SPICE Modeling of TeraHertz Heterojunction bipolar transistors

Félix Stein

THESE

L'UNIVERSITE DE BORDEAUX

ECOLE DOCTORALE DE SCIENCES PHYSIQUES ET DE L'INGENIEUR

Présentée par

Felix STEIN

POUR OBTENIR LE GRADE DE
DOCTEUR

Spécialité : **Electronique**

SPICE MODELING OF TERAHERTZ HETEROJUNCTION BIPOLAR TRANSISTORS

Soutenue le : 16 Décembre 2014

Après avis de :

M. LALLEMENT Christophe	Professeur, Université de Strasbourg	Rapporteur
M. QUERE Raymond	Professeur, Université de Limoges	Rapporteur

Devant la Commission d'examen formée de :

M. LALLEMENT Christophe	Professeur, Université de Strasbourg	Rapporteur
M. QUERE Raymond	Professeur, Université de Limoges	Rapporteur
M. ZIMMER Thomas	Professeur, Université de Bordeaux	Président du jury
M. CELI Didier	Ingénieur, STMicroelectronics	Examinateur
M. FREGONESE Sébastien	Chargé de recherche, CNRS	Examinateur
Mme. MANEUX Cristell	Professeur, Université de Bordeaux	Directeur de thèse
M. DERRIER Nicolas	Ingénieur, STMicroelectronics	Co-directeur de thèse

Acknowledgements

I would like to first thank all the members of my dissertation committee for their time and patience as well as intellectual contributions to my development as a researcher and scientist. I would like to thank the reviewers Christophe Lallement and Raymond Quéré for their time, their valuable comments and suggestions to improve this thesis.

I would like to express my deepest thanks to Cristell Maneux for her contributions to this work and take this opportunity to thank you for your guidance and supervision. I greatly appreciated your help and advice over these past three years. Your input and assistance was very helpful.

Next I want to thank my industrial supervisor Nicolas Derrier. I enjoyed being taught by you immensely. I appreciate all your hard work. It has meant so much to me. You have introduced me to the art of modeling and supported me at times when I needed help. You taught me how to stay focused on success while navigating the challenges of working with new programs, and giving me insight on how to accomplish everyday corporate work assignments. Thank you for being a role model.

I want to give a special acknowledgment to my research advisor, Didier Céli. You were an excellent teacher and have inspired me to continue learning with an open and positive mind and I enjoyed being taught by you immensely. Thank you for sharing your wisdom and experience so generously with me along the way. You have given me words of consolation when I needed them. I am so thankful to have had a caring, motivating and inspirational mentor like you. Thank you for believing in me, encouraging me and strengthening me. You have absolutely been a pleasure to work with.

I want to offer my heartfelt thanks to my colleagues, inside STMicroelectronics and the IMS laboratory. I have had the honor of knowing you, working with you, and learning from you in the last years. You taught me more than I ever imagined possible when I came to this program three years ago. The creative environment and resources available combine to make this place stimulating, exciting, convivial and productive.

I also want to acknowledge my friends, both here and at a distance, who have helped me throughout this process by patiently listening to me complaining about the work ahead, helping me thinking positive about small accomplishments, and successfully

gotten me out of the house, enjoying the beautiful scenery of the mountains and nature from time to time. It is obvious how much I learned, gained, and enjoyed from the work in France. Without all the people there, it would not have been possible.

My family has helped me all along the way and has been a great source of strength and inspiration to me. I would like to thank them for always supporting me in everything I have chosen to do.

I am especially indebted to my parents for their constant sacrifice and support, not only during my doctoral program but throughout my life. Their caring in my success has made me the successful individual I am today.

Finally, I want to thank Feli for always believing in me and helping me believe I can achieve more than I ever dreamed I could. Coming home to you at the end of each day has made all of this hard work worthwhile.

Abstract

The aim of BiCMOS technology is to combine two different process technologies into a single chip, reducing the number of external components and optimizing power consumption for RF, analog and digital parts in one single package. Given the respective strengths of HBT and CMOS devices, especially high speed applications benefit from advanced BiCMOS processes, that integrate two different technologies.

For analog mixed-signal RF and microwave circuitry, the push towards lower power and higher speed imposes requirements and presents challenges not faced by digital circuit designs. Accurate compact device models, predicting device behaviour under a variety of bias as well as ambient temperatures, are crucial for the development of large scale circuits and create advanced designs with first-pass success.

As technology advances, these models have to cover an increasing number of physical effects and model equations have to be continuously re-evaluated and adapted. Likewise process scaling has to be verified and reflected by scaling laws, which are closely related to device physics.

This thesis examines the suitability of the model formulation for applicability to production-ready SiGe HBT processes. A derivation of the most recent model formulation implemented in HICUM version L2.3x, is followed by simulation studies, which confirm their agreement with electrical characteristics of high-speed devices. The fundamental geometry scaling laws, as implemented in the custom-developed model library, are described in detail with a strong link to the specific device architecture.

In order to correctly determine the respective model parameters, newly developed and existing extraction routines have been exercised with recent HBT technology generations and benchmarked by means of numerical device simulation, where applicable. Especially the extraction of extrinsic elements such as series resistances and parasitic capacitances were improved along with the substrate network.

The extraction steps and methods required to obtain a fully scalable model library were exercised and presented using measured data from a recent industry-leading $55nm$ SiGe BiCMOS process, reaching switching speeds in excess of $300GHz$. Finally the extracted model card was verified for the respective technology.

Titre : Modélisation compacte des transistors bipolaires fonctionnant dans la gamme TeraHertz

Résumé

Les études qui seront présentées dans le cadre de cette thèse portent sur le développement et l'optimisation des techniques pour la modélisation compacte des transistors bipolaires à hétérojonction (TBH). Ce type de modélisation est à la base du développement des bibliothèques de composants qu'utilisent les concepteurs lors de la phase de simulation des circuits intégrés. Le but d'une technologie BiCMOS est de pouvoir combiner deux procédés technologiques différents sur une seule et même puce. En plus de limiter le nombre de composants externes, cela permet également une meilleure gestion de la consommation dans les différents blocs digitaux, analogiques et RF. Les applications dites rapides peuvent ainsi profiter du meilleur des composants bipolaires et des transistors CMOS. Le défi est d'autant plus critique dans le cas des applications analogiques/RF puisqu'il est nécessaire de diminuer la puissance consommée tout en maintenant des fréquences de fonctionnement des transistors très élevées.

Disposer de modèlés compacts précis des transistors utilisés est donc primordial lors de la conception des circuits utilisés pour les applications analogiques et mixtes. Cette précision implique une étude sur un large domaine de tensions d'utilisation et de températures de fonctionnement. De plus, en allant vers des nœuds technologiques de plus en plus avancés, des nouveaux effets physiques se manifestent et doivent être pris en compte dans les équations du modèle. Les règles d'échelle des technologies plus matures doivent ainsi être réexaminées en se basant sur la physique du dispositif.

Cette thèse a pour but d'évaluer la faisabilité d'une offre de modèle compact dédiée à la technologie avancée SiGe TBH de chez ST Microelectronics. Le modèle du transistor bipolaire SiGe TBH est présenté en se basant sur le modèle compact récent HICUM version L2.3x. Grâce aux lois d'échelle introduites et basées sur le dessin même des dimensions du transistor, une simulation précise du comportement électrique et thermique a pu être démontrée.

Ceci a été rendu possible grâce à l'utilisation et à l'amélioration des routines et méthodes d'extraction des paramètres du modèle. C'est particulièrement le cas pour la détermination des éléments parasites extrinsèques (résistances et capacités) ainsi que celle du transistor intrinsèque..

Finalement, les différentes étapes d'extraction et les méthodes sont présentées, et ont été vérifiées par l'extraction de bibliothèques SPICE sur le TBH NPN Haute-Vitesse de la technologie BiCMOS avancée du nœud 55nm, avec des fréquences de fonctionnement atteignant 320/370GHz de f_T / f_{max}.

Keywords

Heterojunction bipolar transistors, silicon-germanium, SiGe, semiconductor device models, analytical geometrical model, charge-based model, ICCR, integrated circuit design, small-signal model, vertical bipolar junction transistor, device simulation, circuit simulation, test structure, base resistance, collector resistance, emitter resistance, Early effect

Mots clés

Transistors bipolaires à hétérojonction (TBH), TBH à base de SiGe, modélisation du transistor, simulation du circuit, modèle géométrique, conception des circuits intégrés (CI), paramètres en régime de petit signal, transistor bipolaire vertical auto-aligné, transistor bipolaire dans une simulation SPICE, structure de test, résistance de base, résistance collecteur, résistance d'émetteur, l'effet Early

Résumé substantiel

Les études qui sont présentées dans le cadre de cette thèse portent sur le développement et l'optimisation des techniques pour la modélisation compacte des transistors bipolaires à hétérojonction (TBH). Ce type de modélisation est à la base du développement des bibliothèques de transistors utilisés par les concepteurs de circuit lors de la phase de simulation et d‘optimisation des circuits intégrés. Les applications dites rapides peuvent ainsi profiter du meilleur des transistors bipolaires et des transistors MOS. Le défi est d'autant plus important dans le cas des applications analogiques/RF pour lesquelles il est nécessaire de diminuer la puissance consommée tout en maintenant des fréquences de fonctionnement très élevées typiquement proche de 400GHz.

L’augmentation de la fonctionnalité des circuits requiert l’utilisation de circuits RF à haute densité d’intégration comme dans le cas des circuits numériques CMOS qui authorizent la réduction de taille des puces électroniques. Le but d'une technologie BiCMOS est de pouvoir combiner deux procédés technologiques sur une seule puce pour bénéficier à la fois des fonctions hautes fréquences des transistors bipolaires et de la densité d’intégration des transistors MOS. Ainsi, le procédé de fabrication par micro-usinage de couches compatible avec les procédés CMOS, est très favorable à l’intégration des transistors bipolaires grande vitesse. Cette approche permet de réduire les coûts de fabrication des circuits fonctionnant au-delà de 60 GHz.

Cette bande de fréquence est très utilisée par plusieurs applications grand public. C’est le cas des communications sans fil qui intègrent des objets de communication sans fil, des réseaux de communication locaux (WiFi), des systèmes de communication optiques et les communications par satellites. Par ailleurs, les applications radar pour l’industrie automobile est intègrent également ce type de transistors et circuits qui nécessitent des technologies de plus en plus sophistiquées.

Pour la réalisation de circuits intégrés, les transistors doivent être modélisés selon un standard : la modélisation compacte. Le but est de décrire le comportement électrique d’un transistor élémentaire par des équations analytiques à l’aide d’un simulateur numérique. Face à l’évolution continue des technologies il faut sans cesse vérifier et valider les équations des modèles pour vérifier leur sens physique. Ces modèles électriques sont également capables de

reproduire avec précision le comportement des transistors dans les régimes de fonctionnement statiques et dynamiques pour une gamme de températures comprises entre -40C et 150C.

Les modèles compacts donc contiennent d'une part des expressions basées sur la physique et d'autre part une simplification empirique favorable à la réduction du temps de simulation des circuits. Ces modèles s'adaptent aux différents technologies à partir de jeu paramètres spécifiques à chacune d'elle.

Ce thèse est spécialement dédiée au modèle HICUM qui est un modèle compact de type contrôle de charges, initialement propose par Gummel et Poon. Le modèle HICUM intègre tous les mécanismes de transport de charges nécessaire à la description du fonctionnement du transistor bipolaire à homojonction. Il a été généralisé pour permettre de modéliser les technologies bipolaires à hétérojonction récentes qui sont basées sur un profil graduel de germanium dans la base.

La précision des modèles compacts implique une étude de son comportement sur une large gamme de tensions d'utilisation et de températures de fonctionnement. Etant un modèle semi-physique, théocratiquement l'ensemble des paramètres de HICUM peuvent être calculés à partir des données technologiques et des mesures électriques. La valeur de chaque élément du circuit équivalent est ainsi reliée à une partie physique du transistor à partir des données technologiques ou des modes de fonctionnement.

En allant vers des nœuds technologiques de plus en plus avancés, de nouveaux effets physiques se manifestent et doivent être pris en compte dans les équations du modèle compact. Les règles d'échelle des technologies plus matures doivent ainsi être réexaminées en se basant sur les mécanismes physiques qui gouvernent le fonctionnement du transistor de nouvelle génération.

Ainsi, un modèle s'accompagne de stratégies d'extraction de paramètres. Ceci est rendu possible grâce à l'utilisation et à l'amélioration des routines et méthodes d'extraction des paramètres du modèle. C'est particulièrement le cas pour la détermination des éléments parasites extrinsèques (résistances et capacités) mais aussi ceux du transistor intrinsèque.

L’augmentation des densités de courant pour atteindre de fortes fréquences de coupure provoque le déclenchement du mécanisme d’auto-échauffement qui est autant plus important que la puissance dissipée est élevée. L’extraction de la résistance thermique par des mesures électriques est basée sur les approximations du comportement du transistor à fort courant. Ainsi, une partie de cette thèse porte sur la réalisation et l’analyse de mesures pulsées dont l’objectif est d’améliorer la prise en compte du phénomène d’auto-échauffement à fort courant et d’évaluer les valeurs des paramètres liés. En appliquant de courtes impulsions (80ns), il est possible de réduire l’impact du mécanisme d’auto-échauffement pour tendre vers des meures quasi-isothermes et discriminer ainsi le comportement électrique pur de celui thermo-électrique. Les bases de l’extraction de paramètre adaptées aux mesures pulsées sont présentées.

Le manuscrit comporte trois chapitres principaux. Apres une courte introduction, le deuxième chapitre est dédié aux les effets physiques des TBH fabriqués en processus BiCMOS optimisés pour les hautes fréquences. Il s’agit d’une analyse aussi des points critiques liés aux mécanismes physiques spécifiques de ces transistors. En particulier, ce chapitre présente une analyse basée sur des simulations numériques 2D de type TCAD.
Une attention particulière est apportée sur la description et l’effet de la base graduelle constitué d’un profil de Germanium triangulaire et d’un profil de Germanium trapézoïdal donnant lieu à la réduction progressive de la bande interdite le long de la base. Le but principal de ce type de base graduelle est d’accélérer les porteurs à travers la base et de diminuer le temps de transit dans la base. Grace au schéma électrique équivalent et aux équations physiques des TBH SiGe, les effets visibles sur les caractéristiques électriques sont reproduits et les équations pour le modèle HICUM version 2L3x sont dérivées.

Cette thèse avait pour but d'évaluer la faisabilité d'une offre de modèle compact dédiée à la technologie avancée SiGe TBH de chez ST Microelectronics. Spécialement les technologies les plus récentes du nœud technologique 130nm (BiCMOS9MW, B3T, B4T et B5T) ainsi que le nœud 55nm (BiCMOS55). Ces technologies TBH Si/SiGe:C présentent des fréquences de coupure f_T et f_{max} supérieures à 300GHz. Le modèle du transistor bipolaire SiGe TBH est présenté en se basant sur le modèle compact récent HICUM version L2.3x qui prend en compte finement les charges dans les différentes parties du transistor. Grace aux lois d'échelle introduites et basés sur les dimensions du transistor, une simulation précise du comportement électrique et thermique a pu être démontrée. Dans l’avant-dernier chapitre les règles d'échelle

spécifiques à la librairie de STMicroelectronics sont détaillées. Les effets couverts par le modèle sont reliés à la région du transistor spécifique permettant de développer le lien entre d'une part dimensions et données technologiques et d'autre part paramètres du modèle. Il s'agit de l'adaptation de lois physiques aux règles empiriques qui permettent d'avoir le meilleur compromis entre un sens physique et la simplicité de d'extraction des paramètres. Les structures spécifiques de test sont présentées et la précision des méthodes d'extraction est démontrée.

Finalement, dans le dernier chapitre, l'accent est mis sur la cohérence de l'ensemble de la stratégie d'extraction des paramètres du modèle en se basent sur les résultats de la caractérisation électrique statique et dynamique ainsi que sur les simulations numériques 2D. Les différentes étapes de la procédure extraction sont détaillées pour le TBH NPN haute-vitesse de la technologie BiCMOS du nœud 55nm, avec des fréquences de fonctionnement atteignant 320/370GHz de f_T/f_{max}. Même si les solutions présentées sont particulièrement adaptées à la technologie bipolaire de STMicroelectronique, les résultats obtenus sont facilement généralisables à l'ensemble des TBH SiGe.

Contents

Chapter 1

Introduction

1.1 Device Modeling for Heterojunction Bipolar Transistors

The development of heterojunction bipolar transistors (HBTs) started over two decades ago with the first demonstration of a silicon-germanium (SiGe) based HBT in 1987 [1]. In modern bipolar transistor technology, germanium (Ge) is alloyed with silicon (Si) to practice the so called band-gap-engineering, in order to extend the capabilities of conventional silicon bipolar devices.

SiGe based bipolar devices have a competitive advantage over classical III-V HBTs, that have lower levels of integration, are more difficult to manufacture, and have lower yield resulting in much higher cost. Compared to III-V devices Si-based technologies are advantageous given their excellent compatibility with CMOS processes, making SiGe technology a strong contender in several different market segments. In order to realize the integration of fast SiGe based bipolar transistors into an existing CMOS process flow, certain constraints are imposed (cf. App. B). Yet the advantage of simultaneous integration of analog and digital circuitry to build highly integrated mixed signal system-on-chip (SoC) solutions outweighs the effort for additional process steps. By maintaining the advantages of conventional silicon processing, BiCMOS technology is regarded as a key enabler for the introduction of fast HBT devices to mass production and thus the widespread use of analog mixed-signal circuits [2].

In the last years silicon based transistor development has focused on addressing the area of 'More than Moore' technologies for applications beyond $60GHz$. Starting in July 2011, the European RF2THz SiSoC[1] research project focuses on the establishment of silicon system-on-chip technology platforms for emerging radio frequency (RF), millimeter-wave (mmW) and TeraHertz (THz) consumer applications [3]. The frequency limit of modern SiGe based heterojunction bipolar transistors was pushed towards 0.5 TeraHertz in the DOTFIVE project [4] and following industry projects are targeting even higher operating frequencies in the near future [5].

[1] 'from *RF to* mmW and *THz Si*licon *S*ystem-*o*n-*C*hip technologies'

With over 30 research partners, the stated aim within the RF2THz project is to address new applications and markets such as $77GHz$/$120GHz$ automotive radars, mmW imaging and sensing, fast measurement equipment, $60GHz$ wireless networking as well as fast optical and RF wireless communication. All of them require high performance devices (regarding transmitted power, consumption, integration, isolation). Given their reliability under extreme environmental conditions, HBT based circuits are especially suited for the growing market area of automotive electronics.

A key in European research is to establish state-of-the-art microchip technologies and drive innovation through extension of large-scale semiconductor microchip manufacturing. One major goal was the creation of a $300mm$ BiCMOS silicon technology platform tailored for a wide range of everyday consumer applications and a variety of circuits and systems for the mass-market [6]. In order to address this need, STMicroelectronics, as part of the research project, was first to introduce a HBT technology to an advanced $300mm$ manufacturing process [7].

The technologies studied in this work (cf. App. C.2) have been developed by STMicroelectronics and manufactured within the 8" (B5T) and the new 12" (BiCMOS55) wafer manufacturing facility located in Crolles (France). The analyzed SiGe HBT technologies are explicitly tailored for high-frequency operation in a region, applicable for mmW technology. The B5T technology has been the last evolution of a $0.13\mu m$ prototype technology (bipolar only) for performance anticipation, yet will not go into mass production [8]. In contrast the BiCMOS55 platform is embedded in a CMOS 55nm node with the introduction of the high-speed HBT developped in the B5T technology. It suited for mass markets and scheduled to reach sufficient process stability and desired performances for series production by the end of the year 2015.

The ability to accurately simulate and analyze integrated circuits has become one of the most critical issues in SoC design. Integration of a large amount of mixed-signal circuitry however makes RF simulation a challenging field and prone to functional failures. Hence full-chip SPICE (Simulation Program with Integrated Circuit Emphasis) simulations of RF circuits are increasingly demanded, including accurate transistor level simulation. Efficient and accurate mixed-signal transistor level simulation is therefore a key asset for SoC success and first-pass RFIC designs.

In this context, compact modeling refers to modeling of large scale dynamical systems to predict the behavior of a circuit. A compact model targets the description of the electrical characteristics using terminal currents, charges and capacitances. Instead of using strictly physical and complex relations, engineers can often accept some approximations, in order to get a more robust and simple model and thereby reduce model complexity and computation effort to speed up the simulation of large designs. However to do so, a sound understanding of the underlying device physics is required. The relevance of physical effects and constraints imposed through model simplification have to be carefully considered in order to keep the model simple yet make it applicable to a variety of materials, technologies and processes.

1.2 The Limitations of Device Modeling

Advanced circuits nowadays contain several billion transistors and other electronic components in one chip package, while area consumption is reduced to a few square millimeters. Device performance compared to discrete parts is high because the components consume little power as a result of the small size and close proximity. However, given the higher process complexity with the move to more advanced nodes, higher manufacturing costs increase the demand for accurate designs with first-pass design success. The fast technology development therefore comes along with the need for accurate characterization methods, improved compact device models as well as robust and accurate model extraction, to enable reliable predictive simulation of transistor operation.

