupt in the simulation. Overall, this results in an underestimation of the maximum f_T by about 10 %.

The quasi-static transit time analysis has been used to examine the impact of the profile variations on f_T. It revealed that a narrow BE depletion region leads to a higher f_T by reducing minority charge storage in the emitter and that the optimum position of the heterojunction (relative to the BE depletion region) is determined by a trade-off between charge storage in the emitter and in the base. Moreover, it was shown that even for such advanced vertical profiles, maximum f_T is still dominated by the transit times of the base and the base-collector depletion region. A significant improvement of f_T can thus be expected by further scaling of the base width and the BC junction doping. The simulation suggests that moderate changes of the vertical profile can increase f_T of the present HBTs by about 30 %.

Considerable improvements can also be expected from a reduction of device parasitics such as the series resistance of the emitter and the collector. The simulated transit frequency of the inner 1D transistor is about 90 % higher than f_T of the full device. This shows that the present vertical doping profile offers a significant potential for further performance improvements.

Possible modifications of the lateral device architecture have been investigated by simulation. A strong effect of the external base doping on f_T was found: Enhanced boron out-diffusion from the external base link can create parasitic barriers in the conduction band, which lead to enhanced charge storage in the base link region. Such barriers would have a particularly strong impact on laterally scaled transistors where the relative contribution of the device perimeter to the total charge storage increases. However, further experimental investigations are needed to figure out if this effect is relevant in real HBTs.

The comprehensive comparison between measured and simulated electrical characteristics presented in this thesis reveals weaknesses in the hydrodynamic simulation of advanced SiGe HBTs. The measured $f_T(I_C)$ characteristics could not be reproduced properly by the simulation. This suggests that the hydrodynamic transport model is not sufficiently accurate for a reliable quantitative prediction of the high-frequency performance of the next generation of SiGe HBTs. In future work, it should be investigated if simulations based on BTE are capable to reproduce the measured f_T characteristics. Furthermore, it should be assessed if a more realistic model of the effective bandgap narrowing improves the accuracy of the simulation. Nevertheless, the HD simulations were able to qualitatively reproduce the impact of changes of the vertical profile and variations of the lateral device geometry. Therefore, HD simulation continues to be a useful tool for the investigation and optimization of modern SiGe HBTs.

A. Numerical Parameters of the Physical Models

A.1. Parameter Values of the Energy Relaxation Time Model

Table A.1.: Parameter values are taken from [8].

Parameter	Unit	Value	Parameter	Unit	Value
$\tau_{w,0,Si}$	ps	0.391	$C_{1,Si_{0.7}Ge_{0.3}}$	1	0.0028
$\tau_{w,0,Si_{0.7}Ge_{0.3}}$	ps	0.449	C_c	1	-0.001818
C_τ	ps	-0.05	C_0	1	0
$\tau_{w,1}$	ps	-0.14434	C_2	1	-0.059
$C_{1,Si}$	1	0.00135	C_3	1	0.0107

A.2. Parameter Values of the Mobility Model

Table A.2.: Parameters for Si taken from [8]. Normalized temperature $T_n = T/300K$.

Parameter	Unit	Electrons	Holes
$\mu_{max,0}$	cm^2/Vs	1421.6	485.51
γ		-2.24	-2.49
μ_{0d}	cm^2/Vs	49	$123.34 \cdot T_n^{-1.028}$
μ_{0a}	cm^2/Vs	$205.25 \cdot T_n^{-0.934}$	$46.42 \cdot T_n^{-0.627}$
μ_{1d}	cm^2/Vs	$\frac{24.81 \cdot T_n^2 + 84.6 \cdot T_n - 132.36}{T_n^2}$	$\frac{-68.8 \cdot T_n^2 + 300.19 \cdot T_n - 230.53}{T_n^2}$
μ_{1a}	cm^2/Vs	$\frac{94.533 \cdot T_n^2 + 518.28 \cdot T_n - 419}{T_n^2}$	$\frac{-5.783 \cdot T_n^2 + 81.146 \cdot T_n - 76.02}{T_n^2}$
C_{r1}	cm^{-3}	$8.393 \cdot 10^{16} \cdot T_n^{2.951}$	$1.329 \cdot 10^{17} \cdot T_n^{3.07}$
C_{r2}	cm^{-3}	$5.42 \cdot 10^{16} \cdot T_n^{3.045}$	$1.631 \cdot 10^{17} \cdot T_n^{3.111}$
C_{s1}	cm^{-3}	$1.81 \cdot 10^{19}$	$5.1 \cdot 10^{19}$
C_{s2}	cm^{-3}	$4.2 \cdot 10^{19}$	$5.8 \cdot 10^{19}$
α_1	1	0.68	0.7
α_2	1	0.7	0.77

A. Numerical Parameters of the Physical Models

Table A.3.: Parameters for $Si_{0.7}Ge_{0.3}$ taken from [8]. T_n is the normalized temperature $T_n = T/300K$.