The ultimate goal of device modeling is to provide design kits with support for a commercial mixed signal platforms and design environments to foundry customers. In brief, a mixed-signal/RF process design kit (PDK) is a set of data files, enabling analog circuit designers to simulate integrated circuits (IC) in widely used software tools, dedicated to electronic design automation (EDA) for a given process with a range of available devices.

Compact transistor model parameters, as part of the PDK, allow for a complete design flow and can effectively lead to a first-pass design success, without the need for multiple prototypes or design iterations and first-to-market launch of new products and solutions with profitable success. In contrast to highly integrated CMOS based digital processor designs, where circuit simulation accuracy is optimized for extreme numbers of transistors and high integration levels, circuit simulation models for BiCMOS processes focus on precision for RF accuracy.

Along with the advancement of process technology, compact models have to keep up with the demand for lower switching delay and increased frequency to provide solutions for emerging THz electronics applications. Model limitations with regards to geometry scaling and bias dependence have been reached [9]. Especially the bias and temperature behaviour of the collector current at low and medium current densities was found to be unsatisfactory reproduced by the model. Concepts to overcome these apparent constraints have been proposed [10–12].

This work focuses on key aspects of device modeling for state-of-the-art SiGe bipolar junction transistors using the widespread HICUM/L2 device model. With the strong intention to use physics-based analytical approaches as far as possible, the release of a new HICUM version L2.3x brought increased flexibility [13]. However, the model simplifications had to be evaluated with regards to their suitability for modeling of high-speed integrated circuits.

In a first part this work focuses on the theoretical and experimental verification and validation of newly implemented model equations. Measured data and experimental results were obtained for recent industry-standard process generations in 130 and 55nm. Additionally, numerical simulations help to bridge the gap between the theory behind the simplification of model equations and the actual reality represented by a non-idealistic vertical device profile.

1.3 Fully Geometry Scalable Compact Device Modeling

Scaling as part of technology advancement introduces new electrical behavior into HBTs, that had been imperceptible or unknown before and hence not covered by hitherto existing models. However fully scalable compact models, describing the physics and operation of state-of-the-art HBT devices, are imperative for technology design and product development.

In order to provide design flexibility, the library of devices covers a variety of high-speed transistors with different spatial dimensions and device configurations. All of them have to be covered in the PDK to allow successful design for mmW and RFIC applications. Given these preconditions, a large part of the work presented in this thesis is devoted to the custom, fully geometry scalable model library, developed and maintained by STMicroelectronics.

Simple curve fitting and optimization based parameter determination for single geometries using measured characteristics may yield satisfactory yet unphysical results. Using a scalable approach, more devices are being used for parameter extraction, thus one can increase the model accuracy, since geometry uncertainty can be averaged out. Furthermore a model hierarchy tailored to an established process technology may be used for predictive modeling. Therefore existing geometry scaling laws implemented in the custom model library are revisited and validated.

Yet multi-device optimization for scalable model parameter extraction relies on a precise assessment of geometry-independent extrinsic parasitics. One of the key assets of this work was therefore the improvement extraction routines for external elements of the HICUM equivalent circuit. Special attention was paid to the improvement of the series resistance extraction (R_E, R_C, R_{Bx}) as well as the substrate network parameters. Furthermore the split of BE and BC related peripheral capacitances was revisited.

For the high-current region, where parameter extraction is difficult due to the strong interrelation of parameters, a geometry scalable procedure for improved extraction of the critical current (I_{CK}) was tested and the influence of device self-heating has been studied using novel measurement approaches with a pulsed RF measurement system.

Since model accuracy depends on both the compact device models itself as well as the determined parameters, the extraction flow was exercised and validated for multi-transistor geometries. A special focus was on the application of extraction routines and parameter determination procedures to the new 55*nm* BiCMOS technology.

1.4 Outline

This work is structured in three major chapters, reflecting the demands in industrial device modeling. The individual parts are focusing on the steps required to put in place a fully scalable model library and obtain the according model parameters. With respect to the model extraction flow used in an industry environment, the issues of physical model basis, the scalable library with their scaling laws and parameter extraction routines are examined in individual chapters.

After a general introduction, chapter 2 will focus on the device itself as well as the concept of compact modeling for RFIC design. Given the close relation of the HICUM model to the device physics of an ideal one-dimensional vertical transistor, basic device physics relevant to the essential model equations are presented, followed by an analysis of model limitations. This assessment of problems focuses on the issues seen with the the previous version of the HICUM model (L2.2x) for modeling of mmW devices. The effects seen in measured data are verified by means of numerical device simulations of realistic device profiles. The improved model formulation found in the latest release version of HICUM (L2.3x) is presented subsequently. By derivation of the model implementation, required simplifications are shown and their physical origin is retraced. Thereafter the new model formulation is benchmarked by comparison with simulated data. The general capability to reproduce the effects seen in measured data are confirmed versus bias and temperature, which were most critical limitations with the previous model release.

The following chapter 3 focuses on the basis of scalable device modeling. The fundamental concept of a fully geometry scalable model library for HBT transistors is explained. Thereafter the scaling laws, as implemented in the library, are presented in detail for each section of the device architecture. At the same time this is the fundamental basis for the development of extraction routines, since these are specifically tailored to the implemented scaling equations. In addition, the last section of this chapter details the basic set of test structures, required to determine the respective unitary and scalable model parameters from on-wafer measurements.

The last chapter focuses on the procedures to reliably determine parameter values for the scalable model library, starting with the external elements of the equivalent circuit. Afterwards the model parameter extractions strategies for the description of the intrinsic device are detailed for low and medium current range followed by high current parameters with a close relation to RF behaviour of the device. Since extraction for a scalable model library relies on a high number of geometries, recommendations for device selection are given along with each group of parameters. Best practices and guidelines for parameter optimization will be given, if direct extraction methods are not existing. Furthermore new concepts and strategies, based on novel extraction structures and measurement techniques, are evaluated. The chapter will be concluded with a final model verification, showing the general agreement of the extracted model with the measured data for a recent process generation.

Chapter 2

Physics and Modeling of Bipolar Junction Transistors

The increasing amount of multimedia functions in communications and consumer markets drive the need for speed and functionality at low power consumption, while cost reasons are limiting die area. Generally bipolar transistors meet these demands by offering high speed and gain at low noise (properties for high-frequency analog amplifiers and circuits). However CMOS technology is best suited for digital applications such as low-power logic gates.

Thus a desirable configuration would be the bipolar junction transistor (BJT) as amplifier with low noise on the one hand and the complementary metal oxide semiconductor (CMOS) dedicated to digital applications on the other hand realized in one integrated circuit. Common application examples include but are not limited to transmit/receive (TR) interfaces to radio antenna of mobile telemetry devices as well as modern radar systems. Yet such applications are demanding mass availability of fast devices at low cost.

For high density digital circuits planar transistor scaling (usually involving an advancement of lithographic nodes) has been one of the means to increase package density and reduce power consumption, yet new materials, new fabrication technologies as well as architectural changes allow performance improvements as well. Even though shrinking is a common technique to increase functionality of digital desgins, analog circuitry does not benefit from process node advancements in the same way as digital circuits.

In contrast to lateral transport devices such as classical MOS-FETs, the device speed is not directly linked to the lithographic dimensions of the process. Especially for bipolar transistors simple lateral scaling of horizontal device dimensions is hence not the only way of process advancements. Being a device with vertical current flow the device profile is equally important [14]. Thus a good balance is required between profile optimization for transit time reduction (τ) as well as scaling of lateral dimensions to keep capacitances low.

Furthermore analog circuits usually incorporate a large number of passive components (capacitances, inductors, resistances), that do not directly scale with lithography. Additionally more advanced technology nodes are more expensive in manufacturing, thus it is often-times favourable to realize circuits in larger manufacturing processes. SiGe BiCMOS technology hence is in a sweet-spot for mm-wave applications due to the combination of high RF performance with low cost, integrability and quality of passives.

Todays SiGe technologies offer considerable performance advantages for mixed-signal applications. Given the demand of mass market applications, especially the combination of SiGe technology with the wide availability of CMOS possesses, increased the development speed in HBT technologies . With BiCMOS performance typically being two process nodes ahead of RF CMOS [15–17], devices reaching speeds comparable or exceeding those provided by RF CMOS at smaller process nodes became available. Technology optimizations for low power, low or high voltage and high frequency provide plenty of options to mixed signal device designers.

In order to make the transition to physics-based device modeling for todays transistors, a good understanding of working principles and effects is required. In this chapter a brief overview of the physical operation principles of SiGe based HBT devices is given to build a solid understanding of the constraints encountered in modeling and to better understand presented solution approaches. The fundamental operation of SiGe bipolar transistors is elaborated, whereafter the most critical issues for modern device modeling and the most important model equations as well as latest model improvements are presented. In addition to the physical equations, the encountered phenomena are reproduced in numerical device simulations and simulation, studies using realistic device profiles.

2.1 Fundamentals of Bipolar Transistor Device Physics

2.1.1 Basics of Silicon npn BJT Operation

The silicon bipolar junction transistor (BJT) consists of three adjacent semiconductor regions: the emitter region (n-type), the base region (p-type), and the collector region (n-type). Such a transistor is called an npn transistor in short notation and the device terminals are labelled Emitter (E), base (B), and collector (C).

The term transistor action refers to the control of a large current (collector-emitter current) by a smaller (base-emitter back injection) current in the forward active operation. The so called 'forward' device operation mode refers to the base-emitter (BE) junction which is forward biased ($V_{BE} > 0$). For npn devices, the base region, being doped with p-type material, has an abundance of free holes (h^+). With an external bias applied, electrons are injected from the (n-type) emitter into the base. The holes reaching the emitter will recombine there with free electrons (e^-). The emitter, being n-type material, has an abundance of electrons. The same electrical field forces holes from the base to the emitter. The aim of a transistor is to have only few holes injected from the base into the emitter. Therefore in a classical transistor design the emitter has a much higher dopant level (N_E) compared to the base impurity concentration (N_B).

There are different modes of device operation depending on bias conditions of both junctions. The main states used in practical circuit design are cut-off, active-linear, and saturation.

The first differentiator is the bias of the BE junction. When the BE junction is reverse-biased, the are no charge carriers that enter the base and move to the collector. Hence a voltage applied between collector and emitter has no significant effect. This region is called the cut-off region. When the BE junction is forward-biased, the transistor is active. The so called *forward active* or *active-linear* mode is generally the intended operation leveraging the transistor as amplifier. The behaviour of the transistor, then depends on the voltage applied between collector and emitter (CE) junction:

- Emitter-Base junction forward biased ($V_{BE} > 0$)
- Base-Collector reverse biased ($V_{BC} < 0$)

The so called *saturation* mode is defined by the voltage between the collector and emitter being less than the forward bias voltage of the base-emitter junction ($V_{CE} < V_{BE}$ or $V_{BC} > 0$ respectively). A large amount of minority carrier charge (e^-) is accumulated in the base region. As the transistor switches from saturation to cut-off, this charge initially remains in the base and a collector current will remain until this charge is removed by recombination. This mode is characterized by:

- Both junctions forward biased ($V_{BE} > 0$ and $V_{BC} > 0$)

When the BE junction is reverse-biased, the are no charge carriers that enter the base and move to the collector. Hence a voltage applied between collector and emitter has no significant effect. A

third mode thus is the *reverse active* operation where function of the emitter and the collector are reversed resulting in very poor efficiency. This mode is characterized by:

- Emitter-Base reverse biased ($V_{BE} < 0$)
- Base-Collector forward biased ($V_{BC} > 0$)

The operating regions can be sketched in the four quadrants of the respective terminal bias as shown in Fig. 2.1. In the output characteristic (cf. Fig. 2.2) the collector current I_C measured for varying V_{CE} shows the operating modes as well.

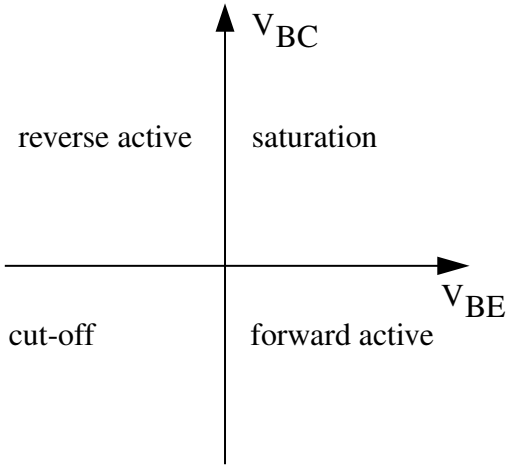


Figure 2.1: Mode of operation determined by the bias condition of the respective quadrant

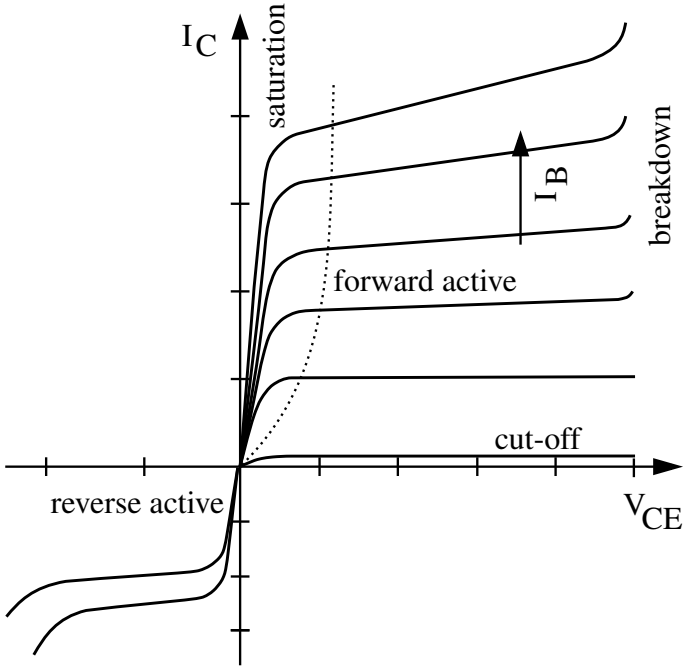


Figure 2.2: Output characteristic with collector current I_C as a function of V_{CE} bias; *dashed line: separation of saturation and active region*

Focusing on the most common use of a BJT as an amplifier, Fig. 2.3 shows the carriers contributing to the current flow. The positive voltage V_{BE} causes the p-type base to be higher in potential than the n-type emitter, thus forward-biasing the emitter-base junction. Electrons are injected from the (n-type) emitter into the base. Likewise, holes from the base are injected into the emitter.

Electrons injected into the base diffuse away from the emitter-base junction towards the base-collector junction. As they move through the base, some of the electrons encounter holes and recombine with them. However the majority of electrons get to the base-collector junction where they encounter a strong force due to the electric field associated with the reverse bias ($V_{BC} < 0$), sweeping these electrons out of the base and into the collector. Given that the base is p-type with available free holes, charge neutrality is maintained.

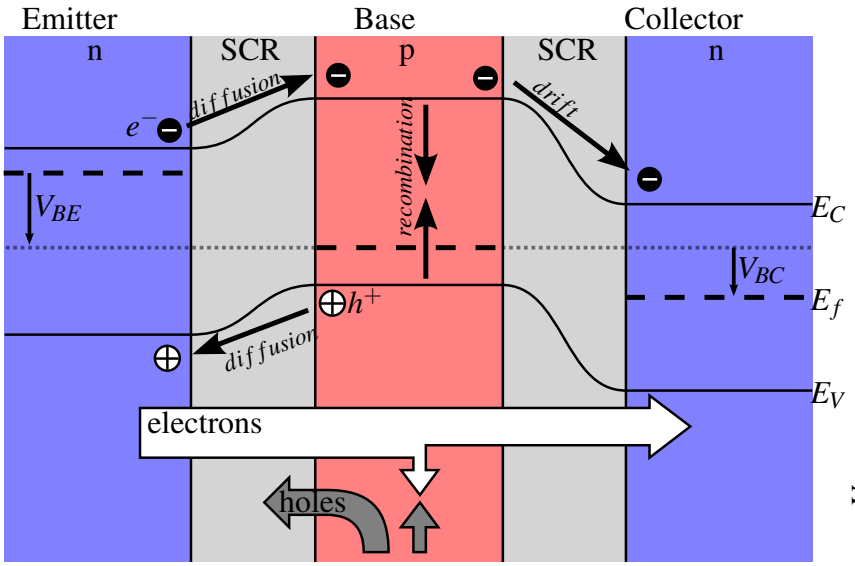


Figure 2.3: Carrier components contributing to current flow

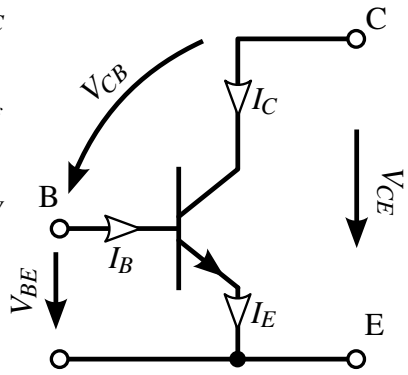


Figure 2.4: Circuit symbol

Current flow in the BE junction is consisting of two components: electrons (e^-) injected from the emitter into the base, and holes (h^+) injected from the base into the emitter. The current that has to be supplied to the base contact comes from two main sources: Recombination in the base and injection of minority carriers into the emitter (reverse injection of holes into the emitter). In order to achieve highest amplification this current is to be kept at a very low level. Hence the device design is intended to favor a high density of electrons in the emitter and a low density of holes in the base.

Electrons are majority carriers in the emitter region (n-type) where they have a high mobility (μ_n). As the electrons pass through the base region (p-type), they are minority carriers and have a relatively low mobility (μ_p). Upon arrival in the collector they are majority carriers again with a high mobility again (n-type).

For representation in equivalent circuit diagrams the symbol of the npn is presented in Fig. 2.4. The direction of I_E is defined out of the emitter terminal (in the direction of the hole current and opposite to the direction of the electron current) with both currents I_B and I_C into the respective terminal.

2.1.2 The SiGe HBT Transistor

The dilemma of classical BJT design is given by the base doping: for high frequency operation a low base resistance is desired, usually achieved by high base doping. However in order to reach a high current gain, the hole injection into the emitter has to be minimized, calling for a low base doping. One solution to this trade-off is the use of band-gap engineering in the Si_{1-x} Ge_x system, allowing for higher base doping at constant gain.

The *H*eterostructure *B*ipolar *T*ransistor (HBT) differs from the classical *Si*licon (Si) based bipolar transistor (exclusively using Si homojunctions) in a way that one (or both) junctions are formed between dissimilar semiconductor materials.

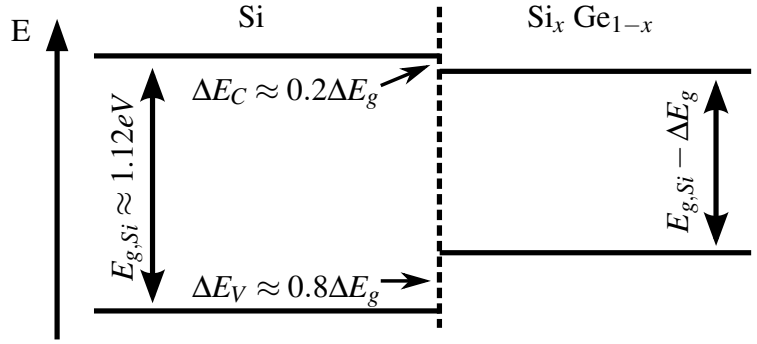


Figure 2.5: Difference of band-gap E_g due to Ge incorporation into silicon; *Example of 20%Ge incorporation, most of the band-gap difference (ΔE_g) seen in the valence band edge (E_V)* [18]

The Si_xGe_{1-x} (with x being the Si fraction) alloy has a smaller band-gap between electron and valence band than that of Si ($E_{g,Ge} = 0.66eV$ compared to $E_{g,Si} = 1.12eV$ at 300K) due to its larger lattice constant. This largely influences the usefulness for transistor engineering. The band offset compared to pure Si is predominantly in the valence band (cf. Fig. 2.5). The goal of HBT technology is to leverage the advantageous material effects due to the band-gap offset. It is a common practice to use a narrower band-gap in the base compared to the emitter. A wider emitter band gap compared to the lower in the base results in a lowered barrier encountered by holes injected to the emitter compared to the barrier for electrons injected into the base. This reduces base current and increases emitter injection efficiency with the desired effect of higher current amplification [19].