Parameter	Unit	Electrons	Holes
$\mu_{max,0}$	cm^2/Vs	453.23	641.08
γ	1	-1.14	-2.118
μ_{0d}	cm^2/Vs	$91.59 \cdot T_n^{-1.0547}$	$130.24 \cdot T_n^{-1.332}$
μ_{0a}	cm^2/Vs	$191.58 \cdot T_n^{-0.92012}$	$40.848 \cdot T_n^{-0.633}$
μ_{1d}	cm^2/Vs	$\frac{5.494 \cdot T_n^2 + 95.87 \cdot T_n - 96.0}{T_n^2}$	$\frac{-90.178 \cdot T_n^2 + 325.95 \cdot T_n - 214.83}{T_n^2}$
μ_{1a}	cm^2/Vs	$\frac{49.072 \cdot T_n^2 + 87.32 \cdot T_n - 131.97}{T_n^2}$	$\frac{3.184 \cdot T_n^2 + 59.611 \cdot T_n - 57.091}{T_n^2}$
C_{r1}	cm^{-3}	$3.8812 \cdot 10^{16} \cdot T_n^{2.0845}$	$1.151 \cdot 10^{17} \cdot T_n^{4.2018}$
C_{r2}	cm^{-3}	$1.913 \cdot 10^{16} \cdot T_n^{2.4096}$	$1.3873 \cdot 10^{17} \cdot T_n^{3.2117}$
C_{s1}	cm^{-3}	$6 \cdot 10^{19}$	$2 \cdot 10^{20}$
C_{s2}	cm^{-3}	$5.4 \cdot 10^{19}$	$7 \cdot 10^{19}$
α_1	1	0.76	0.59
α_2	1	0.7	0.65

Table A.4.: Parameters for interpolation between Si and $Si_{0.7}Ge_{0.3}$ taken from [8].

Parameter	Unit	Electrons	Holes
α	1	0.487	0.548
C_μ	cm^2/Vs	2379.4	556.4

Table A.5.: Fitted parameters for the high-field mobility.

Parameter	Unit	Ge Mole Fraction	Electrons	Holes
β_0	1	0	1.4	1.2
		0.1	1.55	1.2
		0.2	1.75	1.2
		0.3	1.86	1.2
β_{exp}	1	0	0.1	0.1
		0.1	0.1	0.1
		0.2	0.1	0.1
		0.3	0.1	0.1
$v_{sat,0}$	10^7 cm/s	0	1.05	1.19
		0.1	0.77	1.01
		0.2	0.66	0.9
		0.3	0.62	0.84
$v_{sat,exp}$	1	0	0.3	0.3
		0.1	0.3	0.26
		0.2	0.3	0.23
		0.3	0.23	0.23

A.3. Parameter Values of the Effective DOS Model

Table A.6.: Parameter values are taken from [9].

Parameter	Unit	Value	Parameter	Unit	Value
$N_C^{Si}(300\,\mathrm{K})$	cm^{-3}	2.94×10^{19}	ΔE_C	eV	-0.6
$N_V^{Si}(300\,\mathrm{K})$	cm^{-3}	2.24×10^{19}	ΔE_{V1}	eV	-0.31
α_C	1	1.62	ΔE_{V2}	eV	-0.315
α_V	1	1.79	ΔE_{V3}	eV	-0.044

B. Depth Profiling Techniques

In the following, the basic principles of the experimental techniques that are used for the measurement of the the vertical Ge profile are briefly described.

B.1. X-ray diffractometry (XRD)

X-ray diffractometry is a technique to analyze lattice properties of crystalline materials. This method is based on the fact that the lattice acts as a diffraction grating for X-rays. The waves, that are scattered at adjacent planes in the crystal, interfere constructively when the Bragg's law $(2d \sin \theta = n\lambda)$ is satisfied, which allows to determine the lattice constant d. XRD can be used to determine the Ge content of pseudomorphic SiGe layers indirectly via the strain [79]. The strain, which is the change of the lattice constant relative to the substrate, depends on the Ge content of the SiGe layer. If the relation between Ge content and strain is known, one can determine the Ge content from the measured strain.

The determination of the Ge content is based on a comparison of the measured diffraction spectra with calculated spectra from a suitable strain model. Such a model requires an assumption on the shape of the Ge profile. Practically, it has to be possible to describe the Ge profile by a small set of parameters. These parameters can then determined by fitting the calculated diffraction spectra to the measurements.
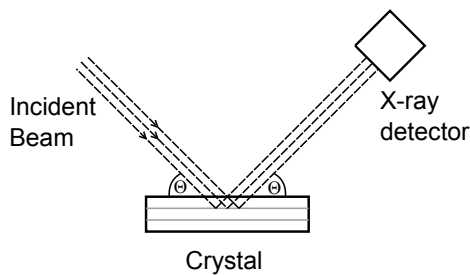
Figure B.1: Basic setup of X-ray diffraction measurements.

The major drawback of this method is that the lattice constant in the SiGe layer is also influenced by a possible carbon incorporation or by dopants. For this reason, the Ge profile of doped SiGe:C layers cannot be measured with such a high accuracy as in undoped SiGe layers.

B.2. **Secondary Ion Mass Spectroscopy (SIMS)**

Secondary ion mass spectroscopy is an analytical technique for the characterization of the chemical composition of surfaces. In SIMS, atoms and molecules are removed from the surface by bombardment with ions. Typical primary ions for the measurement of doping profiles in a silicon matrix are O_2^+ and Cs^+. A small fraction of the removed particles is emitted as (secondary) ions, which makes it possible to analyze them in a mass spectrometer. The rest is emitted as neutral atoms. The efficiency of the ionization (referred to as ion yield) strongly depends on the matrix material and the sputtering energy.