Without further derivation three simplified approximations give the key advantages of SiGe technology over a comparable Si BJT. Using the amount of incorporated band-gap difference ($\Delta E_{g,SiGe}$) those can be summarized in three important normalized Figures of Merit (gain β, Early voltage V_A and transit time τ). The exponential dependence on the band-offset can be directly related to the intrinsic carrier density (n_i, cf. eqn. (E.5)).

$$\left.\begin{aligned}
\frac{\beta_{SiGe}}{\beta_{Si}} &= \frac{\Delta E_{g,SiGe}/k_BT}{1-\exp(-\Delta E_{g,SiGe}/k_BT)} \\
\frac{V_{A,SiGe}}{V_{A,Si}} &= \frac{k_BT}{\Delta E_{g,SiGe}} \cdot [\exp(-\Delta E_{g,SiGe}/k_BT) - 1] \\
\frac{\tau_{B,SiGe}}{\tau_{B,Si}} &= \frac{2k_BT}{\Delta E_{g,SiGe}} \cdot \left[1 - \frac{k_BT}{\Delta E_{g,SiGe}}(1-\exp(-\Delta E_{g,SiGe}/k_BT))\right]
\end{aligned}\right\} \text{improvements through Ge [20]} \tag{2.1}$$

A design goal for bipolar transistors is to have almost all of the current across the base-emitter junction consisting of electrons being injected into the base so that very little emitter current is made up of holes coming from the base into the emitter. Transistors are therefore specifically designed so that almost all of the electrons, which are injected into the base, diffuse to the reverse-biased base-collector junction (yet recombination in the narrow base region is unavoidable).The basic idea of band-gap engineering thus is to provide an additional energy barrier ΔE_V to holes injected from the base to the emitter and a supporting field to the electrons traversing the base. This in turn allows the realization of a required current gain (β) while maintaining a high base doping (N_B). The higher doping level gives a reduced base resistance and therefore improved high-frequency characteristics desired for RF circuits.

A thin trapezoidal shape of the Ge profile is desired due to the combination of a low base transit time τ_B (through the drift field) and a significant Ge mole fraction at the BE junction giving a low τ_E together with low base resistance r_{sBi} which is desired to obtain best f_{max} values. A graded Ge profile is therefore considered as optimum solution to achieve all these characteristics. Ideally Ge profile whose fraction is progressively raised from the emitter side towards the collector results in a band-gap that progressively reduces towards the collector junction. The high hole conductivity in the base precludes a significant change in the valence band (E_V) and hence the majority of the band-gap narrowing occurs in the conduction band (E_C) as seen in Fig. 2.6.

There are some advantages linked to this approach: At identical base-emitter voltage, a higher collector current (higher β) and smaller transit time (τ_f) is achieved resulting in higher cut-off frequency. Given sufficient gain (β) the transistor may hence be further be optimized for smaller base resistance, decreasing transistor input power consumption (P_{in}) whilst higher overall power gain boasts the maximum oscillation frequency (f_{max}). In addition a higher current gain combined with low base resistance help maintaining a high signal to noise ratio (SNR) when used as amplifier resulting in a low noise figure (NF_{min}).

Various studies of the effect of graded Ge profiles on HBT device performance were carried out in the 90's and the early 21st century [21–24]. Since this time however SiGe technology has seen an enormous development in both scaling areas: lateral dimension as well as vertical device profile. Recent research work demonstrated the feasibility of SiGe bipolar transistors with a maximum operating frequency of 0.5 THz (e.g. dotfive [25]).

Compared to todays industry leading technologies the analyzed HBTs were fabricated within technologies capable of less than 100GHz cutoff frequencies[1]. Compared to these technologies todays HBT device designs use higher molar Ge fractions (reaching about 30%), while the base layer width (w_B) decreased all the way to a critical thickness of less than fifty nanometer.

[1]HBT technology generations are commonly distinguished by their *AC* performance, where the peak common-emitter unity gain cutoff frequency ($f_{T,peak}$) is used to reflect the degree of sophistication in device structural design

The advantage of the compositionally graded Ge profile is visible in Fig. 2.7. As illustrated in the comparison of a graded compared to a box-shaped profile with abrupt Ge-content change, the trapezoidal (or triangular as simplified case) profile results in the reduction of unsteadiness in the band diagram leading to less barriers for the carriers incorporated in the device operation.

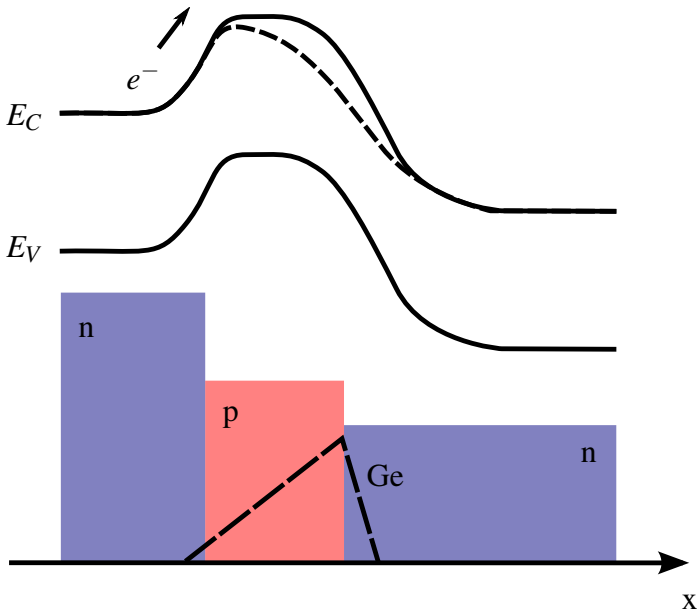


Figure 2.6: Schematic view of compositional graded Ge profile in the base of a npn HBT; *solid lines indicate the band diagram of a npn bipolar transistor without Ge; the dashed line corresponds to a trapezoidal Ge profile in the base*

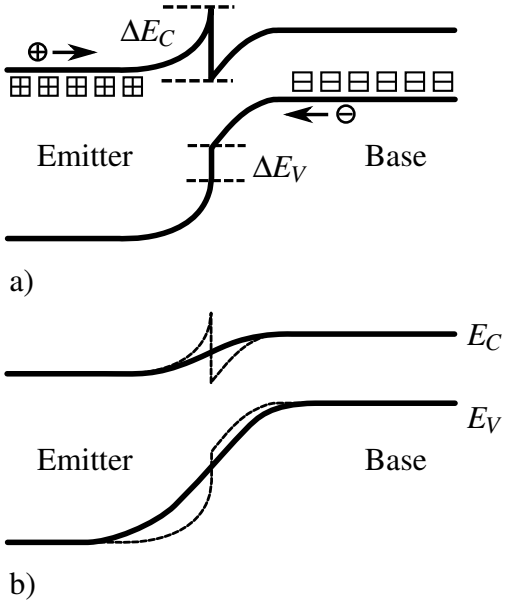


Figure 2.7: Band diagram comparison for BE junction with: a) abrupt Ge profile (Box profile); b) compositional graded Ge profile with trapezoidal shape (dashed: abrupt Ge profile)

A schematic view of the ideal trapezoidal Ge profile incorporated in recent SiGe technologies is shown in Fig. 2.6. The resulting energy band diagram is drawn with the Ge profile in the base. The continuously increasing germanium fraction predominantly lowers the conduction-band edge towards the end of the base. The resulting additional drift field in the neutral base helps carrier transport through the neutral base supporting their diffusion to the collector sided SCR.

In practice a combination of box and linearly graded profile is employed (cf. Tab. App. B.1) in a way that discrete steps of Ge fractions are employed in the base layer, whilst profile-smoothing (smear-out of the Ge profile) is achieved through following annealing steps.

However the undesired heterojunction barrier effect (HBE) at the BC junction (tailing edge of Ge profile) may retard carrier transport and thus cause performance degradation once the device is operated in the high injection regime. Proper device design and modeling of these effects is hence of high importance as vertical profiles are narrowed. To study the influence of the Ge profile on SiGe HBTs and to determine the optimum vertical device doping and Ge profiles, numerical device simulation are the best link to the physics of the HBE (cf. sect. 2.5.2 and Fig. 2.35).

2.2 A Physics-Based Compact Model for BJTs

In early days of bipolar technology there was the classical Ebers-Moll model, a highly idealized model for a bipolar transistors, which was intended to be used in the active mode of operation (forward and reverse). It is based on the the diode current equations (cf. section 2.2.1) and the predecessor of todays computer simulation models. Originally intended for pure silicon BJT these days the models are used for simulation and design of SiGe HBT based circuits as well.

The widespread VBIC (Vertical Bipolar Inter-Company) model introduced in 1995 can be considered as the direct extension of the classic SGP model. The HIgh CUrrent Model (HICUM) as well as MEXTRAM (Most EXquisite TRAnsistor Model) were developed independently in conjunction with bipolar process advancement.

Todays models have been continuously improved, but these days there are hundreds of parameters making correct fitting of device characteristics an increasingly difficult task for model engineers. The right selection of an appropriate model[1] for an intended purpose is hence important. Especially the effects encountered in the SiGe material system require consideration of more complex device phenomena (cf. Section 2.1.2). In particular those modeling issues specific to SiGe HBT devices include a different effective band-gap encountered in the transistor regions and the different temperature dependence. Understanding the physical basis behind the different effects is hence advantageous.

In order to make the system of equations easier and derive computationally efficient algorithms, simplifying the governing equations of device physics is a popular method in device modeling since rational simplifications (while maintaining a physical background) do not necessarily mean a loose of accuracy. A brief summary of essential equations is hence necessary for a deeper understanding of the formulations used and implemented in the device model. Therefore the fundamental set of equations governing the operation of semiconductor devices shall be presented without further derivation of the origins for each relation or definition [26].

Carriers within a semiconductor crystal move as if they were free particles which are unaffected by atoms in the material except for an effective change of the mass of the particle. The carriers in semiconductors can either be electrons (n) or holes (p) which carry a negative or positive unit of charge. Even with no electric field applied these carriers move due to the thermal energy they have at a given ambient temperature.

The motion (drift) of a carrier due to an electric field E is described by a drift current component, due primarily to the majority carrier in an extrinsic semiconductor. During random motion of carriers in a semiconductor (with or without presence of electric field) these constantly change direction and velocity due to scattering.

[1] todays most widespread compact models for bipolar devices available in common commercial simulators are SGPM, HICUM, MEXTRAM or VBIC

With an electric field applied the carriers on average have a net motion along the direction of the field. We define the electric field through the negative gradient of the electrostatic potential ψ

$$E = -\frac{\partial \psi}{\partial x}. \tag{2.2}$$

In this case velocity is proportional to the applied field and a mobility (μ) can be defined as proportionality ratio of velocity (v) to field ($\mu = v/E$)[1]. In consequence a drift current as a backbone of the drift-diffusion model (described in the Appendix, cf. App. I), can then be expressed as a function of mobility, carrier concentration as well as applied field, yielding

$$J_n = qn(x)\mu_n E. \tag{2.3}$$

Characteristically, under thermal equilibrium (without currents in the semiconductor) the Fermi-edge E_f is constant versus the position coordinate (x) throughout the energy-band model and there is only one energy level E_f for electrons and holes. In case of an unbalanced state (e.g. with current flow due to supplied voltages) the Fermi-level is split-up into a quasi-Fermi-level for electrons ($E_{F,n}$ or φ_n) and another one for holes ($E_{F,p}$ or φ_p).

2.2.1 Classical Bipolar Models

Compact models have to be as simple as possible yet meet a number of stringent requirements. Hence there is a long history of compact model development and advancement as device technology brought new challenges as well as opportunities.

Shortly after Bardeen, Brittain and Shockley invented their point-contact transistor in 1947 the first detailed analysis of the p-n junction and the associated derivation of the collector current equation within a silicon based bipolar transistor was done by Shockley in 1948 [27] leading to the Shockley diode equation in the well known form

$$I = I_S \left(\exp\left[\frac{V}{mV_T} \right] - 1 \right), \tag{2.4}$$

where the diode current I is related the applied junction voltage V_D through I_S, the saturation or scale current of the diode (representing the current that flows for negative V_D). The denominator of the exponential factor being V_T, the thermal voltage ($k_B T/q$) and m, the diode ideality factor (or sometimes referred to as emission coefficient). The classical solution by Shockley assumes a constant base doping profile.

In 1954 Moll and Ebers developed a simple composite model that was intended to be usable

[1] The mobility of electrons and holes in silicon is different and decreases with impurity concentration due to scattering with the ionized doping atoms

to predict the operation of a BJT in all of its possible modes [28]. This large-signal model is mathematical model of transistor currents based on the use of only three essential parameters with saturation currents (I_S) and amplification factors (A) for normal (index N) and inverse (index I) operation. The equivalent circuit is composed of two diodes each in parallel with a current-controlled current source (cf. Fig. 2.8). The calculation of currents is performed as follows:

$$I_C = A_N I_{S,N} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) - I_{S,I} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) \tag{2.5}$$

$$I_E = -I_{S,N} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + A_I I_{S,I} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) \tag{2.6}$$

$$I_B = (1 - A_N) I_{S,N} \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + (1 - A_I) I_{S,I} \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) \tag{2.7}$$

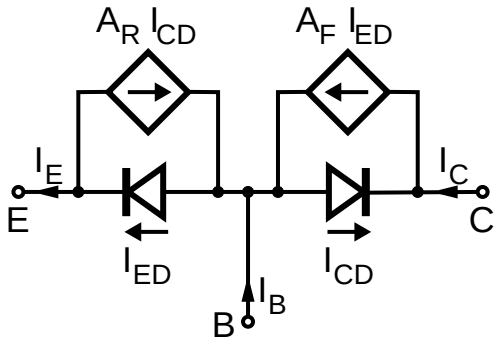


Figure 2.8: Equivalent circuit of simple Ebers-Moll diode model for npn BJT

More sophisticated versions of this basic model include series resistances in order to account for apparent voltage drops yet due to numerous simplifications the model accuracy and application to real devices is very limited. However for some steps in the model parameter extraction flow as well as for simplification these equations are still used in different forms.

2.2.2 The Origin of the Integral Charge Control Relation (ICCR)

In order to model devices adequately at high frequencies, there is a need to account for the charge (Q) stored in the device. Stored charge can be modeled as a capacitance (C) in small signal models. Proportionality between current (I) and charge is given by time (τ). The proper physical interpretation is that this is the transit time for carriers to cross the respective semiconductor region.

$$\tau = \frac{dQ}{dI} \text{ and } Q = \int_0^{I_T} \tau di \tag{2.8}$$

The fundamentals of the Integral Charge Control Relation (ICCR) were presented by Gummel

and Poon in 1970 [29]. Gummel and Poon simplified the charge control approach to a form, where only the inner base region is taken into account. This leads to the well known formulation of the transfer current (I_T) within the bipolar transistor that is controlled directly by the minority charge. Since it presents a fundamental backbone of bipolar modeling the basic theory of the approach behind the ICCR shall be explained in brief [30], [31].

The derivation of the classical ICCR starts with the transport equation of electrons and holes respectively under low injection condition. Low injection is defined as operation when the number of carriers generated are small compared to the majority carriers of the material ($p(x) \ll N_D(x)$). For the electron and hole currents this yields

$$J_n = -qn(x)\mu_n E(x) + qD_n \frac{dn}{dx} = -qn(x)\mu_n \frac{d\varphi_n(x)}{dx} \quad \text{and} \tag{2.9}$$

$$J_p = qp(x)\mu_p E(x) + qD_p \frac{dp}{dx} = qp(x)\mu_p \frac{d\varphi_p(x)}{dx}. \tag{2.10}$$

The carrier concentrations are considered as dependent on position (x) and expressed using the electrostatic potential Ψ and the (position dependent) quasi-Fermi potential $\varphi(x)$

$$n(x) = n_i \exp\left(\frac{\Psi - \varphi_n(x)}{V_T}\right) \tag{2.11}$$

$$p(x) = n_i \exp\left(\frac{\varphi_p(x) - \Psi}{V_T}\right) \tag{2.12}$$

For electrons the substitution of

$$\exp\left(\frac{-\varphi_n}{V_T}\right)\frac{d\varphi_n}{dx} = -V_T \frac{d}{dx}\exp\left(-\frac{\varphi_n}{V_T}\right) \tag{2.13}$$

yields the following differential equation

$$\frac{d}{dx}\exp\left(\frac{-\varphi_n}{V_T}\right) = \frac{J_n \exp\left(\frac{-\Psi}{V_T}\right)}{qV_T n_i \mu_n}. \tag{2.14}$$

With the minority carrier boundary conditions at the depletion edges of the transistor (x_e: the emitter contact and x_c: the collector contact) one obtains

$$\int_{\exp(-\varphi_n(x_e)/V_T)}^{\exp(-\varphi_n(x_c)/V_T)} d\left[\exp\left(\frac{-\varphi_n}{V_T}\right)\right] = \int_{x_e}^{x_c} \frac{J_n \exp\left(\frac{-\Psi}{V_T}\right)}{qV_T n_i \mu_n} dx \tag{2.15}$$

For efficient transistor operation the electron injection component (J_n) is much larger than the hole (back) injection component (J_p). The hole current component hence is assumed to represent

only a small portion when recombination is assumed to be negligible. Hence the hole quasi-Fermi potential can be assumed to be constant throughout the vertical device profile ($\varphi_p = \text{const.}$) simplifying the integration

$$\int_{\exp(\varphi_p-\varphi_n(x_e)/V_T)}^{\exp(\varphi_p-\varphi_n(x_c)/V_T)} d\left[\exp\left(\frac{\varphi_p-\varphi_n}{V_T}\right)\right] = \int_{x_e}^{x_c} \frac{J_n \exp\left(\frac{\varphi_p-\psi}{V_T}\right)}{qV_T n_i \mu_n} dx. \tag{2.16}$$

With the difference between the quasi-Fermi potentials being the junction bias of BE and BC depletion region respectively one can simplify the lhs of the differential equation using the respective voltages (V_{BE} and V_{BC})

$$\begin{aligned}\int_{\exp(\varphi_p-\varphi_n(x_e)/V_T)}^{\exp(\varphi_p-\varphi_n(x_c)/V_T)} d\left[\exp\left(\frac{\varphi_p-\varphi_n}{V_T}\right)\right] &= \exp(\varphi_p-\varphi_n(x_c)/V_T) - \exp(\varphi_p-\varphi_n(x_e)/V_T) \\ &= \exp\left(\frac{V_{BC}}{V_T}\right) - \exp\left(\frac{V_{BE}}{V_T}\right).\end{aligned} \tag{2.17}$$

With no recombination in the structure the current flowing through the device is assumed to be constant, allowing to simplify the rhs. and resulting in the form

$$J_C = \frac{qV_T}{\int_{x_e}^{x_c} \frac{p(x)}{n_i^2\mu_n} dx}\left[\exp\left(\frac{V_{BC}}{V_T}\right) - \exp\left(\frac{V_{BE}}{V_T}\right)\right], \tag{2.18}$$

with the integral in the denominator being referred to as the Gummel number or the total hole charge Q_p respectively

$$Q_p = \int_{x_e}^{x_c} \frac{p(x)}{n_i^2\mu_n} dx \tag{2.19}$$

Models based on this relation are referred to as ICCR models, whereas the treatment of the base charge (integral of weighted hole charge $\int p(x)dx$ in the denominator) is the key differentiating factors between the various variants and compact model implementations respectively.

For device modeling the most important charge contribution in a BJT is given by the minority carrier charge q_p. This charge, being the denominator of the transfer current model equation, consists of two components: the movement of charges of the base-emitter and base-collector depletion region being the first and the injected minority carriers into the base from the emitter being the second (injection from the collector for reverse operation respectively). The former are assumed to be proportional to the change in the normalized depletion charges, the latter being proportional to the transfer current (I_T). The excess minority charges related to the current injection are modeled as a transit time (τ) multiplied by the junction current giving the common implementation of the base charge in the SGPM

$$Q_{pB} = Q_{p0} + Q_{jEi} + Q_{jCi} + \tau_{Bf} \cdot I_{Tf} + \tau_{Br} \cdot I_{Tr} \tag{2.20}$$

In this equation Q_{p0} is defined through charges defined as a combination of diffusion capacitance and space charge related hole charge. Q_{jEi} and Q_{jCi} are the BE and BC depletion charge obtained from the model of junction capacitance and minority charges with their respective current contributions I_{Tf} and I_{Tr}. Subscript indices $_0$ indicate zero-bias (thermal equilibrium) values for charges of the respective junctions and τ_B is the base transit time.