Figure B.2: Experimental setup of SIMS. Depth profiling is achieved by dynamic removal of material during the measurement.
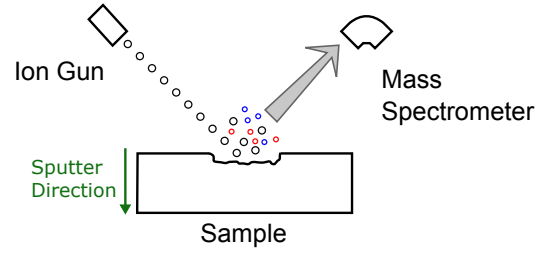
For a quantitative analysis in SIMS, the measured ion intensities has to be converted into material concentrations. For this purpose, the relative sensitivity factor RSF_E for the element E in the matrix material M is defined according to

$$\frac{I_M}{C_M} = RSF_E \cdot \frac{I_E}{C_E}, \tag{B.1}$$

with the intensities I_X and the and concentrations C_X. Depth profiles of the material composition can be obtained by continuous measurement over time, while the sample surface is removed layer by layer. This approach is called dynamic SIMS. It provides a depth resolution down to $0.3\,\text{nm}$ [117].

B.3. **Energy-Dispersive X-Ray Spectroscopy in a TEM**

Another method to determine the elemental composition in solid materials is to use energy-dispersive X-ray spectroscopy in an electron microscope. The electron beam of the microscope generates X-ray emission, which is characteristic of the elements contained in the sample. Energy dispersive spectroscopy of the emitted radiation provides information on the material composition at the position of the electron beam. Using a scanning transmission electron microscope (STEM) in EDX mode allows to create spatial maps of the elements distribution in the sample. In this way, depth profiles of SiGe layers can be measured with a high spatial resolution [78].

A major advantage of this method, compared to SIMS and XRD, is that the material composition can be measured at very small sample areas. This allows the Ge profile to be measured directly in the transistor which is not possible with SIMS or XRD.

Focused
Electron
Beam

EDX Detector

Scanning
Direction

X-ray

Sample
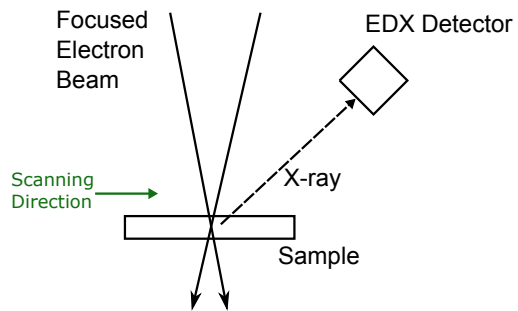
Figure B.3: Principle of element analysis by energy dispersive X-ray spectroscopy in an transmission electron microscope.

Bibliography

[1] J. D. Cressler, "SiGe HBT Technology: A New Contender for Si-Based RF and Microwave Circuit Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 5, pp. 572–589, 1998.

[2] J. D. Cressler, "A retrospective on the SiGe HBT: What we do know, what we don't know, and what we would like to know better," *2013 IEEE 13th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, SiRF 2013 - RWW 2013*, pp. 81–83, 2013.

[3] H. Rücker, B. Heinemann, and A. Fox, "Half-Terahertz SiGe BiCMOS Technology," in *Silicon Monolithic Integrated Circuits in RF Systems (SiRF), 2012 IEEE 12th Topical Meeting on*, pp. 133–136, 2012.

[4] A. Margomenos, "A Comparison of Si CMOS and SiGe BiCMOS Technologies for Automotive Radars," *2009 9th Topical Meeting on Silicon Monolithic Integrated Circuits in RF System, SiRF'09 - Digest of Papers*, pp. 1–4, 2009.

[5] U. R. Pfeiffer and E. Öjefors, "Terahertz imaging with CMOS/BiCMOS process technologies," *ESSCIRC 2010 - 36th European Solid State Circuits Conference*, pp. 52–60, 2010.

[6] W. Liebl, J. Boeck, K. Aufinger, *et al.*, "SiGe Applications in Automotive Radars," *ECS Transactions*, vol. 75, no. 8, pp. 91–102, 2016.

[7] A. Mai and M. Kaynak, "SiGe-BiCMOS based technology platforms for mm-wave and radar applications," *2016 21st International Conference on Microwave, Radar and Wireless Communications, MIKON 2016*, pp. 4–7, 2016.

[8] G. Sasso, N. Rinaldi, G. Matz, and C. Jungemann, "Accurate Mobility and Energy Relaxation Time Models for SiGe HBTs Numerical Simulation," in *2009 International Conference on Simulation of Semiconductor Processes and Devices*, pp. 241–244, 2009.

[9] G. Sasso, N. Rinaldi, G. Matz, and C. Jungemann, "Analytical Models of Effective DOS, Saturation Velocity and High-Field Mobility for SiGe HBTs Numerical Simulation," in *Simulation of Semiconductor Processes and Devices (SISPAD), 2010 International Conference on*, pp. 279–282, 2010.

[10] G. Wedel and M. Schröter, "Hydrodynamic simulations for advanced SiGe HBTs," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2010 IEEE*, pp. 237–244, 2010.

[11] C. Jungemann, A.-T. Pham, S.-M. Hong, L. Smith, and B. Meinerzhagen, "Deterministic solvers for the Boltzmann transport equation of 3D and quasi-2D electron and hole systems in SiGe devices," *Solid-State Electronics*, vol. 84, pp. 112–119, 2013.

[12] R. People and J. C. Bean, "Calculation of crital layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures," *Applied Physics Letters*, vol. 47, pp. 322–324, 1985.

[13] E. J. Prinz and J. C. Sturm, "Analytical Modeling of Current Gain-Early Voltage Product in $Si/Si_{1-x}Ge_x$ Heterojunction Bipolar Transistors," in *IEEE International Electron Devices Meeting*, pp. 853–856, 1991.

[14] P. Ashburn, *SiGe Heterojunction Bipolar Transistors*. John Wiley and Sons, Ltd, 2003.