With a certain loss in physical relation the SGP model representation is then obtained by means of approximation through the model parameters of forward and reverse Early voltage (V_{AF} and V_{AR}) for the respective junctions

$$I_C \propto \frac{1}{Q_B}, \text{ with } Q_B \approx \frac{1}{1 - \frac{V_{BE}}{V_{AR}} + \frac{V_{BC}}{V_{AF}}} \tag{2.21}$$

In addition to this high current degradation parameters in form of knee currents (I_K) are used. The respective model parameters are then determined from measurement data for description of the respective operating regime. It has to be noted that benefit (and weakness at a time) of the model is the de-correlation of DC and capacitance description allowing for a high degree of flexibility. However several critical effects are not taken into account:

- At high current the number of electrons injected into the base reaches the background doping concentration causing the number of holes increase in order to maintain space charge neutrality. This in consequence will cause the gain to drop and was first described by Webster [32]
- In the same way the electron concentration at high current reaches values in excess of the doping concentration in the collector, causing the base push out effect which is significantly increasing the base width w_B whereby the base transit time increases (f_T fall-off first described by Kirk [33])
- So called quasi-saturation appears once the parasitic collector resistance at high currents causes the internal collector node to significantly change potential, eventually forward biasing the base-collector junction [34]

2.2.3 The Generalized Integral Charge Control Relation (GICCR) for Modeling of Heterojunction Transistors

After the presentation and application of the ICCR in SGPM many researchers and engineers retrieved the advantage of the charge-control concept and continuously improved the model formulations. The ICCR approach has therefore been extended with a generalized approach [35, 36]. To make the model applicable for HBT transistors the weighting factors (denoted h) of the space charges Q_{jEi} and Q_{jCi} have been introduced to calculate the total charge within the transistor by

means of eqn. (2.22). This allows for effective modeling of the Early effect. In later model updates to the widely used version L2.1x the model was made applicable to HBT devices as well by generalizing the ICCR and introducing weighting factors [h_{jEi} and h_{jCi}] that account for the highly increased charge dependence on the space charges found in a HBT. In the according publication [37] it is already mentioned that in general the weighting factors should be bias dependent. However at the time being the SiGe HBT process technologies available were found sufficiently described by a formulation without bias dependence

$$Q_{pT} = Q_{p0} + h_{jEi} \cdot Q_{jEi} + h_{jCi} \cdot Q_{jCi} + Q_{fT} + Q_{rT} \tag{2.22}$$

However being closely related to device physics, the different contributions in the model are each attributed to the different regions of the 1D transistor: the charges in the neutral regions are summarized in the forward minority charges within the transistor Q_{pT} which is composed from the base (Q_{fB}), emitter (Q_{fE}) and collector (Q_{fC}) portion as well as the minority charge at low current densities Q_{f0}. The minority charge for operation in reverse mode is denominated Q_{rT} and not dedicated to a special transistor region. At last the charges stored in the space charge regions of both BE as well as BC junction are denominated Q_{jEi} and Q_{jCi} respectively. The various charges directly related to a specific transistor region are shown in Fig. 2.9.

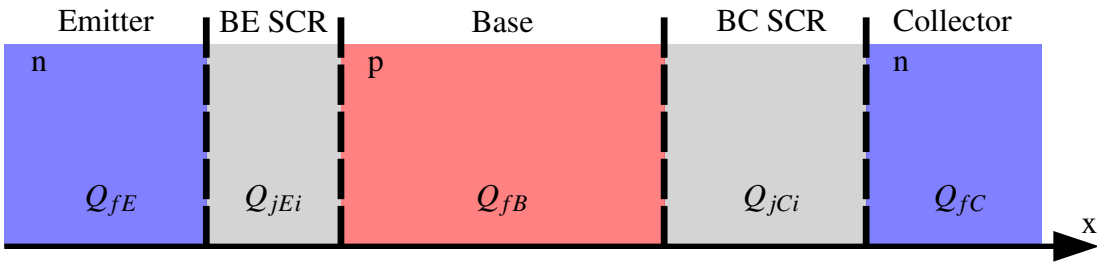


Figure 2.9: Charge contributions in the vertical transistor profile attributed to their respective zone

2.2.4 The HICUM Model

One widely used device model based on the GICCR is the *HI*gh *CU*rrent *M*odel (abbreviated HICUM). It is (together with VBIC and MEXTRAM) part of the ICCR based type of bipolar models originating from the fundamental base charge equations presented in the previous section.

The HICUM model development started in the 1980s with a focus on the high current operating region of a bipolar transistor [31, 38]. In the beginning of the model development the aim was the accurate description of fast bipolar devices used in ECL circuits. Yet the project was successfully developed further to extend the model, making it suitable for high speed applications in different advanced bipolar technologies such as widely used poly-silicon technology as well as the more

recent SiGe based HBTs. The original HICUM model presented in 1987 incorporated an improved but still simple formulation for the transfer current represented by the normalized hole charge (q_{pT}) found inside the 1D transistor and the GICCR constant (c_{10})

$$I_T = \frac{c_{10}}{Q_{pT}} \left[\exp\left(\frac{V_{BEi}}{V_T \cdot m_{CF}} \right) - \exp\left(\frac{V_{BCi}}{V_T \cdot m_{CR}} \right) \right], \text{ with} \tag{2.23}$$

$$Q_{pT} = A \cdot q \int p \, dx = Q_{p0} + Q_e + Q_c + Q_f \text{ and} \tag{2.24}$$

$$Q_{p0} = Q_p(V_{BEi} = V_{BCi} = 0). \tag{2.25}$$

Compared to the Gummel-Poon model, the HICUM model builds on the strengths of the integral charge-control relations making use of the advantage of linking large- and small-signal characteristics. In consequence the model is composed of an equivalent circuit with close physical relation. Thus each model parameter either represents electrical data, technological data, physical data, spatial dimensions or bias and temperature information as evidenced by the equivalent circuit (cf. Fig. 2.10). Likewise through strong relation to device physics a better understanding of the impact of certain effects on circuit characteristics is achieved.

The primary emphasis for the development of the HICUM model was (and still is) circuit design for high-speed/high-frequency applications. Being a semi-physical compact bipolar transistor model, HICUM is based on the extended and generalized Integral Charge Control Relation (ICCR) as presented in the previous section. In contrast to the Gummel-Poon model and its variants, HICUM applies consistently the ICCR concept without simplifications or fitting parameters. Quantities like depletion capacitances, transit times and charges are used to determine the dynamic transistor behavior and hence represent basic quantities of the model.

In HICUM special attention is paid to the contribution of the different charges within the transistor. The charge Q_{jEi} being the total charge between emitter and collector contact is obtained through integration using

$$Q_{jEi} = q \int_{x_B}^{x_{B,0}} N_B dx. \tag{2.26}$$

In this formulation x_B is the end of the BE SCR and depends therefore strongly on the bias of the base-emitter junction (reverse Early effect) that is strongly pronounced in recent HBT technology [39]. Recent work has dealt with the limitations of the previous HICUM model version to overcome issues with advanced bipolar processes [40, 41].

If not stated otherwise the compact transistor model used for any device and circuit simulation presented in this work is the HICUM model in its latest version L2.32[1]. A sound understanding

[1] at time of writing

of model fundamentals (including the equations in the model formulation) is necessary to create, examine and use most of the presented parameter extraction strategies. Those may be very different depending on the used compact model. However given the same principal physical origin of all advanced bipolar device models for some characteristics equivalent implementations are used as well. In consequence some of the parameter extraction strategies are common in between different models and sometimes applicable in the same way (e.g. junction capacitances, cf. section 4.5).

Any update of the model is either due to an improvement of physics-based fundamental model equations or on the other hand a change in implementation in order to improve runtime (speed) or convergence. Before the update to version L2.3x major incremental changes from version L2.1x to L2.2x were already taking into account several effects encountered with increasing device speed and narrow vertical profiles.

The large signal equivalent circuit of the HICUM compact model is shown in 2.10. Compared to the SPICE Gummel-Poon model (SGPM), the equivalent circuit of HICUM contains two additional circuit nodes. Namely these are B* and S'. The node B* separates the operating point dependent internal base resistance from the operating point independent external component and is required to take into account emitter periphery effects.

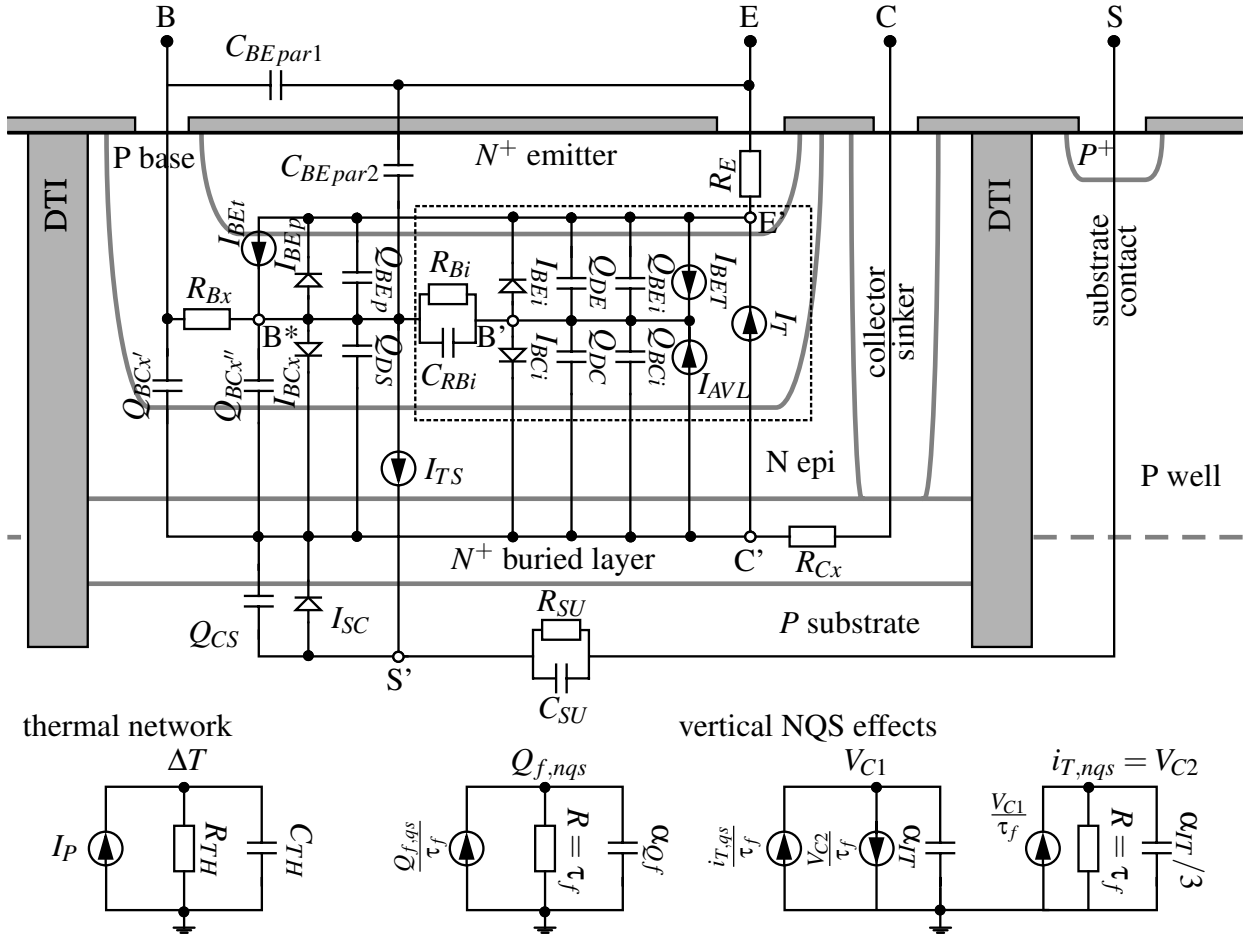


Figure 2.10: The HICUM model equivalent circuit of version L2.3x including additional thermal and NQS networks, *dashed line representing the intrinsic transistor*

The HICUM bipolar transistor model is very specific in a way that, being a physics-based model, it has a completely different transit time approach compared to the other popular SGP and MEXTRAM models. The formulation of charges storage, self-heating as well as the avalanche effect make advanced bipolar models far more applicable to todays advanced SiGe bipolar and BiCMOS technologies than conventional models.

The HICUM formulation links AC and DC components with each other allowing it to be valid under both DC and AC conditions over a wide range of frequency and bias. However advanced model formulations imply a more complicated set of numerical equations compared to the classical SGP model and with respect to model equations, equivalent-circuit and computational effort, and the convergence behavior might be less favorable.

In particular the transit time components for AC operation are influencing the charge Q_{pT} that on the other hand represents a major influence for DC-characteristics. In consequence the separation of AC and DC operation as known from other models (like SGP) is not given and the

interdependence of I_{TF} and Q_{pT} initiates a loop (cf. Fig. 2.11).

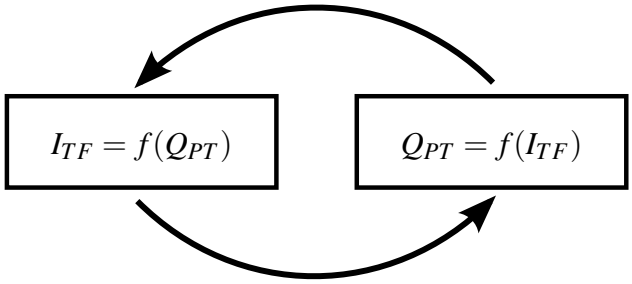


Figure 2.11: Interdependence of DC and AC characteristics in the HICUM model; *model parameters I_{TF} and Q_{pT} are functions of each other*

$$I_T = I_{TF} + I_{TR} \textit{ transfer current}$$

$$I_{TFL} = \frac{C_{10}}{Q_{pT}} \exp\left(\frac{V_{BE}}{M_{CF} \cdot V_T}\right) \textit{ low current}$$

$$Q_{pT} = Q_{p0} + \cdots + Q_{FT} \textit{ total hole charge}$$

$$\text{with } Q_{FT} = f(Q_{F0}),\ Q_{F0} = \tau_{F0} \cdot I_{TF}$$

approximation of interrelation

The consequence of this interrelation of both quantities is that the transfer current is not an explicit function of branch voltages and a iteration-loop (internal Newton solver) in the HICUM model code has to be solved for each simulation. Due to this complexity, simulations using the HICUM L2 model are generally slower than those employing other bipolar device models.

Basic Currents and Components in HICUM

Except for the transfer current I_T, all junction related current components are represented with a classical Shockley diode equation (cf. section 2.2.1). In this diode model equation, the saturation current (I_S) is exponentially dependent on the applied junction bias (V_j) and the thermal voltage (V_T). However non-ideal components arise from recombination in the depletion region of forward biased junctions (leakage currents). Therefore a non-ideality factor (m_j) is introduced. The general expression hence reads

$$I_j = I_{jS} \exp\left[\frac{V_j}{m_j \cdot V_T} - 1\right]. \tag{2.27}$$

Related model parameters for the respective junctions including their respective temperature coefficients as well as the interdependencies with the corresponding effective band-gap voltage are summarized in Tab. 2.1. For the base-emitter junction in particular, another node (B*) is defined in order to take into account peripheral effects (index p). This perimeter base node hence is different from the intrinsic transistor nodes B' and E' providing a partitioning option (cf. Fig. 2.10).

I	V	I_S	m	V_{geff}	ζ_T	component description
I_{BE}	$V_{B'E'}$	I_{BEiS}	M_{BEI}	V_{gEeff}	Z_{ETABET}	internal BE current
I_{BE}	$V_{B'E'}$	I_{REiS}	M_{REI}	V_{gBEeff}	-	BE recombination current
I_{BEp}	$V_{B*E'}$	I_{BEpS}	M_{BEP}	V_{gEeff}	Z_{ETABET}	quasi-static peripheral base current
I_{BEp}	$V_{B*E'}$	I_{REpS}	M_{REP}	V_{gBEeff}	-	peripheral BE recombination current
I_{BCi}	$V_{B'C'}$	I_{BCiS}	M_{BCi}	V_{gCeff}	Z_{ETACIT}	internal BC current
I_{BCx}	$V_{B*C'}$	I_{BCxS}	M_{BCx}	V_{gCeff}	Z_{ETACXT}	external BC current
I_{TS}	$V_{S'C'}$	I_{TSf}	M_{Sf}			fwd. parasitic substrate transistor
I_{TS}	$V_{S'C'}$	I_{TSr}	M_{Sr}			rev. parasitic substrate transistor
I_{SC}	$V_{S'C'}$	I_{SCS}	M_{SC}	V_{gSeff}	Z_{ETASCT}	substrate diode saturation current

Table 2.1: Currents and components in HICUM [*The factors used are denominated as follows: m: non-ideality factor; I_S: saturation current; V_{geff} band-gap description*; ζ_T exponent coefficient of temperature dependence; voltages are denoted according to the nodes as specified in the HICUM EC (cf. Fig. 2.10)]

As an example, the forward BE current I_{jBEi} is calculated using the following formula

$$I_{BEi} = I_{BEiS}\left[\exp\left(\frac{V_{B'E'}}{m_{BEi}\cdot V_T}\right) - 1\right]. \tag{2.28}$$

The temperature dependence is then taken into account by the general formulation

$$I_S(T) = I_S\left(\frac{T}{T_0}\right)^{Z_{ETA}} \exp\left[\frac{V_{geff}(0)}{V_T}\left(\frac{T}{T_0} - 1\right)\right], \tag{2.29}$$

where the parameter V_{geff} is modeled temperature-dependent itself.

Since often times identical manipulations of the model formulation behind a specific operating condition or bias range are employed for model parameter extraction, other model formulations will be presented in the respective section dedicated to the parameter extraction (c.f. chapter 3).

2.3 Effects in Advanced SiGe HBT Technologies

2.3.1 The Early Effect

The so called forward Early effect was first described by Early in [42]. The effect describes the variation of the neutral base width (w_B) with the variation of the BC junction bias. In a standard npn BJT technology the collector is fabricated with an impurity concentration that is some orders of magnitude lower than the doping concentration of the base ($N_B >> N_C$). Hence with a variation of the BC voltage the majority of the SCR variation happens within the collector. The base narrowing caused by bias variation has two consequences for the current. The neutral base width (w_B) is reduced, resulting in a reduction of recombination. The more significant effect however is the increased charge gradient across the base increasing the number of minority carriers injected across the emitter junction. In consequence at constant BE junction bias a change of V_{CE} increases the transfer-current (slope in the output characteristic).

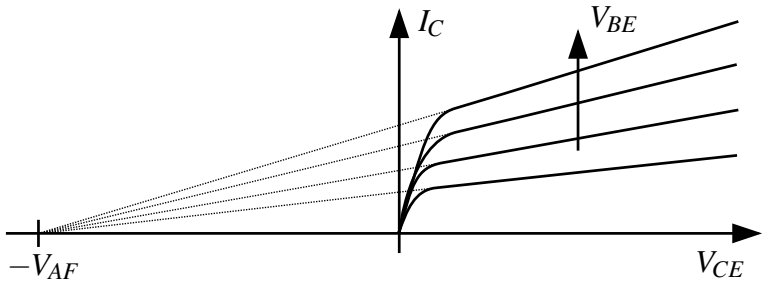


Figure 2.12: Output characteristic with forward Early effect and associated Early voltage V_{AF}

The reverse Early effect (impact of V_{BE} bias on w_B) became more pronounced in recent technologies. Contrary to the forward Early effect the reverse Early effect is linked to the BE junction of the BJT. In analogy to the base-collector junction the space charge region (SCR) of the BE junction varies with bias. Even though the variation of the BE SCR is less pronounced, the forward biased base-emitter junction influences the effective neutral base width (w_B) and thus the collector current as well. In pure silicon bipolar technology the Early effect was predominating and the reverse Early effect had less of an influence on the transfer current. However in recent technologies the incorporation of germanium in the base of SiGe HBTs increased the influence of the BE junction on the device characteristics.

Usually the doping level of the base is much lower than the doping concentration of the emitter. The variation of the depletion layer is hence predominantly in the base region (x_e). In analogy to the forward Early voltage the characteristic reverse Early voltage V_{AR} can be extracted from measured data. Since this effect is visible in DC measurement and directly related to the vertical profile, it can be regarded as process specific and rather independent of the lateral device geometry.

The reverse Early voltage V_{AR} is hence another measure allowing to easily evaluate and compare device performance. It can be derived from the forward Gummel characteristic at a BC junction bias of zero ($V_{BC} = 0V$). The collector current is normalized by the ideal diode current-voltage characteristic of the BE junction. V_{AR} can then be extracted from a linear regression on the normalized collector current ($I_{C,n}$) plotted as a function of the BE junction voltage V_{BE} (as shown in Fig. 2.13)

$$I_{C,n} = \frac{I_C}{I_S \cdot \exp\left(V_{BE}/V_T\right)} \approx 1 - \frac{V_{BE}}{V_{AR}}. \tag{2.30}$$

Shrinking vertical profile dimensions brought an apparent reduction of the stability of the current gain versus bias. This is most visible in the normalized collector current characteristics of different device generations manufactured by STMicroelectronics (cf. Fig. 2.14).

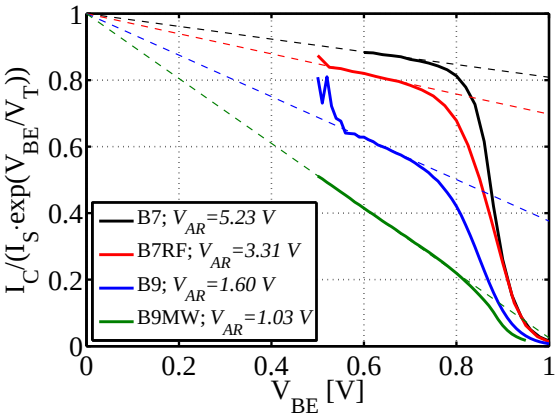


Figure 2.13: Normalized collector current $I_{C,n}$ for different BiCMOS technologies of STMicroelectronics; *solid line: measured data, dashed line: extraction of* V_{AR}

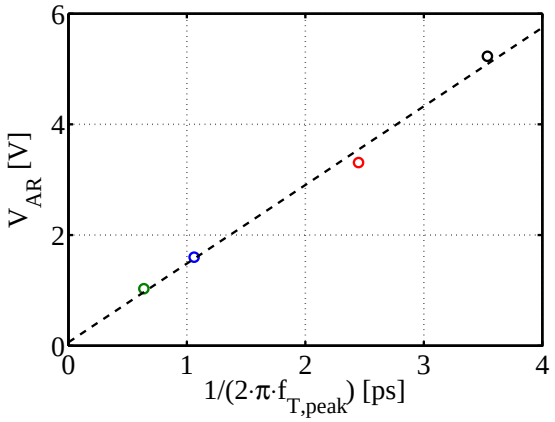


Figure 2.14: Extracted reverse Early voltage versus minimum transit time of different technology generations

As can be seen in Fig. 2.13, the reverse Early voltage (V_{AR}) was reduced significantly with the advancement of the SiGe technology. Setting the FoMs of the compared technology generations into relation the link can be made from the evolution of the transit frequency ($f_{T,peak}$) to the lowering of V_{AR} as shown in Fig. 2.14. It becomes evident that reverse Early voltage decreased linearly with the lowering of the transit time for the past generations of commercially available SiGe technologies manufactured by STMicroelectronics and HICUM in the former version L2.2x was not capable of reflecting this strong bias dependence. However the situation improved with the arrival of the most recent development of the BiCMOS55 device generation.

2.3.2 Observations From Process Splits

The non-ideality of the collector current detected in the BiCMOS9MW device generation raised the question of the physical origin of the effects evidenced by electrical characteristics. First a misplacement of E-B junction with respect to Si/SiGe hetero-junction was assumed. To further study this effect, different trials were made to analyze the impact of a change in the process route on the device characteristics. The manipulations in the process flow were done with the same mask set keeping lateral dimensions identical throughout the process splits. However the vertical profile was changed from the reference process. In a dedicated experimental lot a set of three wafers each were processed with one and the same process flow before the process flow was changed.

The focus was on the SiGe base epitaxial growth. The root process is based on a two-step Ge profile as shown in table 2.2. Hence the most important changes were the total amount of germanium in the transistor as well as the difference in concentration between both Ge plateaus (referred to as ΔGe). During the processing of the wafer the ideal profile, which is deposited (with its two discrete steps) becomes indistinct. The reason for this effect is natural as well as thermally activated out-diffusion. In consequence the ideal step profile becomes nearly trapezoidal after the full process flow (cf. Fig. 2.15).

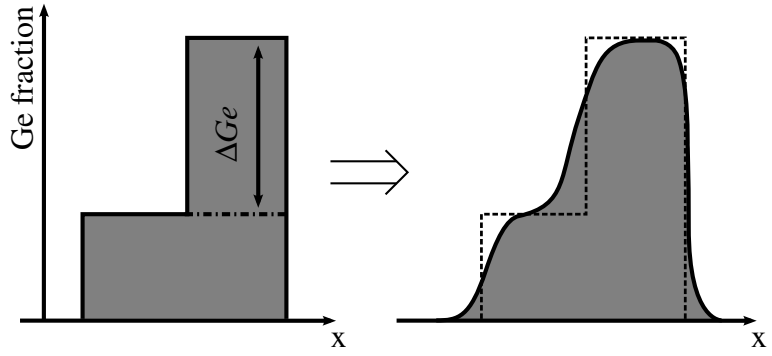


Figure 2.15: Germanium profile shape: *idealized deposited SiGe base composition (left) compared to the Ge profile after wafer processing and thermal annealing (right)*

The trials for variation of the SiGe profile are summarized in table 2.2. Opposed to the root process with a nominal Ge fraction of 10% and 25% respectively resulting in a slope of $\Delta Ge = 15\%$ was altered within a range of $\Delta Ge = 5\%$ through $\Delta Ge = 20\%$. Except for the split for $\Delta Ge = 5\%$ the total amount of Ge in the base was kept at the fixed value of $Ge_{max} = 25\%$ in order to allow a fair comparison of the device performance.

The measured electrical characteristics from the trials is shown in Fig. 2.16. The reverse Early voltage was extrapolated from the measured data at V_{BC0}. As becomes evident in the normalized collector current ($I_{C,n}$) curves the increase of the Ge profile steepness significantly decreases the reverse Early characteristics.

One can clearly see the impact of the steepness of the rising edge in the Ge profile between the first and the second plateau. For a relatively flat Ge profile with an increase of only 5% the reverse Early voltage has a value of more than 3 volt whereas the variation of the collector current with the V_{BE} bias is largely visible through a small V_{ER} value of only about one volt for an increase of 20%.

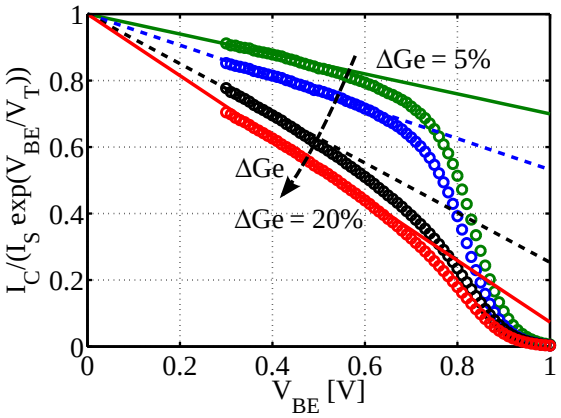


Figure 2.16: Normalized collector current $I_{C,n}$ for the different Ge profiles from process splits

wafer	Ge comp.	ΔGe	$V_{AR}/[V]$
w23	15%, 20%	5%	3.33
w24	15%, 25%	10%	2.25
w02	10%, 25%	15% (ref.)	1.34
w09	05%, 25%	20%	1.12

Table 2.2: Summary of changes in the Ge profile of the process split

2.4 Numerical Device Simulation

In contrast to compact modeling the domain of device modeling refers to the representation of device operation using detailed physical representation. Hence device modeling is usually incorporating TCAD (Technology Computer Aided Design) and helps supporting device and process modeling whereas compact modeling seeks to be predictive. The software solution used for all simulation trials is presented in [43]. A sample input file is found in the appendix (cf. App. L).

Since the reverse Early voltage was very low in the BiCMOS9MW technology generation and the effect of variation of the normalized collector current with BE bias was highly visible in this technology it was subject of detailed analysis. With the known increase of the Ge fraction incorporated with advance of device generations the worst-case scenario is represented by the BiCMOS9MW technology. The technology is based on a two-step Ge profile topping out at 25% and analyzed in detail by means of numerical device simulation. The background doping profile was obtained from SIMS measurement as well as calibrated process simulation conducted in a TCAD campaign. Even though the apparent band-gap narrowing ($\Delta E_{g,app}$) through high doping effects in the base is known to have an influence on device characteristics it is taken out of consideration in the subsequent analysis.

The trials were hence simplified for a better comparability in between the results. In real silicon processing the Ge fraction as well as the position has a high influence on the stability of the boron

profile of the base. The incorporation of different germanium profiles would alter the boron profile in the base yet this correlation was ignored. Instead the net doping profile as shown in Fig. 2.17 was kept unchanged for all variations of the Ge profiles.

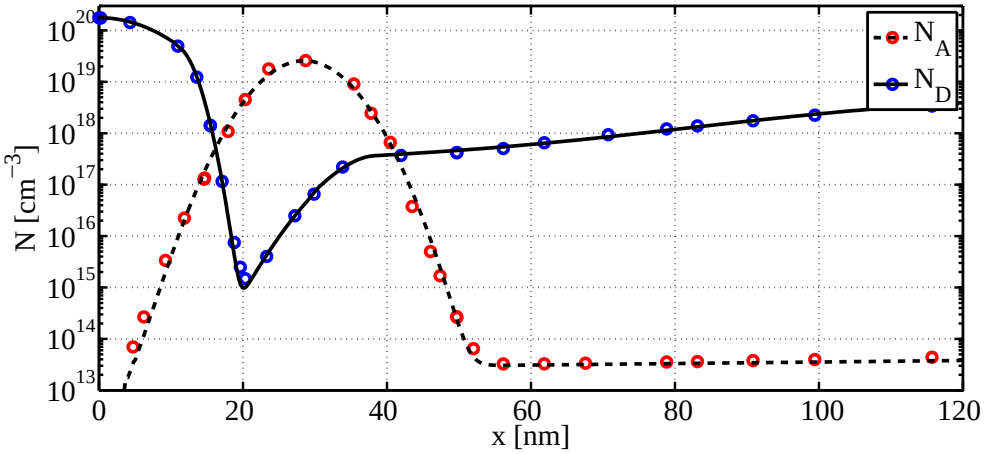


Figure 2.17: BiCMOS9MW vertical doping profile for device simulation [*discrete points: calibrated process simulation results, lines: doping profile for numerical device simulation*]

The Influence of the Ge Profile on Device Characteristics

To further investigate the effect seen in measurements of recent SiGe device generations and to determine how it is linked to the germanium profile, several different device simulation trials were performed. The results of the numerical simulations were compared to measured data in a first step.

As shown in Fig. 2.21 the Ge profile obtained as a result of technological process simulation was re-produced (modeled by the error function $erfc$) for vertical quasi-one-dimensional device simulation (denoted *fit*). Since the measured data is normalized by a saturation current a comparison with measured data is feasible (cf. Fig. 2.22). However it was evident that the initial 1D Ge profile does not allow to re-produce the electrical data obtained from the manufactured silicon. An *adapted* Ge profile was hence built by calibration to measured data, that allowed to re-simulate the measured data from the process.

2.4.1 Definition of the Boundaries in the Vertical Device Profile

The definition of the space charge region boundaries is adopted as described in [44] and calculated from the results of a small-signal analysis with variation of the terminal voltages in numerical device simulation as follows. δm is defined as the change of mobile carriers. For the base region of the HBT those minority carriers are holes (δp). δ here is indicating a deviation wrt terminal voltage or the response of the carriers to a terminal voltage change ($\delta = \partial/\partial V$). In a so called

regional analysis approach (regap) regional components defined by their corresponding minority and depletion charge are subdivided, defining the boundaries between a neutral region (NR) and a space-charge region (SCR).

In classical theory for abrupt profiles these boundaries are well defined whereas in real devices abrupt boundaries between NR and SCR are not existing. Another approach using the change of the space-charge density $d\rho = q(dp - dn)$ is hence required. The peaks to the left and right of the analyzed junction define the SCR boundaries (under low-current) as defined in [45].

Starting at the emitter the neutral emitter width is defined as w_E being the boundary of the emitter sided BE SCR at the same time and so forth until the neutral collector region at $x_{C,c}$. The indication of $|_{B'E'}$ means quasi-static or small-signal simulation with BE short while $|_{B'C'}$ stands for simulation with BC short.

$$w_E = x\left(\left.\frac{\partial \rho}{\partial V_{B'E'}}\right|_{V_{B'C'}} = min\right) \quad \text{emitter sided BE SCR boundary} \tag{2.31}$$

$$x_e = x\left(\left.\frac{\partial \rho}{\partial V_{B'E'}}\right|_{V_{B'C'}} = max\right) \quad \text{base sided BE SCR boundary} \tag{2.32}$$

$$x_{ci} = x\left(\left.\frac{\partial \rho}{\partial V_{B'E'}}\right|_{V_{B'C'}} = min\right) \tag{2.33}$$

with a limit at the metallurgical junction (x_{jc}) to take into account the base push-out

$$x_c = min[x_{ci}, x_{jc}] \quad \text{base sided BC SCR boundary} \tag{2.34}$$

$$x_{C,c} = x\left(\left.\frac{\partial \rho}{\partial V_{B'C'}}\right|_{V_{B'E'}} = min\right) \quad \text{collector sided BC SCR boundary} \tag{2.35}$$

The respective widths of the transistor regions can then be obtained as:

$$w_B = x_c - x_e \quad \text{neutral base} \tag{2.36}$$

$$w_{BE} = x_e - w_E \quad \text{BE SCR width} \tag{2.37}$$

$$w_{BC} = x_{C,c} - x_{ci} \quad \text{BC SCR width} \tag{2.38}$$

$$w_i = max[x_{ci} - x_{jc}, 0] \quad \text{collector injection region width} \tag{2.39}$$

Using the doping profile of the BiCMOS9MW technology the described boundaries of SCR and NR were determined.

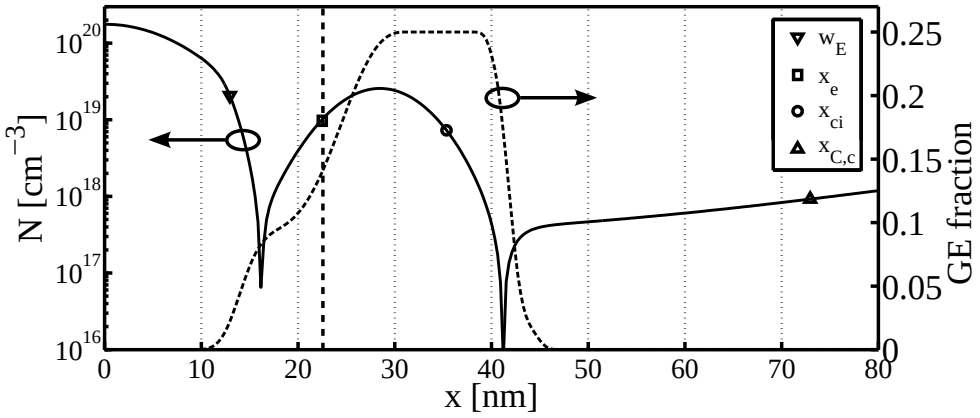


Figure 2.18: BiCMOS9MW vertical doping profile with indication of SCR boundaries by application of REGAP to numerical device simulation under forward bias ($V_{BE} > 0V, V_{BC} = 0V$) [*discrete points: SCR boundaries; solid line: net doping profile; dashed line: Ge profile; vertical dashed line: BE SCR boundary with the neutral base*]

A schematic view of the injection with w_i as used in the HICUM model is given in the zoom on the BC junction in Fig. 2.19 below.

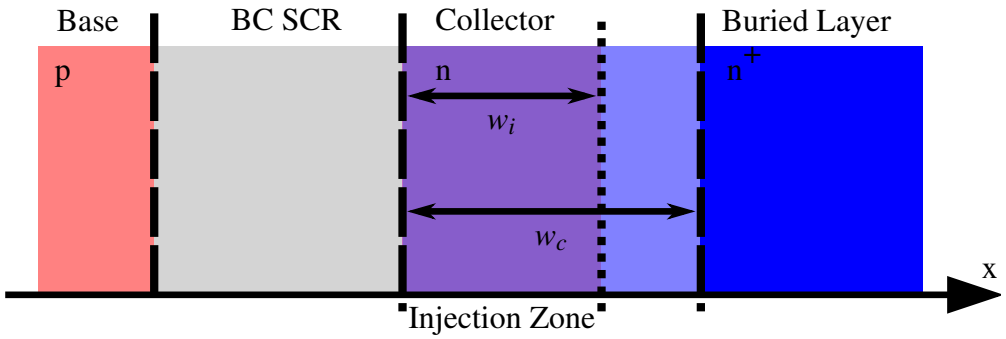


Figure 2.19: Schematic representation of injection width w_i in the collector; *detailed view of the BC junction*

Since special attention shall be given to the BE SCR end indicating the onset of the neutral base, Fig. 2.20 presents a zoom into the BE SCR boundary drawn on the vertical profile with variation of BE bias under low injection condition. One can clearly see how the boundary moves towards the emitter for increased BE bias, lowering the Ge fraction at the BE SCR boundary x_e.

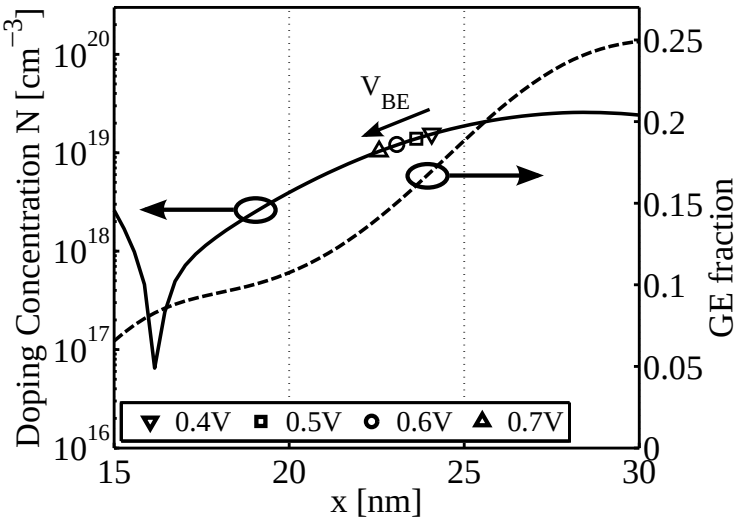


Figure 2.20: BiCMOS9MW vertical doping profile with indication of SCR boundary x_e, marking the onset of the neutral base under low injection for different V_{BE}

Using the methods described above the end of the BE SCR was determined for zero-bias at a vertical device depth of approximately 25nm while the boundary is shifted towards the emitter for increasing forward bias of the BE junction (cf. Fig. 2.20).

2.4.2 Idealization of the Ge Profile

A reference to build the device profile for numerical device simulation was the data from TCAD based process simulation. The process simulation was calibrated to measured data. However device simulation results showed that the initially assumed germanium profile obtained from process simulation was too steep. This profile caused the reverse Early effect to be highly overestimated compared to the actual measured data (cf. Fig. 2.22). The profile denominated 'fit' is in full accordance with the data from process simulation. The 'adapted' germanium profile however is tailored to match the measured data from electrical characterization of the BiCMOS9MW reference wafer.

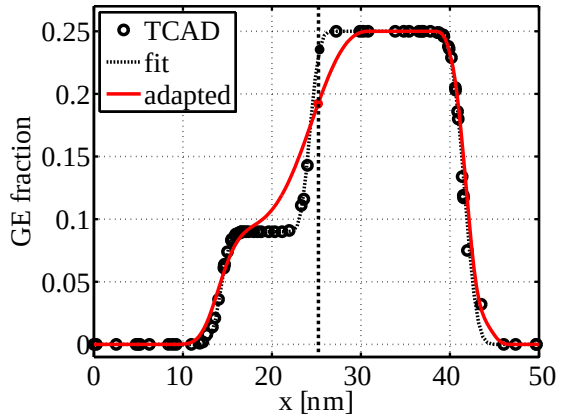


Figure 2.21: Germanium profiles used for numerical device simulation [*solid lines: numerical device simulation, discrete points: reference profile of process simulation*]

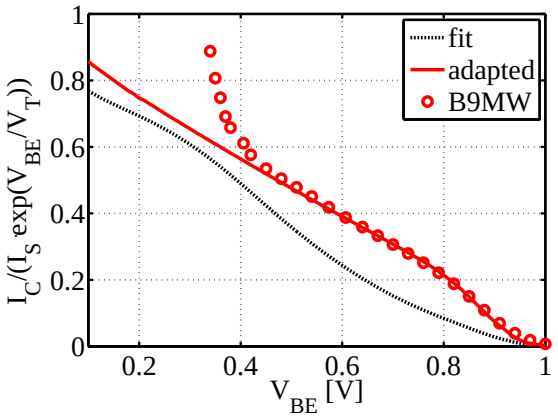


Figure 2.22: Normalized collector current of device simulation (*solid*) compared to measured data from BiCMOS9MW technology (*points*)

In order to show the validity of the Ge profile used for device simulation as well as the used material parameters the simulation results were then compared to the measurement data taken at different temperatures over a range of -20 through 125C. The results are shown in figure 2.24. Comparing the normalized currents, one can conclude that the simulation fits the measured data reasonably well. It is hence feasible to use the 'adapted' germanium profile as a reference for subsequent comparisons.