[15] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Artech House, 2002.

[16] S. Decoutere, S. Van Huylenbroeck, B. Heinemann, *et al.*, "Advanced Process Modules and Architectures for Half-Terahertz SiGe:C HBTs," in *Proceedings of the IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 9–16, 2009.

[17] M. Schroter, H. Tran, and W. Kraus, "Germanium profile design options for SiGe LEC HBTs," *Solid-State Electronics*, vol. 48, no. 7, pp. 1133–1146, 2004.

[18] L. Lanzerotti, A. St. Amour, C. Liu, *et al.*, "$Si/Si_{1-x-y}/Ge_xC_y$/Si heterojunction bipolar transistors," *IEEE Electron Device Letters*, vol. 17, pp. 334–337, jul 1996.

[19] H. Rücker, B. Heinemann, D. Bolze, *et al.*, "Dopant diffusion in C-doped Si and SiGe: Physical model and experimental verification," *Electron Devices Meeting, 1999. IEDM Technical Digest. International*, pp. 345–348, 1999.

[20] Q. Z. Liu, J. W. Adkisson, V. Jain, *et al.*, "SiGe HBTs in 90nm BiCMOS Technology Demonstrating f_T/f_{MAX} 285GHz/475GHz through Simultaneous Reduction of Base Resistance and Extrinsic Collector Capacitance," *ECS Transactions*, vol. 64, no. 6, pp. 285–294, 2014.

[21] B. Heinemann, H. Rücker, R. Barth, *et al.*, "SiGe HBT with f_T/f_{max} of 505GHz/720 GHz," in *IEEE International Electron Devices Meeting 2016*, 2016.

[22] D. L. Harame, G. Freeman, and J. Rieh, "Look into the Future for SiGe HBTs," in *2003 International Symposium on Compound Semiconductors*, pp. 207–208, 2003.

[23] P. Chevalier, T. F. Meister, B. Heinemann, *et al.*, "Towards THz SiGe HBTs," in *2011 IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 57–65, 2011.

[24] M. Khater, J.-S. Rieh, T. Adam, *et al.*, "SiGe HBT Technology with $f_{max}/f_T =$ 350/300 GHz and Gate Delay below 3.3 ps," in *IEEE International Electron Devices Meeting 2004*, pp. 247–250, 2004.

[25] R. Lachner, "Towards 0.7 Terahertz Silicon Germanium Heterojunction Bipolar Technology – The DOTSEVEN Project," *ECS Transactions*, vol. 64, pp. 21–37, aug 2014.

[26] B. Heinemann, R. Barth, D. Bolze, *et al.*, "SiGe HBT technology with f_T/f_{max} of 300GHz/500GHz and 2.0ps CML gate delay," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, pp. 30.5.1–30.5.4, 2010.

[27] J. Böck, K. Aufinger, S. Boguth, *et al.*, "SiGe HBT and BiCMOS Process Integration Optimization within the DOTSEVEN Project," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 121–124, 2015.

[28] P. Chevalier, G. Avenier, G. Ribes, *et al.*, "A 55 nm Triple Gate Oxide 9 Metal Layers SiGe BiCMOS Technology Featuring 320 GHz f_T 370 GHz f_{MAX} HBT and High-Q Millimeter-Wave Passives," in *IEEE International Electron Devices Meeting*, pp. 3.9.1–3.9.3, 2014.

[29] V. P. Trivedi, J. P. John, J. Young, *et al.*, "A 90nm BiCMOS technology featuring 400GHz f_{MAX} SiGe:C HBT," *2016 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 60–63, 2016.

[30] J. J. Pekarik, J. Adkisson, P. Gray, *et al.*, "A 90nm SiGe BiCMOS technology for mm-wave and high-performance analog applications," in *Proceedings of the IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 92–95, 2014.

[31] P. Hurwitz, R. Kanawati, K. Moen, *et al.*, "Advances in RF foundry technology for wireless and wireline communications," *SiRF 2016 - 2016 IEEE 16th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 5–8, 2016.

[32] S. Li and Y. Fu, *3D TCAD Simulation for Semiconductor Processes, Devices and Optoelectronics*. Springer-Verlag New York, 2012.

[33] T. Grasser, H. Kosina, and S. Selberherr, "Investigation of spurious velocity overshoot using Monte Carlo data," *Applied Physics Letters*, vol. 79, no. 12, pp. 1900–1902, 2001.

[34] R. Kosik, *Numerical Challenges on the Road to NanoTCAD*. PhD thesis, Technischen Universität Wien, 2004.

[35] S. Keith, C. Jungemann, S. Decker, *et al.*, "Full-band Monte Carlo device simulation of a Si/SiGe-HBT with a realistic Ge profile," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 219–222, 1999.

[36] S. M. Hong and C. Jungemann, "Electron transport in extremely scaled SiGe HBTs," *Proceedings of the IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 67–74, 2009.

[37] Synopsys, "Sentaurus Device User Guide," 2016.

[38] M. Al-Sa'di, V. D'Alessandro, S. Fregonese, *et al.*, "TCAD simulation and development within the European DOTFIVE project on 500 GHz SiGe:C HBTs," in *Microwave Integrated Circuits Conference (EuMIC)*, pp. 29–32, 2010.

[39] G. L. Patton, S. S. Iyer, S. L. Delage, S. Tiwari, and J. M. Stork, "Silicon-Germanium-Base Heteroiunction Bipolar Transistors By Molecular Beam Epitaxy," *IEEE Electron Device Letters*, vol. 9, no. 4, pp. 65–67, 1988.