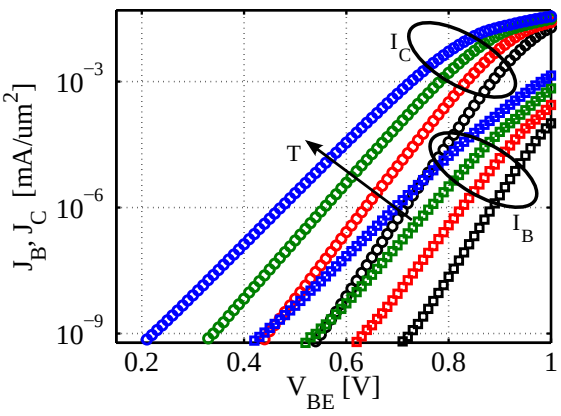


Figure 2.23: Data from forward Gummel measurement versus temperature [*squares: I_B, rectangles: I_C*]

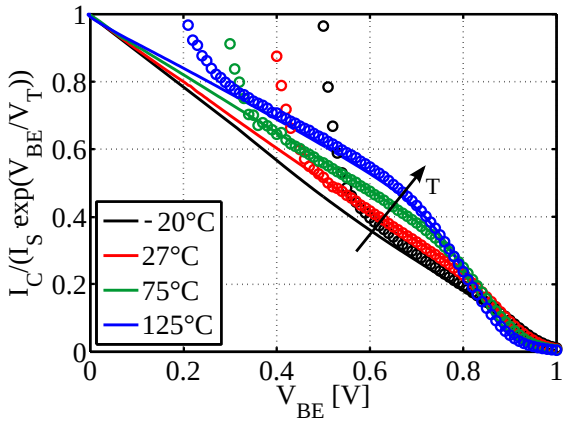


Figure 2.24: Comparison of measured data versus numerical device simulation at different temperatures [*solid lines: numerical device simulation, discrete points: measurement*]

In the subsequent comparisons the profile denominated 'B9MW' represents the *adapted* Ge profile calibrated to measured data. The influence on the most important device characteristics

(f_T, β, V_{AR}) were monitored throughout the simulation trials. The reverse Early voltage was calculated using data at a bias of V_{BC} = 0V (BC junction at zero bias). For each simulation the saturation current I_S, as the zero bias collector current, was linearly extrapolated from the I_C versus V_{BE} curves.

2.4.3 Impact of the On- and Offset Position of the Ge Profile

In a first experiment the impact of the position of the onset as well as the tail of the Ge profile were investigated through shifting them by several nm (cf. Fig. 2.25). However as evidenced by the normalized current ($I_{C,n}$) calculated for the simulation results, a variation of the Ge profile onset position (cf. Fig. 2.26) has no significant impact on the variation of the collector current with the BE junction bias. Since the Ge shape at the the end of the BE SCR x_e does not vary, this is in line with expectations from device physics (cf. equation 2.18).

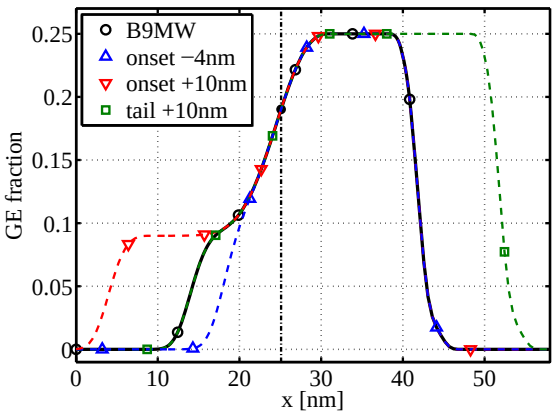


Figure 2.25: Different tested Ge profile shapes to analyze the impact of a shift in onset / tail of the profile [*vertical dashed line: end of the BE SCR (x_e)*]

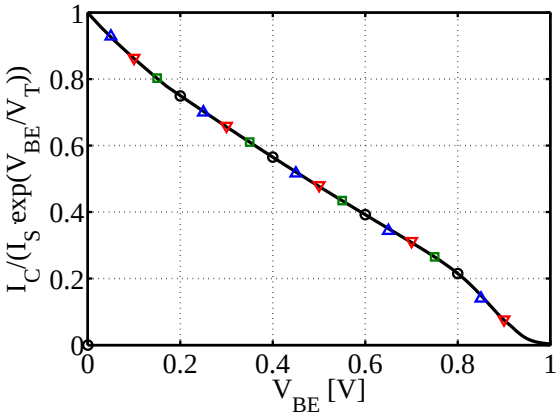


Figure 2.26: Normalized collector current $I_{C,n}$ for variation of the Ge profile onset and tail

However the results of quasi-static small signal device simulation confirm that there is a small impact on the transit time (f_T) as well as the current gain (β) as seen in Fig. 2.27 and 2.28.

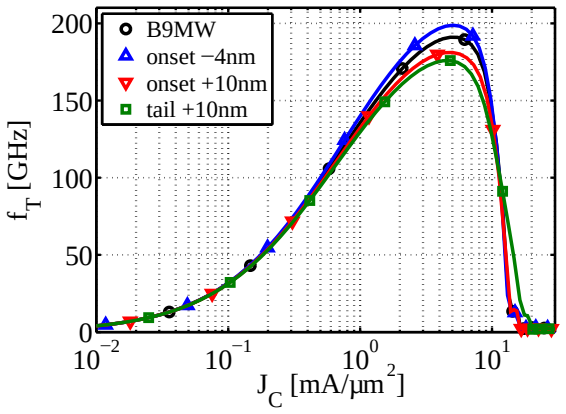


Figure 2.27: Simulated transit frequency (f_T) as a function of current density for profile variation

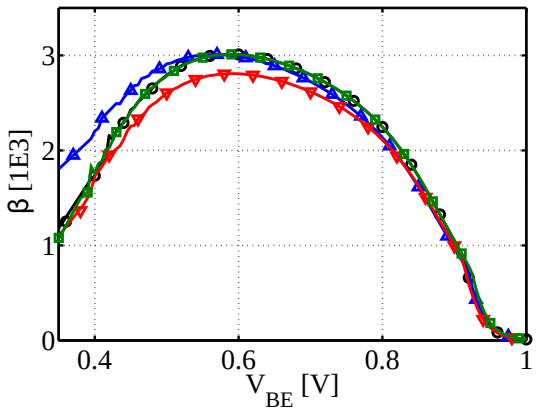


Figure 2.28: Simulated forward current gain β as function of V_{BE} bias

Given the results of the different device simulation trials, one can conclude that the slope of the Ge alloy profile at the BE SCR end is the most important factor of impact for the reverse early voltage, whereas the variation of onset and tailing edge has next to no impact on the DC characteristics. Due to formation of undesired barriers however these variations might have a negative impact on other important characteristics such as f_T or f_{max}.

2.4.4 Idealization of the Ge Profile

For a second trial the doping profile was kept constant again. However the germanium profile was approximated by an ideal trapezoidal shape to study the effect of the profile steepness on the reverse Early voltage. From the adapted profile the slope of the Ge profile at the end of the BE SCR was determined as 1.85% increase of the Ge fraction per one nanometer of vertical profile depth (1.85%/nm).

The absolute value of the Ge concentration was kept constant at 25% yet the smooth curvature obtained from process reproduction was replaced by a sharp, idealized trapezoidal Ge profile. The goal was to match the Ge fraction at the BE SCR border x_e at zero bias (cf. Fig. 2.29). In a first set of simulations the slope of the Ge profile was matched ($\Delta Ge/\Delta x = 1.85\%/nm$). For comparison a second profile with a less aggressive profile using a significantly smaller slope ($\Delta Ge/\Delta x = 0.8\%/nm$) was created.

For the trapezoidal profile with matched germanium slope ($1.85\%/nm$) the reverse Early voltage as well as the normalized collector current are in perfect agreement with the reference profile (cf. Fig. 2.30). This observation matches expectations, showing that a less aggressively scaled Ge profile significantly improves the V_{AR} value.

Hence it can be concluded that the main impact factor for the bias dependence of the collector

current is the gradually increasing Ge profile in the base. To be more precise, the Ge profile slope end of the BE SCR (x_e) is clearly the differentiating impact factor.

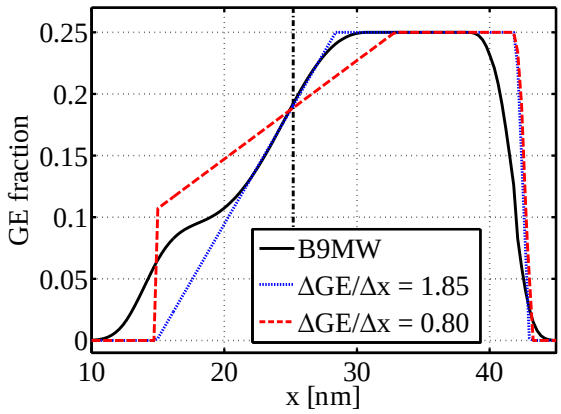


Figure 2.29: Germanium profiles for trapezoidal Ge profile [*vertical dashed line indicates the end of the BE SCR* x_e]

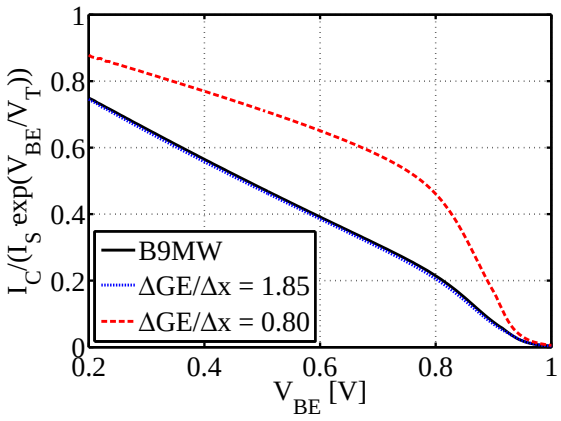


Figure 2.30: Numerical device simulation comparison of normalized collector current $I_{C,n}$ for Ge profile slope variation

To verify the results two more simulations with constant Ge profile as well as without germanium were performed. For the constant Ge profile the concentration was adjusted to the value found at x_e at zero bias (approx. 18%). Except for the high current region the results of the normalized collector current in figure 2.32 are matching for both scenarios with and without germanium. This confirms the observations from the previous trial. The bias dependence of I_C is not distinctively linked to the concentration of the germanium fraction but rather to the relative variation ($\Delta Ge/\Delta x$) with the depth of the vertical profile.

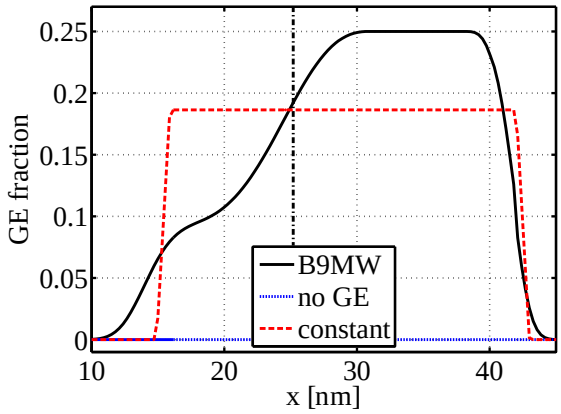


Figure 2.31: Germanium profiles for constant Ge profile with BiCMOS9MW data for comparison, [*black dashed line indicates the end of the BE SCR* x_E]

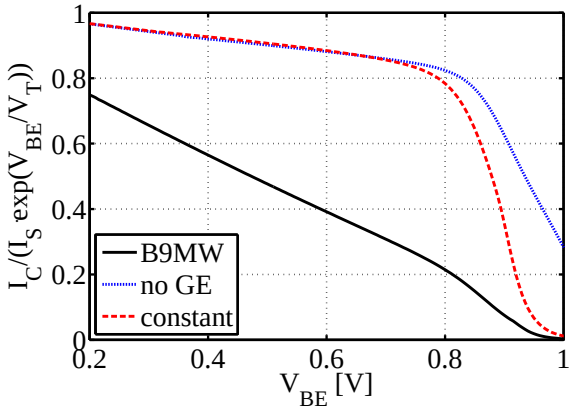


Figure 2.32: Normalized collector current $I_{C,n}$ comparison for constant Ge profiles

To show the impact on the speed of the device, figures 2.33 and 2.34 indicate that a constant Ge profile is almost as bad for the transit time as no germanium at all whereas the slope of the Ge profile does not have a significant impact on f_T or the transit time respectively. However it has to be noted that the simulations were performed using a drift-diffusion based device simulator [43]. This approach is not as accurate as a simulation using a hydrodynamic transport model or a statistical numerical technique as the Monte Carlo technique solving the Boltzmann transport equation with the dynamics of the individual charge carriers. However it provides reasonable accurate results with less requirements of extensive computing.

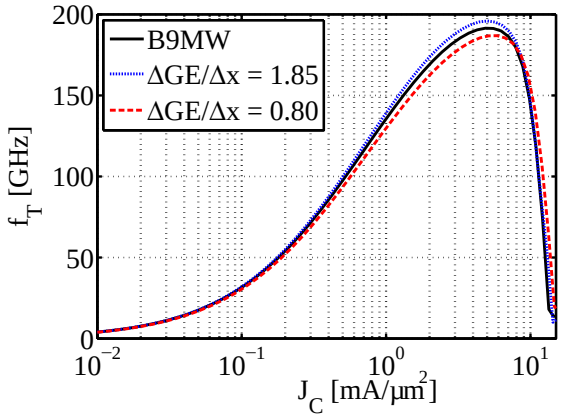


Figure 2.33: Device simulation results of RF FoM f_T for trapezoidal Ge profiles

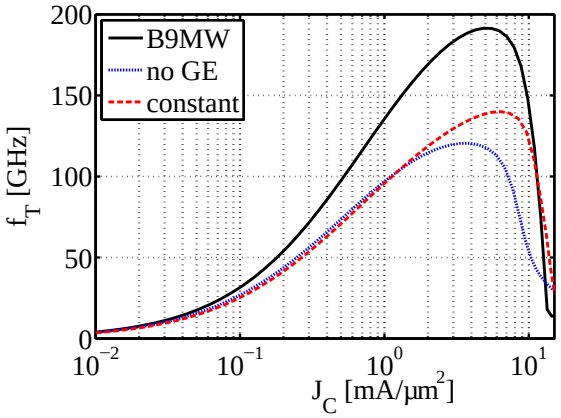


Figure 2.34: Device simulation results of RF FoM f_T for constant Ge profiles

2.5 Model Adaptations to Improve Model Accuracy with Advanced HBT Technologies

Given recent technology advancements in the framework of recent research projects targeting higher speed devices and circuits the device models were due for an update [46]. As shown in the previous section the most critical issue encountered in SiGe bipolar device modeling was the variation of the BE SCR with the applied BE junction bias as well as the temperature dependence. Therefore in this section the most important model improvements shall be summarized, emphasizing on the new formulation of the transfer current under low injection condition. The bias dependent BE depletion charge weighting factor in HBTs (denoted h_{jei}).

The latest major model release is the HICUM version L2.3x presented in the framework of dotfive [25]. After the last major update of the model to version 2.20 in 2005 [47] the model update to version L2.30 was introduced in 2011 [13] followed by minor incremental updates to the current[1] release version L2.32 [48].

2.5.1 Derivation of the Formulation of the BE Space Charge Weighting Factor introduced in HICUM L2.30

A problem analysis as well as new extension for the HICUM formulation of the transfer current I_T, effective at medium current densities was presented in 2009 [49]. New physics-based model equations with additional model parameters were proposed in order to add bias dependence to the charge weighting factor for accurate Q_{p0} modeling. As a result the new model formulation calculates h_{jEi} as a function of the internal BE junction bias V_{BEi}.

The starting point for the derivation of the bias dependent weighting factors are the simplifications given in [50]. The implementation used for the HICUM model follows closely the model derivation used in the MEXTRAM model as presented in [51, 52], where the changes were introduced to improve the description of SiGe based HBT devices [53]. To find a solution with a simple formulation the fundamental assumptions (idealizations respectively) are as follows:

- constant doping profile within the base region [$p(x) = N_B = \text{const.}$]
- carrier mobility is constant throughout the base region [$\mu_n = \overline{\mu_n} = \text{const.}$]
- doping-induced band-gap narrowing in the base is considered negligible compared to band-gap change due to the Ge incorporation

As a consequence the case of the abrupt pn junction is applicable and hence the simplified

[1] at the time of writing

relation between junction voltage and space charge region width w_{SCR} reads as follows [54]:

$$w_{SCR}(V_a) = w_0\sqrt{1-\frac{V_a}{V_D}} \text{ and } x_p(V_a) = x_0\sqrt{1-\frac{V_a}{V_D}}. \tag{2.40}$$

The weighting factors introduced for the first major model update to version 2.1 [37] remained unchanged during the transition model version 2.2x. Both charge weighting factors are calculated as average values independent of junction bias by means of the ratio of mobility and intrinsic carrier density between two neighboring regions:

$$h_e = \frac{\overline{\mu_{nB} n_{iB}^2}}{\overline{\mu_{nE} n_{iE}^2}} \text{ and } h_c = \frac{\overline{\mu_{nB} n_{iB}^2}}{\overline{\mu_{nC} n_{iC}^2}}. \tag{2.41}$$

However as shown in section 2.4, with higher Ge fraction and steeper profiles found in recent HBT technologies the model did not sufficiently describe the physical behavior of the devices. Hence the model formulation of the parameter h_{jEi} was adapted to reproduce the dependence on the bias of the BE junction.

The starting point for the derivation of the weighting function $h(x)$ is the transfer current description itself.

$$I_T \propto \frac{1}{\int_{x_1}^{x_2} h(x)p(x)dx} \exp\left(\frac{V_{BEi}}{V_T}\right) \tag{2.42}$$

The same weighting function $h(x)$ in the denominator of the transfer current equation is found in the formulation of the charge definition (cf. equation (2.22)). The BE charge weighting factor h_{jEi}, that accounts for the reverse early effect, is derived by application of the above mentioned simplifications to the general equation describing the relation between base and emitter

$$h_{jEi} = \frac{\int_{x_1}^{x_2} q \frac{\mu_{n0} {n_{i0}}^2}{\mu_n {n_i}^2} N_B \, dx}{\int_{x_1}^{x_2} q \, p(x) \, dx}. \tag{2.43}$$

With all constant contributions taken out of the integrals over the BE space charge region $[x_1, x_2]$ the equation can be simplified to the following form where the doping cancels out of the equation

$$h_{jEi} = \frac{\frac{\overline{\mu_{n0}}}{\overline{\mu_n}} \overline{n_{i0}}^2 \int_{x_1}^{x_2} \frac{1}{{n_i}^2} dx}{\int_{x_1}^{x_2} dx}. \tag{2.44}$$

The integration limits involved in the calculation of the BE junction weighting factor h_{jEi} are the metallurgical junction between base and emitter on the one side $[x_1 = x_{p0}]$ and the end of the BE SCR on the other side $[x_2 = x_p]$. The boundary of the BE SCR and the neutral base however changes with the BE bias. One hence needs to replace the limits of the integral by equation (2.40). Representing the intrinsic carrier concentration by a relative change with the carrier concentration

at the metallurgical junction x_{p0} as reference gives

$$n_i(x) = n_i(x_{p0})^2 \exp\left(\frac{x}{a_{ni}}\right), \tag{2.45}$$

with the factor a_{ni} representing a the position-normalized band-gap difference. The full weighting-factor formulation then reads

$$\begin{aligned} h_{jEi} &= \frac{\frac{\overline{\mu_{n0}}}{\overline{\mu_n}}\overline{n_{i0}}^2 \int_{x_{p0}}^{x_p} \frac{1}{n_i(x_{p0})^2 \exp\left(\frac{x}{a_{ni}}\right)} dx}{x_{p0} - x_{p0}\sqrt{1-\frac{V_{BEi}}{V_{DEi}}}} \\ &= \frac{c\, a_{ni}\left[\exp\left(\frac{-2x_p}{a_{ni}}\right) - \exp\left(\frac{-2x_{p0}}{a_{ni}}\right)\right]}{x_{p0}\left[1-\sqrt{1-\frac{V_{BEi}}{V_{DEi}}}\right]} \text{ with } c = \frac{\overline{\mu_{n0}}}{\overline{\mu_n}}\overline{n_{i0}}^2 \end{aligned} \tag{2.46}$$

In a next step the weighting factor is normalized to his zero-bias value h_{jEi0} that is obtained at a base-emitter bias of zero volt. However since h_{jEi0} is given in a indeterminate form, l'Hôpital's rule is applied for the limit calculus.

$$\begin{aligned} h_{jEi0} &= \lim_{V_{BEi}\to 0} h_{jEi} = \lim_{V_{BEi}\to 0} \frac{f(x)}{g(x)} = \lim_{V_{BEi}\to 0} \frac{f'(x)}{g'(x)} \\ &= \lim_{V_{BEi}\to 0} \frac{c}{n_{i,BE}^2} \frac{\exp\left(-\frac{2x_{p0}}{a_{ni}}\sqrt{1-\frac{V_{BEi}}{V_{DEi}}}\right)}{1-\frac{V_{BEi}}{V_{DEi}}} = \frac{c}{n_{i,BE}^2}\exp\left(-\frac{2x_{p0}}{a_{ni}}\right) \end{aligned} \tag{2.47}$$

In a last step the formulations obtained from equations (2.46) and (2.47) are merged into the final form as shown below:

$$\begin{aligned} h_{jEi} &= h_{jEi0}\frac{h_{jEi}}{\frac{c}{n_{iBE}^2}\exp\left(\frac{-2x_{p0}}{a_{ni}}\right)} \\ &= h_{jEi0}\frac{\exp\left(\frac{2x_{po}}{a_{ni}}\left[1-\sqrt{1-\frac{V_{BEi}}{V_{DEi}}}\right]\right)-1}{\frac{2x_{po}}{a_{ni}}\left[1-\sqrt{1-\frac{V_{BEi}}{V_{DEi}}}\right]} \end{aligned} \tag{2.48}$$

$$h_{jEi} = h_{jEi0}\frac{\exp(u)-1}{u} \quad \text{with} \quad u = \frac{2x_{p0}}{a_{ni}}\left[1-\sqrt{1-\frac{V_{BEi}}{V_{DEi}}}\right] \tag{2.49}$$

To make the formulation fully compatible with the existing HICUM model equations the square root is replaced by the exponential factor z_{Ei} that is used to adjust the voltage dependence of the junction capacitance C_{jBE}. The factor $2x_{p0}/a_{ni}$ is then summarized in a parameter a_{hjEi} being new,

temperature dependent, model parameter with its according temperature exponent $\xi_{h_{jEi}}$

$$u = a_{hjEi}\left(1 - \left(1 - \frac{v_j}{V_{DEi}}\right)^{z_{Ei}}\right), \tag{2.50}$$

$$\text{with } a_{hjEi}(T) = a_{hjEi}(T_0)\left(\frac{T}{T_0}\right)^{\xi_{h_{jEi}}} \tag{2.51}$$

2.5.2 High Injection Effects

At *low level injection* the change in minority carrier concentration m outside the space charge region that is caused by the injected minority carriers is small compared to the ionized carriers (e.g. $n \ll N$). For the case of a ideal pn junction this means that there is no electric field outside the SCR and the transport of injected carriers is exclusively due to the diffusion mechanism. Hence the condition for low level injection can be defined as the diode operation where the excess minority carrier concentration is much less than the equilibrium majority carrier concentration.