[40] C. King, J. Hoyt, and J. Gibbons, "Bandgap and transport properties of $Si_{1-x}Ge_x$ by analysis of nearly ideal $Si/Si_{1-x}Ge_x/Si$ heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 36, no. 10, pp. 2093–2104, 1989.

[41] S. Jain, J. Poortmans, S. Iyer, *et al.*, "Electrical and optical bandgaps of Ge_xSi_{1-x} strained layers," *IEEE Transactions on Electron Devices*, vol. 40, no. 12, pp. 2338–2343, 1993.

[42] Z. Matutinovic-Krstelj, V. Venkataraman, E. J. Prinz, J. C. Sturm, and C. W. Magee, "Base resistance and effective bandgap reduction in n-p-n $Si/Si_{1-x}Ge_x/Si$ HBTs with heavy base doping," *Electron Devices, IEEE Transactions on*, vol. 43, no. 3, pp. 457–466, 1996.

[43] I. Anteney, G. Lippert, P. Ashburn, *et al.*, "Electrical determination of bandgap narrowing and parasitic energy barriers in SiGe and SiGeC heterojunction bipolar transistors," in *Workshop on High Performance Electron Devices for Microwave and Optoelectronic Applications*, vol. 2, pp. 55–60, 1997.

[44] J. Eberhardt and E. Kasper, "Bandgap narrowing in strained SiGe on the basis of electrical measurements on Si/SiGe/Si hetero bipolar transistors," *Materials Science and Engineering: B*, vol. 89, pp. 93–96, feb 2002.

[45] M. Lundstrom, *Fundamentals of Carrier Transport.* Cambridge University Press, 2000.

[46] T. Grasser, T. W. Tang, H. Kosina, and S. Selberherr, "A Review of Hydrodynamic and Energy-Eransport Models for Semiconductor Device Simulation," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 251–273, 2003.

[47] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh, "Multidimensional Spherical Harmonics Expansion of Boltzmann Equation for Transport in Semiconductors," *Applied Mathematics Letters*, vol. 5, pp. 85–90, may 1992.

[48] T. Grasser, H. Kosina, M. Gritsch, and S. Selberherr, "Using six moments of Boltzmann's transport equation for device simulation," *Journal of Applied Physics*, vol. 90, no. 5, pp. 2389–2396, 2001.

[49] K. Blotekjaer, "Transport equations for electrons in two-valley semiconductors," *IEEE Transactions on Electron Devices*, vol. 17, no. 1, pp. 38–47, 1970.

[50] A. Schenk, *Advanced Physical Models for Silicon Device Simulation.* Springer-Verlag Wien, 1998.

[51] R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers," *Phys. Rev.*, vol. 126, pp. 2002–2014, jun 1962.

[52] E. Azoff, "Generalized energy-momentum conservation equations in the relaxation time approximation," *Solid-State Electronics*, vol. 30, pp. 913–917, sep 1987.

[53] Dotfive, "Website of the Dotfive Projekt," 2011.

[54] B. Gonzalez, V. Palankovski, H. Kosina, A. Hernandez, and S. Selberherr, "An energy relaxation time model for device simulation," *Solid-State Electronics*, vol. 43, pp. 1791–1795, sep 1999.

[55] D. M. Caughey and R. E. Thomas, "Carrier Mobilities in Silicon Empirically Related to the Doping and Field," *Proceedings of the IEEE*, vol. 55, no. 11, pp. 2192 – 2193, 1967.

[56] W. Hänsch, M. Orlowski, and W. Weber, "THE HOT-ELECTRON PROBLEM IN SUBMICRON MOSFET," *Le Journal de Physique Colloques*, vol. 49, pp. C4–597–C4–606, sep 1988.

[57] G. Masetti, M. Severi, and S. Solmi, "Modeling of Carrier Mobility Against Carrier Concentration in Arsenic-Doped, Phosphorus-Doped, and Boron-Doped Silicon," *IEEE Transactions on Electron Devices*, vol. 30, no. 7, pp. 764–769, 1983.

[58] S. Reggiani, M. Valdinoci, L. Colalongo, *et al.*, "Electron and hole mobility in silicon at large operating temperatures - Part I: Bulk mobility," *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 490–499, 2002.

[59] G. Sasso, *Transport Models and Advanced Numerical Simulation of Silicon-Germanium Heterojunction Bipolar Transistors.* PhD thesis, University of Naples Federico II, 2010.

[60] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region," *Solid State Electronics*, vol. 28, no. 11, pp. 1101–1103, 1985.

[61] D. B. M. Klaassen, J. W. Slotboom, and H. C. De Graaff, "Unified apparent bandgap narrowing in n- and p-type silicon," *Solid-State Electronics*, vol. 35, no. 2, pp. 125–129, 1992.

[62] F. Bärwolf, I. Costina, Y. Yamamoto, and A. Scheit, "DSIMS Quantitative Calibration and Analysis of Dopant Concentration in Homogeneously and Graded SiGe Layer," in *SIMS Europe Münster*, 2016.

[63] B. Barbalat, T. Schwartzmann, P. Chevalier, *et al.*, "The effect of carbon on neutral base recombination in high-speed SiGeC heterojunction bipolar transistors," *Semiconductor Science and Technology*, vol. 22, no. 1, pp. 99–102, 2007.

[64] J. Sturm, "Si/SiGe/Si heterojunction bipolar transistors," in *Properties of strained and relaxed Silicon Germanium* (E. Kasper, ed.), ch. 7, pp. 193–204, inspec, 1995.

[65] S. Swirhun, Y.-H. Kwark, and R. Swanson, "Measurement of electron lifetime, electron mobility and band-gap narrowing in heavily doped p-type silicon," *1986 International Electron Devices Meeting*, vol. 32, pp. 0–3, 1986.

[66] E. J. Prinz, P. M. Garone, P. V. Schwartz, X. Xiao, and J. C. Sturm, "The effect of base-emitter spacers and strain dependent densities of states in $Si/Si_{1-x}Ge_x/Si$ heterojunction bipolar transistors," in *IEEE International Electron Devices Meeting*, pp. 639–642, 1989.

[67] F. M. Bufler, *Full-Band Monte Carlo Simulation of Electrons and Holes in Strained Si and SiGe*. PhD thesis, Universität Bremen, 1998.

[68] C. G. Van de Walle, "Theoretical calculations of semiconductor heterojunction discontinuities," *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures*, vol. 4, no. 8, p. 1055, 1986.

[69] A. S. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt, "Defect-free band-edge photoluminescence and band gap measurement of pseudomorphic SiGeC alloy layers on Si (100)," *Applied Physics Letters*, vol. 67, no. 26, p. 3915, 1995.

[70] K. Yuki, K. Toyoda, T. Takagi, *et al.*, "Bandgap and Strain Engineering in SiGeC Heterojunction Bipolar Transistors," *Japanese Journal of Applied Physics*, vol. 40, no. 4, pp. 2633–2636, 2001.

[71] Y. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, pp. 149–154, jan 1967.

[72] R. Braunstein, A. Moore, and F. Herman, "Intrinsic Optical Absorption in Germanium-Silicon Alloys," *Physical Review*, vol. 109, no. 3, pp. 695–710, 1958.

[73] M. Schröter, T. Rosenbaum, P. Chevalier, *et al.*, "SiGe HBT Technology: Future Trends and TCAD-Based Roadmap," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–19, 2016.

[74] T. Rosenbaum, O. Saxod, V. T. Vu, *et al.*, "Calibration of 1D doping profiles of SiGe HBTs," in *2015 IEEE Bipolar/BiCMOS Circuits and Technology Meeting - BCTM*, pp. 64–67, 2015.

[75] M. Schröter, G. Wedel, B. Heinemann, *et al.*, "Physical and Electrical Performance Limits of High-Speed SiGeC HBTs—Part I: Vertical Scaling," *IEEE Transactions on Electron Devices*, vol. 58, no. 11, pp. 3687–3796, 2011.

[76] J. Korn, H. Rücker, B. Heinemann, *et al.*, "Experimental and theoretical study of f_T for SiGe HBTs with a scaled vertical doping profile," in *Bipolar/BiCMOS Circuits and Technology Meeting - BCTM, 2015 IEEE*, pp. 117–120, 2015.

[77] A. Pawlak, S. Lehmann, and M. Schröter, "A Simple and Accurate Method for Extracting the Emitter and Thermal Resistance of BJTs and HBTs," in *Bipolar/BiCMOS Circuits and Technology Meeting - BCTM, 2014 IEEE*, pp. 175–178, 2014.

[78] M. A. Schubert, P. Zaumseil, I. Costina, and H. Rücker, "Quantitative energy dispersive X-ray spectroscopy on thin SiGe layers," in *The 16th European Microscopy Congress*, 2016.

[79] P. Zaumseil, D. Krüger, and R. Kurps, "Precise Measurement of Ge Depth Profiles in SiGe HBT's - a Comparison of Different Methods," *Solid State Phenomena*, vol. 95, pp. 473–482, 2003.

[80] Synopsys, "TCAD Sentaurus Release Notes Version M-2016.12 (STAR No. 9001056792)," tech. rep., Synopsys Inc., 2016.

[81] H. Rücker, B. Heinemann, R. Barth, and M. Lisker, "Impact of back end stress on SiGe bipolar transistors," in *Electron Devices Meeting (IEDM), 2011 IEEE International*, pp. 34.5.1–34.5.4, 2011.

[82] D. J. Paul, "Si/SiGe heterostructures: from material and physics to devices and circuits," *Semiconductor Science and Technology*, vol. 19, no. 10, pp. R75–R108, 2004.

[83] M. Schröter and S. Lehmann, "The rectangular bipolar transistor tetrode structure and its application," *IEEE International Conference on Microelectronic Test Structures*, pp. 206–209, 2007.

[84] T. Grasser, R. Quay, V. Palankovski, and S. Selberherr, "A global self-heating model for device simulation," in *European Solid-State Device Research Conference*, pp. 324–327, 2000.

Bibliography

[85] A. Magnani, G. Sasso, V. Alessandro, *et al.*, "Advanced thermal resistance simulation of SiGe HBTs including backend cooling effect," in *Thermal Investigations of ICs and Systems (THERMINIC), 2015 21st International Workshop on*, pp. 1–5, 2015.

[86] H. Kamrani, D. Jabs, V. Alessandro, *et al.*, "A Deterministic and Self-Consistent Solver for the Coupled Carrier-Phonon System in SiGe HBTs," *IEEE Transactions on Electron Devices*, vol. 64, no. 2, pp. 361–367, 2017.

[87] V. D'Alessandro and N. Rinaldi, "A critical review of thermal models for electrothermal simulation," *Solid-State Electronics*, vol. 46, no. 4, pp. 487–496, 2002.

[88] S. J. Mason, "Power Gain in Feedback Amplifier," *Transactions of the IRE Professional Group on Circuit Theory*, vol. CT-1, no. 2, pp. 20–25, 1954.