Contrary with increasing device current, more carriers contribute to the charge transport. If the excess minority carrier concentration approaches the equilibrium majority carrier concentration ($n \approx N$), the device shows *high level injection* [55]. High injection effects occur in a bipolar junction transistor, just like in a pn diode causing to invalidate the approximations made in the derivation of the ideal diode characteristics. For the case of a forward biased bipolar device one can define that high injection occurs when the minority carrier density in the base is equal or larger than the base doping ($n \geq N_B$).

The importance of the high injection regime for device modeling in silicon based BJTs is significant, since transistors operated in the high injection regime show several phenomena that cause a significant deviation from the ideal current characteristics derived at low injection ($I_B, I_C \sim exp(V_{BE}/V_T)$). Those effects include the Kirk-effect [33], the Webster-effect [32], voltage drops in the neutral regions of emitter and collector as well as quasi-saturation behavior due to the collector series resistance. Given the fact that bipolar transistors achieve their maximum RF performance at very high collector current densities[1], the accurate description of transit time under high injection is crucial for accurate circuit simulation.

In addition to the impact on the general transfer current behavior, the Ge profile also has an impact on the bias dependence of the stored minority charge at high current. Although the critical current density (J_{CK}) itself does not change, the physical mechanism that increases the transit time τ_f is different in SiGe HBTs compared to classical bipolar devices. Given the additional band-gap barrier resulting from the usually sharp drop of the Ge at the BC junction, holes accumulate on the base side of the junction, leading to a rapid increase of the stored minority charge in the base

[1] The term *high* generally refers to collector current densities in the order of a few $mA/\mu m^2$

region once J_C approaches J_{CK}, so that the electric field at the junction starts to drop.

Kirk Effect and Heterojunction Barrier Effect (HBE)

The barrier effect (HBE) is strongly related to the Kirk effect. The effect that occurs under high injection was fist described in [33] and deals with the apparent base push out at high current densities.

Physically, at high injection the injected electrons add to the negative space charge on the base side of the base collector junction hence the minority carrier concentration in the CB SCR exceeds the space charge on the collector side of the region causing the SCR to collapse and the base to be pushed out into the collector. In turn this reduces the depletion width and increases the quasi-neutral base width w_B. Hence the Kirk effect is often referred to as the base-pushout effect.

Under this condition the effective width of the base layer equals the width of the base and collector space charge layer leading to a significant rise in the base transit time τ_B leading to the well known trade-off of high frequency device performance and operating voltage described by the "Johnson-limit" [56].

Another significant high injection phenomena encountered in bipolar transistors is the Heterojunction Barrier Effect (HBE). This high injection effect is related to the base-collector heterojunction leading to a conduction band barriers and a non-negligible base current component due to neutral-base recombination.

Several studies were carried out to investigate and describe the barrier effect encountered in SiGe HBTs [57–59]. Yet experimental verification is difficult due to the multitude of high current effects occurring at high forward bias. The effect is best seen in high-speed SiGe BJT that exhibit a significant base current increase at collector current densities exceeding the critical current. Hence in the forward Gummel plot of measured data this is visible at high current, where the base current increases significantly while the collector current drops.

A device simulation comparison using the BiCMOS9MW doping profile with and without Ge background shows clearly how the base current at high injection deviates due to the barrier at the BC junction.

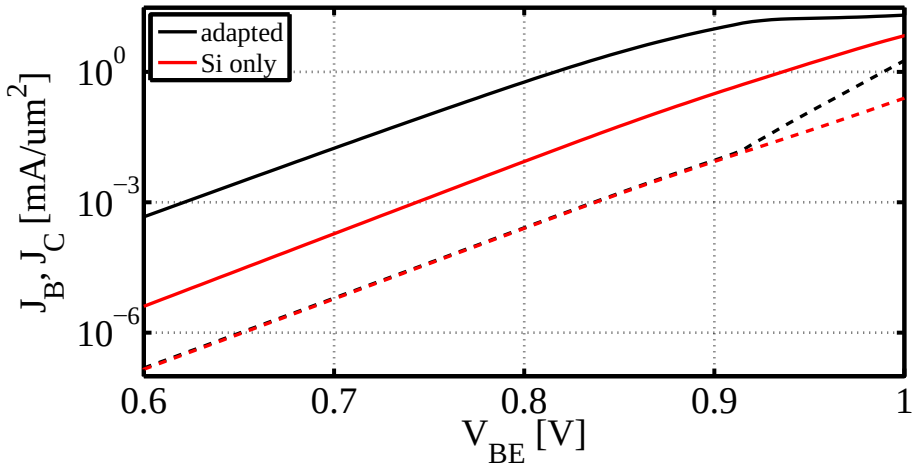


Figure 2.35: Forward Gummel characteristic for identical doping profile with and without Ge background [*apparent influence on currents I_C (solid) and I_B (dashed) in high current operation ($V_{BE} > 0.9V$)*]

The undesired HBE and neutral-base recombination effects are usually avoided by proper SiGe HBT profile optimization. A design goal is not to have a sharp Ge profile drop in the BC junction. However this can not easily be reproduced by means of device simulations. A profile based on the B4T device generation was used for two more trials. In this technology the Ge profile is having a less steep Ge ramp-up in the BC junction by default.

However the total Ge concentration in the base epitaxial layer was increased to 30%[1] giving rise to the BC barrier. In order to study the influence of the shape of the Ge profile at the BC junction a second Ge profile (denoted 'long tail') was created (cf. Fig. 2.36). Here the same impurity profile as for the standard process is used while modifying the Ge tail. The intention of using a more gradual Ge is to reduce the discontinuous band in the BC junction band barrier. However as evidenced by the DC characteristics even a significant modification of the BC profile shape does neither improve the base current nor influence the HBE (cf. Fig. 2.37).

[1] compared to 25% in the BiCMOS9MW generation

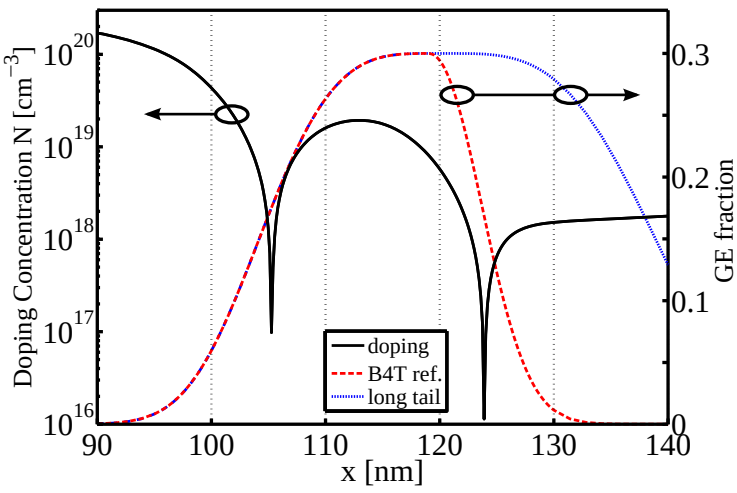


Figure 2.36: Base region zoom with Ge profile background for device simulation of B4T SiGe HBT doping profile [*two different Ge profiles with steep (red) and gradual (blue) BC tail*]

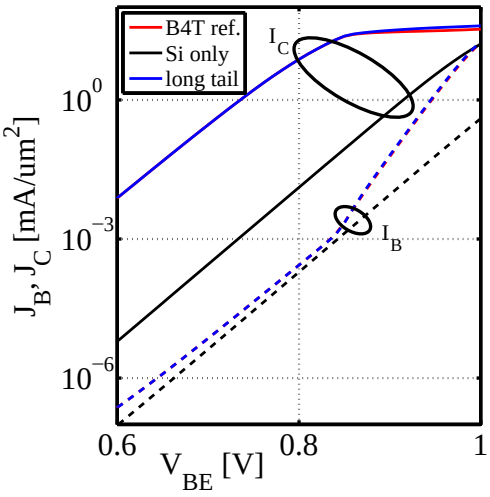


Figure 2.37: Forward Gummel characteristic for identical doping profile with variation of the Ge background [*Ge profile influence on device currents I_C (solid) and I_B (dashed)*]

The HBE occurs in bipolar technologies with varying band-gap. The HBE can be influenced by the design of the collector. More precisely it can be altered by the composition of the SiGe-Si hetero-interface in the BC junction where the transition of the narrow base band-gap base to the larger band-gap Si introduces a valence band offset blocking carrier transition. At low injection the band offset is superposed by the band bending in the BC SCR (cf. Fig. 2.6).

Hence under high injection, when the electric CB field collapses, the effect is more or less pronounced depending on the device design. Yet these days it plays an important role in the device characteristics and due to the fact that SiGe based technologies achieve their maximum performance at high collector currents, the accurate modeling of the HBE and it's impact on key figures of merit (g_m, f_T, f_{max}) is crucial. The aforementioned effects hence have gradually been taken into account by the advanced bipolar compact transistor model HICUM.

2.5.3 Formulation of the Critical Current I_{CK}

Special attention is paid to the current flow in both vertical and lateral directions for operation in high injection. Under high injection condition the electron density in the base reaches the level of the hole concentration ($n \backsimeq p \gg N_B$). Several effects are no longer negligible and have to be taken into account.

Especially the reduction of the transit time with increasing current density is critical. A critical current I_{CK} is defined that presents the onset of the high injection region.

$$I_{CK} = \frac{v_{ceff}}{r_{Ci0}} \frac{1}{\left[1 + \left(\frac{v_{ceff}}{V_{lim}}\right)^{\delta_{CK}}\right]^{1/\delta_{CK}}} \cdot \left[1 + \frac{x + \sqrt{x^2 + a_{ickpt}}}{2}\right], \tag{2.52}$$

with a_{ickpt} being a fixed parameter at a default value of 10^{-3}. The factor new parameter δ_{CK} was introduced within the model update to HICUM version 2.30. Setting its value to 2 defaults the equation back to the original formulation using a square root. Within the smoothing function in brackets the factor $x = (v_{ceff} - V_{lim})/V_{PT}$ makes for transition between low and high electric fields in the collector defined by the voltage v_{lim} and the punch-through voltage V_{PT}.

$$v_{ceff} = \left[1 + \frac{u + \sqrt{u^2 + a_{vceff}}}{2}\right] = V_T \text{ , with } u = \frac{v_c - V_T}{V_T} \tag{2.53}$$

2.5.4 Model Validation for HICUM L2.3x

Low Current Charge Weighting Factor h_{jEi}

The measurement of the collector current for zero BC junction bias (fwd. Gummel, V_{BC0}) was one of the most critical characteristics and a driver for model improvement. The measured data for the reference device with a drawn emitter length of $5\mu m$ is shown in Fig. 2.38 below. The lines show the simulation result of a single device geometry with a discrete model card for the old (ver. L2.2x) as well as the new (ver. L2.3x) HICUM implementation. In the old model version the model parameters c_10, q_p0 as well as h_{jei} were optimized. For simulation with version L2.3 however the new parameter a_{hjei} was included in the extraction and binned to a value of $a_{hjei} = 1.45$ (cf. Chapter 3).

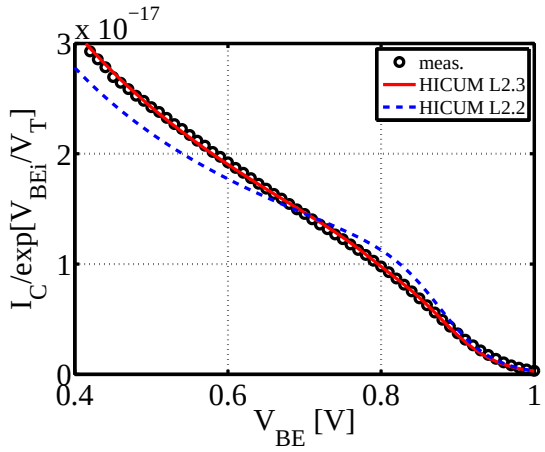


Figure 2.38: Normalized collector current $I_{C,n}$ versus V_{BEi} for BiCMOS9MW technology generation; *Model comparison of L2.2 and L2.3*

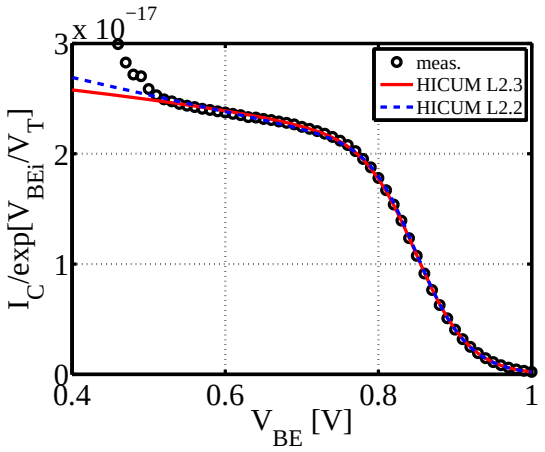


Figure 2.39: Normalized collector current $I_{C,n}$ versus V_{BEi} for BiCMOS55 technology generation; *Model comparison of L2.2 and L2.3*

One can clearly see that for the older device generation of BiCMOS9MW devices the significant slope of the collector current with BE junction bias cannot be reproduced with a constant weighting factor h_{jei} as used in the old model L2.2x. However through process optimization and careful design of the vertical profile the effect seen in medium bias regions was largely diminished making the more recent BiCMOS55 generation less susceptible to the model version.

The Heterojunction Barrier Effect (HBE)

Improved formulations taking into account the barrier effect have been proposed [60]. The approach used in the HICUM formulation uses an additional current offset ΔI_B. The model replicates the increase of the total transit time observed at the onset of high current region in HBT. The charge $\Delta Q_{Bf,b}$ is added to the total charge Q_f. Here the according Kirk-effect related collector charge and transit time increase is delayed by the barrier voltage which is represented by the model parameter V_{cBar}. By setting $V_{cBar} = 0$ the entire barrier effect formulation is turned off.

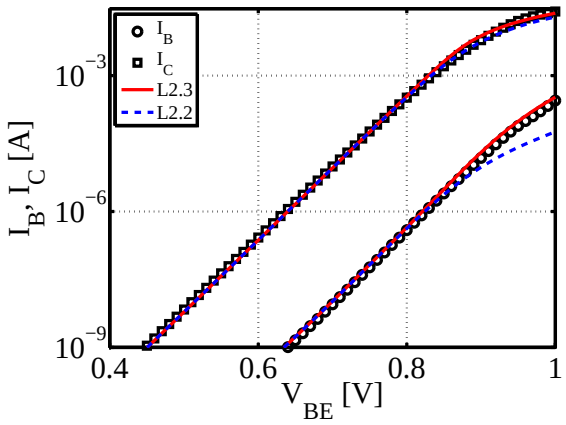


Figure 2.40: Forward gummel characteristic with I_B and I_C versus V_{BE} for BiCMOS9MW technology generation; *Model comparison with L2.2 ($V_{cBar} = 0$) and L2.3 ($V_{cBar} \neq 0$)*

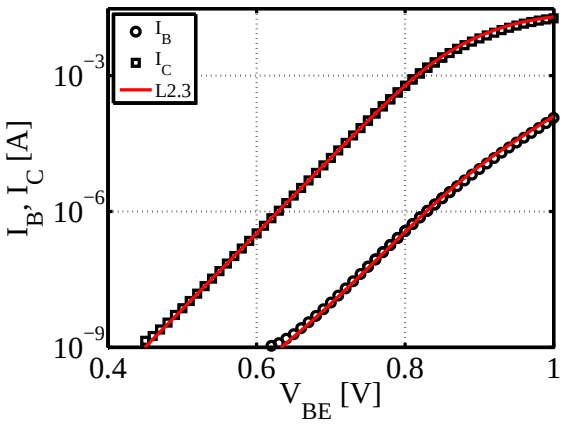


Figure 2.41: Forward gummel characteristic with I_B and I_C versus V_{BE} for BiCMOS55 technology generation; *Model comparison with simulation using $V_{cBar} = 0$*

In the above figure (cf. Fig. 2.40) another validation of the model formulation is made comparing measured data to simulation. For the older BiCMOS9MW one can see that even given the limited bias range of V_{BE} up to $1V$ the impact on the base current cannot be modeled by the older formulation in version L2.2x. Activating and fitting the parameters for the new formulation with otherwise unchanged parameters effectively solves the problem.

Given the improved Ge profile design already evidenced by the less significant reverse Early effect of the BiCMOS55 generation one can see in Fig. 2.41 that the old implementation (deactivating the barrier effect with $V_{cBar} = 0$) is sufficient to have a good agreement of base and collector current of the model compared to measured data.

2.6 Conclusion

For a clear understanding of device physics, this section summarized the basic semiconductor working principles as well as essential equations. The mechanisms of carrier transport, junction theory and fundamental charge-relations for bipolar transistor operation have been presented. The origin of the fundamental concept of charge storage models and their extension to heterojunction devices was shown.

The basic description of static and dynamic behavior of bipolar devices has been presented. In a comparison study of one-dimensional device simulations, the effects seen in measured devices have been re-produced, with a focus on the reverse Early voltage. The original device profile has been calibrated to match the measured characteristics and was used to re-simulate. The origin of newly implemented model equations suitable for recent SiGe based RF/microwave device technologies was verified and the applicability of the new model implementation has been tested with regards to bias dependence and temperature scaling.

As with every other compact model most of the newly introduced model features and extensions added flexibility and extend the validity to new phenomena or technologies. Yet the suitability and applicability of new equations has to be validated by derivation of the respective model equations and experimental comparison of the model with device simulation or measured characteristics. Furthermore, adding new model parameters increases the extraction effort due to the higher number of parameters to be correctly determined from measured or simulated characteristics. Lastly, scalability of the new model parameters with device geometry as well as ambient conditions has to be validated.

Since the native HICUM formulation is only applicable to a single device or transistor, the model as described so far cannot cover the full selection of devices available in a process related design kit. The aim of compact modeling for industry standard technologies however is to provide a complete library. Thus individual scaling laws have to be defined for each parameter and implemented in a preprocessor, that is capable to generate individual model cards for various desired device configurations.

Chapter 3

Scalable Device Modeling with HICUM

With circuit simulators predicting the behavior of a given IC before manufacturing, transistor models are used for the most part of modern electronic design work. Recent advances have made BiCMOS technology a viable platform for RF/analog and millimeter-wave circuits. For any given technology the design of these circuits however requires accurate, scalable compact models for active transistors as well as passive components.

With the transition from micro- to nanoelectronics, the lateral and vertical scaling of devices brings challenges due to approaching basic physical limits. Approaching the limits of the used technology requires accurate modeling of devices with high speed, low noise and careful consideration of power consumption in case of mobile applications. For a first pass design success it is hence crucial to have a good agreement between measured characteristics and the simulated model over wide geometry, bias and temperature range.

The employment of geometry scaling is mandatory for cost-efficient design of analog integrated circuits. Depending on the specific need of a circuit element, the IC designer has the choice of process variants (vertical profile) and transistor configurations (number of contacts and emitter fingers) as well as geometries (scaling of junction area).