[89] G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, "A New Recombination Model for Device Simulation Including Tunneling," *IEEE Transactions on Electron Devices*, vol. 39, no. 2, pp. 331–338, 1992.

[90] A. Gruhle, "The Influence of Emitter-Base Junction Design on Collector Saturation Current, Ideality Factor, Early Voltage and Device Switching Speed of Si / SiGe HBT's," *IEEE Transactions on Electron Devices*, vol. 41, no. 2, pp. 198–203, 1994.

[91] H. K. Gummel and H. C. Poon, "An Integral Charge Control Model of Bipolar Transistors," *Bell System Technical Journal*, vol. 49, no. 5, pp. 827–852, 1970.

[92] M. Schröter and A. Chakravorty, *Compact Hierarchical Bipolar Transistor Modeling with HICUM*. World Scientific, 2010.

[93] A. Pawlak, M. Schröter, and A. Fox, "Geometry Scalable Model Parameter Extraction for mm-Wave SiGe-Heterojunction Transistors," in *Proceedings of the IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 127–130, 2013.

[94] A. Pawlak and M. Schröter, "Modeling of SiGe HBTs with (f_T, f_{max}) of (340, 560) GHz based on physics-based scalable model parmeter extraction," in *IEEE 17th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 100–102, 2017.

[95] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Springer Vienna, 1984.

[96] H. K. Gummel, "A Charge Control Relation for Bipolar Transistors," *The Bell System Technical Journal*, vol. 49, no. 1, pp. 115–120, 1970.

[97] B. Heinemann, *2D-Bauelementesimulation der elektischen Eigenschaften von SiGe-HBTs*. PhD thesis, Technische Universität Berlin, 1997.

[98] J. J. H. van den Biesen, "A Simple Regional Analysis of Transit Times in Bipolar Transistors," *Solid State Electronics*, vol. 29, no. 5, pp. 529–534, 1986.

[99] M. Schröter and H. Tran, "Charge-storage related parameter calculation for Si and SiGe bipolar transistors from device simulation," in *WCM, International Nano-Tech Meeting, Boston (MA)*, vol. 3, pp. 735–740, 2006.

[100] P. Agarwal, G. Hurkx, J. Donkers, and J. Slotboom, "On the delay times in vertically scaled SiGe HBTs," *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting, 2005.*, pp. 264–267, 2005.

[101] P. Palestri, M. Mastrapasqua, A. Pacelli, and C. A. King, "A Drift-Diffusion / Monte Carlo Simulation Methodology for $Si_{1-x}Ge_x$ HBT Design," *IEEE Transactions on Electron Devices*, vol. 49, no. 7, pp. 1242–1249, 2002.

[102] Y. Shi and G. Niu, "Vertical Profile Design and Transit Time Analysis of Nano-Scale SiGe HBTs for Terahertz f_T," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 213–216, 2004.

[103] J. Kirk, C.T., "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities," *IRE Transactions on Electron Devices*, vol. 9, no. 2, pp. 164–174, 1962.

[104] R. J. E. Hueting, J. W. Slotboom, A. Pruijmboom, *et al.*, "On the Optimization of SiGe-Base Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 43, no. 9, pp. 1518–1524, 1996.

[105] G. Wedel, "Terahertz SiGe HBT TCAD and numerical device simulation," in *IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS) - Short Course*, 2016.

[106] W. Kloosterman, J. Paasschens, and D. Klaassen, "Improved extraction of base and emitter resistance from small signal high frequency admittance measurements," *Bipolar/BiCMOS Circuits and Technology Meeting, 1999. Proceedings of the 1999*, pp. 93–96, 1999.

[107] J. Berkner, *Kompaktmodelle für Bipolartransistoren.* Expert Verlag Renningen, 2002.

[108] H. Rücker, B. Heinemann, W. Winkler, *et al.*, "A 0 .13μm SiGe BiCMOS Technology Featuring f_T/f_{max} of 240GHz/330GHz and Gate Delays Below 3ps," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 9, pp. 1678–1686, 2010.

[109] J. Slotboom, G. Streutker, a. Pruijmboom, and D. Gravesteijn, "Parasitic energy barriers in SiGe HBTs," *IEEE Electron Device Letters*, vol. 12, no. 9, pp. 486–488, 1991.

[110] Z. A. Shafi, C. J. Gibbings, P. Ashburn, *et al.*, "The Importance of Neutral Base Recombination in Compromising the Gain of Si/SiGe Heterojunction Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 38, no. 8, pp. 1973–1976, 1991.

[111] A. J. Joseph, J. D. Cressler, D. M. Richey, and R. C. Jaeger, "Neutral Base Recombination and its Influence on the Temperature Tependence of Early Voltage and Current Gain-Early Voltage Product in UHV/CVD SiGe Heterojunction Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 44, no. 3, pp. 404–413, 1997.

[112] S.-I. Kim, B. R. Ryum, S.-W. Kang, and W. Kim, "The Effect of a Parasitic Potential Barrier on the Neutral Base Recombination Current of Si/SiGe/Si DHBTs," *Solid-State Electronics*, vol. 37, pp. 517–519, mar 1994.

[113] G. Niu, J. D. Cressler, and A. J. Joseph, "Quantifying Neutral Base Recombination and the Effects of Collector-Base Junction Traps in UHV/CVD SiGe HBT's," *IEEE Transactions on Electron Devices*, vol. 45, no. 12, pp. 2499–2504, 1998.

[114] S. Reggiani, E. Gnani, M. Rudan, *et al.*, "Measurement and Modeling of the Electron Impact-Ionization Coefficient in Silicon Up to Very High Temperatures," *IEEE Transactions on Electron Devices*, vol. 52, no. 10, pp. 2290–2297, 2005.

[115] J. D. Hayden, D. Burnett, and J. Nangle, "A Comparison of Base Current Reversal and Bipolar Snapback in Advanced N-P-N Bipolar Transistors," *IEEE Electron Device Letters*, vol. 12, no. 8, pp. 407–409, 1991.

[116] A. Schenk, "Finite-temperature full random-phase approximation model of band gap narrowing for silicon device simulation," *Journal of Applied Physics*, vol. 84, no. 7, pp. 3684–3695, 1998.

[117] H. Czichos, T. Saito, and L. Smith, *Spinger Handbook of Materials Measurement Methods.* Springer, 2006.

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

BC	Base-Collector
BE	Base-Emitter
BGN	Bandgap Narrowing
BJT	Bipolar Junction Transistor
BTE	Boltzmann Transport Equation
CMOS	Complementary Metal-Oxide-Semiconductor
DD	Drift-Diffusion
DOS	Density of States
DPSA	Double Poly-Silicon Self-Aligned (emitter/base)
EAG	Evans Analytical Group
EDX	Energy Dispersive X-ray Spectroscopy
ET	Energy Transport
FOM	Figure of Merit
HBT	Heterojunction Bipolar Transistor
HD	Hydrodynamic
HICUM	High Current Model
ITRS	International Technology Roadmap for Semiconductors
MC	Monte Carlo
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
NBR	Neutral Base Recombination
QS	Quasi-Static
RF	Radio Frequency
SCR	Space Charge Region
SEG	Selective Epitaxially Grown (base)
SHE	Spherical Harmonics Expansion
SiGe(C)	Silicon Germanium (Carbon)
SIC	Selectively Implanted Collector
SIMS	Secondary Ion Mass Spectroscopy
SRH	Shockley-Read-Hall
TCAD	Technology Computer Aided Design
TEM	Transmission Electron Microscope
RSF	Relative Sensitivity Factor
RTA	Rapid Thermal Annealing
XRD	X-ray Diffractometry

Symbols

β	Current gain
C_{jEB}	Base-emitter depletion capacitance
C_{dEB}	Base-emitter diffusion capacitance
C_{EBi}	Internal base-collector capacitance
C_{EBx}	External base-collector capacitance
$\Delta E_{g,dop}$	Doping induced bandgap narrowing
$\Delta E_{g,Ge}$	Bandgap reduction due to germanium
D_n	Electron diffusivity
E_C/E_V	Conduction/valence band energy
E_{Fn}/E_{Fp}	Quasi Fermi energy of electrons/holes
E_g	Bandgap
$E_{g,eff}$	Effective bandgap in SiGe (including bandgap narrowing)
$E_{g,Si}$	Bandgap in (undoped) silicon
f_{max}	Maximum oscillation frequency
f_T	Transit frequency
f^{hf}	Heat flux parameter
f^{td}	Thermal diffusion parameter
g_m	Transconductance
h_{21}	Small-signal current gain
I_B/I_C	Base/collector current
k	Boltzmann constant
L_E	(Lateral) length of the emitter window
μ_e/μ_e	Electron/hole mobility
n	Electron density
N_C/N_V	Effective DOS in the conduction/valence band
n_{I_C}	Collector current ideality factor
p	Hole density
Q_b	Hole charge in the base
r	Energy flux coefficient
$R_E/R_B/R_C$	Emitter/base/collector resistance
R_{sbi}	Base sheet resistance
R_{sbx}	External base resistance
τ_f	Forward transit time
T_C	Carrier Temperature
T_L	Lattice Temperature
τ_w	Energy relaxation time
U	Mason's unilateral gain
v_{sat}	Saturation velocity
w_b	(Vertical) width of the neutral base
W_E	(Lateral) width of the emitter window

List of Publications

Publications

J. Korn, H. Rücker, B. Heinemann, A. Pawlak, G. Wedel and M. Schröter, "Experimental and theoretical study of f_T for SiGe HBTs with a Scaled Vertical Doping Profile", *2015 IEEE Bipolar/BiCMOS Circuits and Technology Meeting - BCTM*, Boston, MA, 2015, pp. 117-120.

J. Korn, H. Rücker and B. Heinemann, "Experimental Verification of TCAD Simulation for High-Performance SiGe HBTs," *2017 IEEE 17th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, Phoenix, AZ, 2017, pp. 94-96.

Co-authored Publications

H. Rücker, B. Heinemann, J. Korn, "Optimization of Vertical Doping Profiles for High-Speed SiGe HBTs", *2015 10th European Microwave Integrated Circuits Conference (EuMIC) - Workshop on SiGe for mm-Wave and THz*, Paris, 2015

B. Heinemann et al., "SiGe HBT with f_T/f_{max} of $505\,\mathrm{GHz}/720\,\mathrm{GHz}$," *2016 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, 2016, pp. 3.1.1-3.1.4.

J. Schmidt, J. Korn, G. G. Fischer and R. Sorge, "Impact of Breakdown Voltage on Gamma Irradiation Effects in $0.13\,\mu$m and $0.25\,\mu$m SiGe HBTs," in *IEEE Transactions on Nuclear Science*, vol. 64, no. 4, pp. 1037-1041, April 2017.

H. Rücker, J. Korn, J. Schmidt, "Operation of SiGe HBTs at Cryogenic Temperatures", *2017 IEEE Bipolar/BiCMOS Circuits and Technology Meeting* (accepted).