In order to achieve highest performance and bring competitive products to market, todays ICs for mixed-signal applications operate close to the performance limits of each device. Circuits must therefore be carefully optimized by selecting adequate transistor configurations. To drive a high current for example a wide transistor with long window length gives a large emitter area capable of providing the desired transfer current. Yet strong self-heating limits the applicability of large emitter areas and thus for high speed operation at increased current densities a designer may prefer to use a smaller transistor (e.g. $w_{E,min}$) with multiple emitter fingers. The delivered process design kit (PDK[1]) hence needs to cover many scenarios and take into account all choices in the offer. The associated scalable model card needs to give very good agreement for DC and small-signal

[1] the PDK usually comprises: design rules, transistor models for circuit simulation and layout information

characteristics.

One hence desires a complete and continuous description of all relevant device characteristics over the full range of devices and configurations provided to circuit designers in the DK. In recent years the initial scaling approach using exclusively drawn device dimensions has been extended progressively and became more and more complex. Therefore this chapter is dedicated to special methods focusing on fully geometry-scalable model parameter extraction for the HICUM model.

3.1 Geometry-Scalable Compact Device Modeling

A compact model represents the bridge between a given technology and circuit simulation for IC design. Model complexity is ranging from physical models (with parameters based upon physical properties) over empirical models (based on curve fitting) to simple tabular models (look-up table). Even though they are more complex, analytical or semi-numerical descriptions of devices hold the advantage of capturing the essential device physics, distilling them into a tractable form. For accurate simulation nonlinear physics-based models are preferred given their ability to describe the entire operating area of a transistor.

The HICUM model, used to describe fast HBTs, is such a physics-based model with a sound physical base as well as physical description of the parts of a transistor. As a result, the model parameter values are expected to scale excellently with geometry. This makes it ideally suited for physically-based geometry scaling of its parameters and description of the various configurations and sizes available in a technology.

In contrast to the modeling of MOS transistors, where geometry scaling is partly a native feature of the global RF-CMOS model (e.g. through dedicated parameters for physical geometrical scaling in BSIM or PSP), extraction methodologies for geometry-scalable parameters are different for bipolar devices. The focus in model development is to preserve general applicability. In consequence none of the main industry standard compact models for bipolar transistors (namely SGP, VBIC, MEXTRAM, HICUM) features a built-in geometry scaling.

In addition actual, complete and accurate geometrical scaling is only possible if the technological process and the geometrical layout of the transistors to be modeled is known in some detail, which most of the time is restricted to the individual foundry or manufacturer. Furthermore advancements in technology and process (shrinking device geometry and profiles etc.) often lead to physical effects not properly covered by earlier implementations causing them to become inaccurate or obsolete. To overcome these limitations, new or improved versions of existing scaling equations need to be implemented as technology advances, making individual scaling approaches (maintained and developed per manufacturer / foundry) highly desirable.

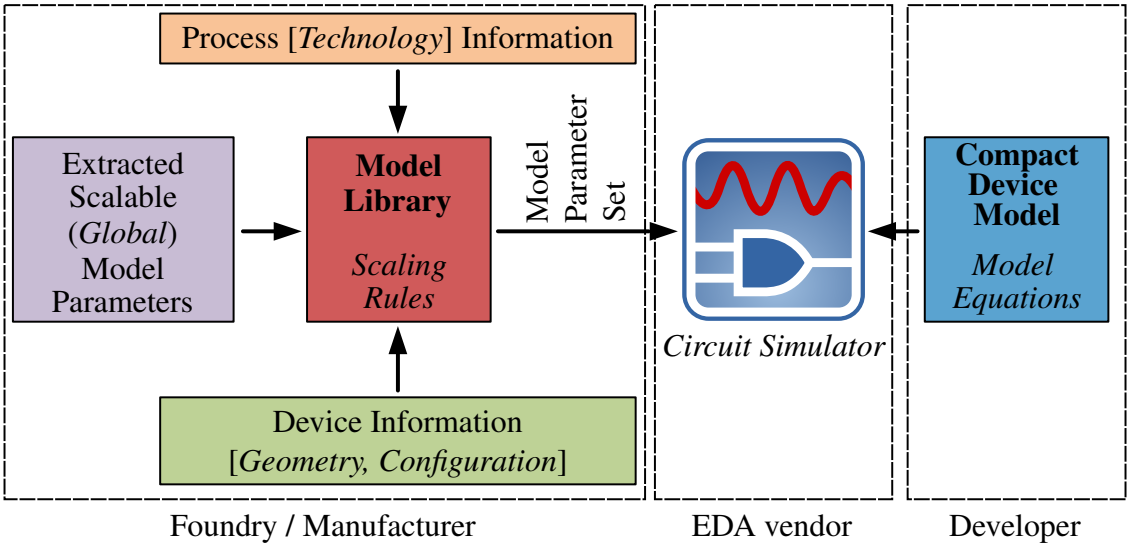


Figure 3.1: Prerequisites for scalable device simulation

A fully geometry scalable model is defined as a *hierarchical model*. The basis of geometry scaling is a *local model*, describing the characteristics of a single device with a particular device geometry. This so called mono-geometry approach comprises a full set of model parameters for a single geometry, giving complete description of the electrical characteristics and behavior of one device (e.g. currents and capacitances as a function of bias conditions and temperatures).

Since individual model parameters exhibit a certain dependence on device geometry, one can derive scaling rules. In a next level of the model hierarchy a scaling rule for each *local parameter* is attributed. Even though empirical scaling equations (derived from curve progression) may be defined, it is highly preferred to find these scaling rules based on device physics as far as possible.

Some of the compact model parameters are defined as a ratio of two physical quantities scaling with geometry, canceling out the area dependence. Hence in addition to scalable parameters there is a number of parameters that are independent of lateral device dimensions and constant for a given process. These parameters are called *global parameters*.

The fully scalable or *global model* is then comprised of global model parameters, customized scaling rules as well as device specific information in order to calculate sets of local parameters for each geometry. The scaling equations required for this are contained in the model library. Additional external inputs are certain information about the process (e.g. doping level and sheet resistances) as well as device-specific information such as the transistor configuration and size (cf. Fig. 3.1).

3.1.1 The Scalable Model Library

Supplemented with physically based geometry scaling of its parameters one single extracted nominal model may cover the whole offer, which is desired from a designers point of view.

Since existing bipolar models are not geometry-scalable by default, the geometrical scaling model is supposed to be added to the model, as a shell. These technology-specific custom solutions (scaling parameters and scaling equations) are usually included in the model library (cf. Fig. 3.1). A geometry pre-processor in the circuit simulator or in a sub-circuit is required to generate the appropriate model parameters.

However with an eye on the protection of IP (especially important for foundries) the scaling equations might as well be embedded in a software suite (as for example the TRADICA software suite [61]) where scaling equations are built into the (protected) program code.

To take full advantage of scaling, not only the emitter/base junction width and length must be scaled. Input parameters for geometry scalable NPN devices are linked to the offer given to designers in the so called parameterized cells (PCells). The individual PCell represents a design block (basic unit of functionality) that is automatically generated by the EDA tool based on the value of its instance parameters. The PCell library for automated design includes different device types to choose from (transistors, resistors, capacitors, diodes etc.). The specification of a Pcell for HS NPN HBTs is linked to parameters such as length, width and the finger number (multiplier). In addition the number of contacts asserts different device configurations. A single-sided base contact ($nbb = 1$) might for example be desired in order to use a transistor with reduced collector resistance ($R_{CX}(\mathrm{CBEBC}) < R_{CX}(\mathrm{CBE})$). The applicable ranges for the most recent technology are summarized in Tab. 3.1.

Input	Description	Default*	Range*
w_E	drawn emitter width	$0.20\mu m$	$w_{min} = 0.2\mu m \leq w_E \leq w_{max} = 0.42\mu m$
l_E	drawn emitter length	$5.56\mu m$	$l_{min} = 0.45\mu m \leq l_E \leq l_{max} = 10.0\mu m$
nbe	Number of emitter fingers	1	≤ 5
nbb	Number of base contact stripes	1	≤ 10
nbc	Number of collector contacts	1	≤ 6

Table 3.1: Set of input parameters for geometry scalable model of a high speed NPN SiGe HBT, [* *Default values and parameter range given for BiCMOS55 technology*]

3.1.2 Emitter Window Scalability

General Emitter Window Scalability

In general the aim of the emitter window scaling approach is to express a device characteristic (e.g. the transfer current) by the emitter dimensions as well as the corresponding set of model parameters. The fundamental assumption of W-L (width and length respectively) scaling is that a parameter (e.g. capacitance) is a linear combination of a bottom-area (A) proportional and a perimeter (P) proportional component with

$$\mathrm{Par} = \mathrm{Par}_A \cdot A + \mathrm{Par}_P \cdot P, \tag{3.1}$$

which corresponds to a simple one dimensional regression equation. This technique is referred to as perimeter-area (P/A) scaling.

Since the (main) current of high-speed HBTs flows in vertical direction, lateral dimensions often-times present the reference for calculation of total device current (I) from current densities ($J = I/A$). Hence the emitter window is taken as a reference for a big part of the scalable model equations embedded in the design kit. Therefore it is important to have an accurate measure at hand.

Scaling based directly on the drawn device dimensions would be the most convenient concept. Yet this is not always possible, keeping in mind the different process steps involved in the manufacturing of SiGe HBTs. In modern devices emitter edge current crowding as well as corner rounding effects (esp. for short devices) is of non-negligible importance. In consequence the actual electrical device size could be quite different for the model parameters representing different physical quantities.

Due to the complexity of SiGe processing the transistor dimensions fixed in the design rules are not representing the real dimensions in silicon. The emitter window opening (using the so called nominal 'drawn dimensions' w_{Ed} and l_{Ed}) is etched laterally under the hard mask. Further mask processing and non-idealities result in a different emitter window opening area A_E on top of the surface for the base epitaxy. Event though more accurate control over critical dimensions becomes available with advancing lithography technology, as device feature sizes become smaller, it is increasingly difficult to accurately control the dimensions of features such as the effective emitter window opening.

In addition, the FSA/SEG technology used for the described process makes use of silicon-nitride inside wall spacers. These spacers are formed to protect the extrinsic base polysilicon and provide the SiGe:C base link separation [62]. The spacer is deposited on top of a TEOS oxide cap that separates the final poly-silicon emitter deposition from the base. A trade-off has to be made between the protection of the TEOS oxide by the spacer nitride and a thin spacer dimension to

increase the effective base-emitter junction area.

To account for these effects the drawn dimensions (e.g. handed over as parameter to the Pcell) are corrected in several steps to arrive at a emitter area that corresponds more to the junction actually found in silicon.

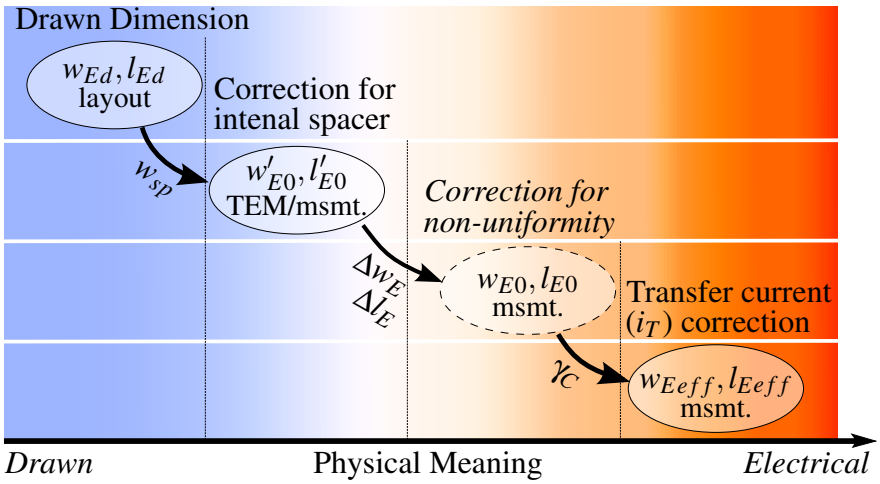


Figure 3.2: Correction of emitter window dimensions defined in layout in order to take into account deviations and physical effects due to silicon processing [*intermediate step of non-uniformity correction optional,* (*default:* $\Delta w_E = \Delta l_E = 0$)]

In a first step the inside spacer is subtracted from the drawn dimensions (A_{Ed}). The value of the spacer width w_{sp} might be obtained either from a TEM imaging analysis or as a by-product of the extraction of the internal base sheet resistance (R_{SBi}, cf. section 4.4.2). The inside spacer value is assumed to be of uniform size for lateral x and y direction (cf. Fig. 3.6).

Other effects influencing the emitter window size are linked to the manufacturing process or lithography respectively. The image patterning of the photoresist is non-ideal. Even though photo-lithography enhancement techniques such as proximity correction are common in modern processes, sharp corner features (such as a rectangular emitter window) will be rounded (corner rounding) due to the limitation of the diffraction from the mask and narrow line ends are usually shortened (line-end pullback) [63, 64]. For example the top view of a 130*nm* lighography process shows that the emitter window opening of a drawn emitter width of $w_{Ed} = 250nm$ structure is significantly reduced to about 220*nm* and emitter window edges do not correspond to their rectangular shape (cf. Fig. 3.3 and 3.4).

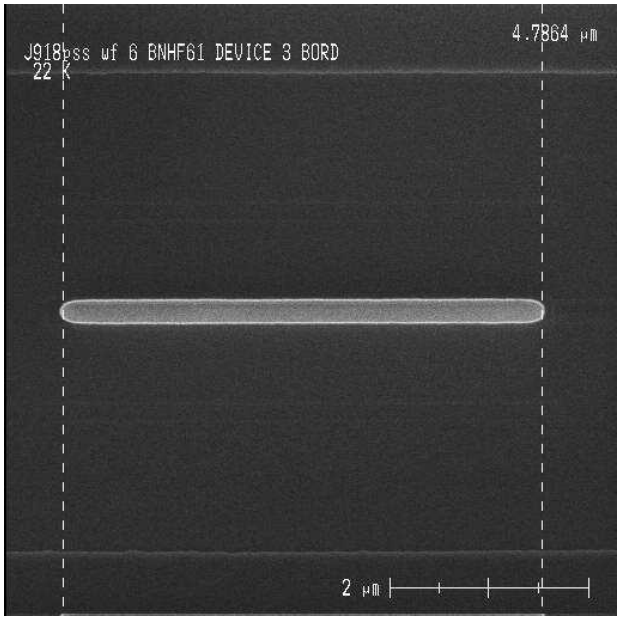


Figure 3.3: Top view of a SiGe HBT device in 130nm BiCMOS technology with a drawn emitter length $l_{E,d}$ of 5μm

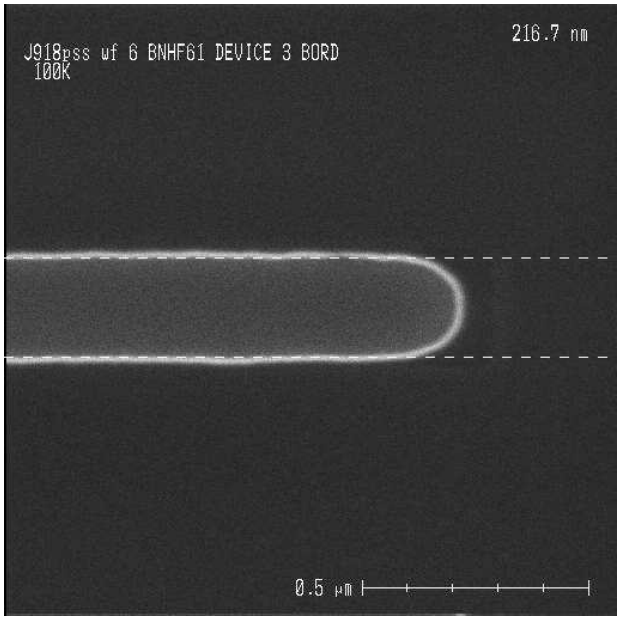


Figure 3.4: Detailed top view with section cut of the emitter window edge with a drawn emitter width $w_{E,d}$ of 0.25μm

Difficulties may arise from the fact that non-idealities become even more pronounced as lithography and device size is reduced. In consequence the actual emitter dimensions w_{E0} and l_{E0} might not be known exactly and in recent technologies the lateral emitter doping profile seems to exhibit non-uniformities across the device sizes. Different wide- and narrow-emitter effects have been discovered leading to deviations from standard geometry scaling. It was found that this phenomenon is linked to a arsenic (As) accumulation in the amorphous E region above the spacer, reducing As diffusion into the mono-Si at the emitter perimeter [65]. The effect is referred to as non-uniform doping or laterally recessed perimeter junction. In order to account for effects encountered in recent investigations of geometry scaling in advanced BiCMOS technologies, more flexibility for device scaling was required [66, 67].

As one of the suggested solutions, an additional correction step using specific electrical parameters for the different lateral directions (Δw_E and Δl_E) is hence introduced[1]. In sum the drawn emitter window is hence corrected in two steps by the direction independent spacer (w_{sp}) and additional electrical offset parameters (cf. Fig. 3.6).

[1]However the additional step for electrical offset parameters is not mandatory, (default: $\Delta w_E = \Delta l_E = 0$)

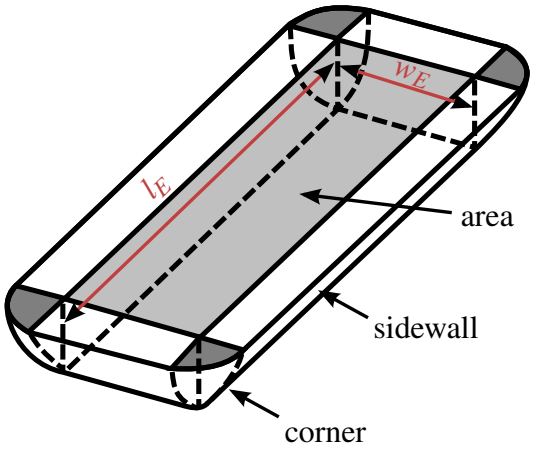


Figure 3.5: Three-dimensioal view of emitter window with corners and and sidewall roundings

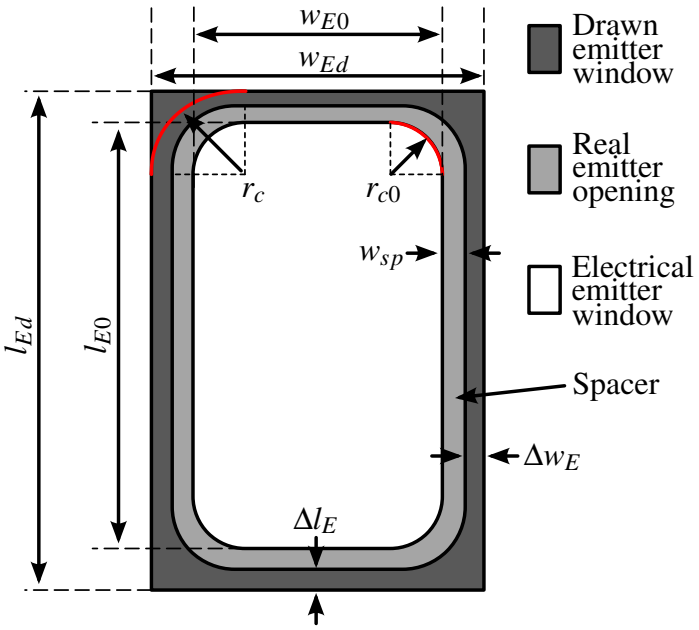


Figure 3.6: Difference of drawn emitter and actual emitter window, index 0 refers to real emitter dimensions between sidewall spacers; w_E and l_E are dimensions of the effective electrical (injecting) emitter

The principal dependence of the scalable model parameters for a vertical transport device is known for a long time (cf. e.g. [68]) and can be expressed using the dimensions given in the device layout as well as process-specific (correction) parameters. The area of each junction within a bipolar transistor can hence be separated into three parts: a rectangular area, combined with cylindrical sidewalls and spherical corners [69] as shown in Fig 3.5. In order to take the corner rounding into account the area of a rounded rectangle has to be determined.

Rectangles that have rounded corners instead of square corners however are a bit more complicated to calculate, taking into account the area of the rounded corners ($A_{corner} = \pi r^2$). The rounded rectangle is characterized by smoothed out corners represented by quarter-circle arcs. Generally the area and perimeter of a rounded rectangle depends on the overall width (w) and length (l) of the shape as well as the radius (r) of curvature at the corners as follows

$$A = l \cdot w - 4 \cdot r^2 + \pi r^2 = l \cdot w - (4 - \pi) r^2. \tag{3.2}$$

Likewise the perimeter formula expressed in terms of w, l, and r is

$$P = 2l + 2w - 8r + 2\pi r = 2(w + l) - (8 - 2\pi) r. \tag{3.3}$$

The Effective Electrical Emitter Window

In addition to non-uniformity effects already mentioned, the nature of fabrication (e.g. dopant out-diffusion during the high-temperature annealing process) imposes that the doping in the BE junction is nonuniform along with a decrease of the current contribution towards the perimeter. Consequently the actual emitter available for vertical current flow (A_{Eeff}) will be significantly smaller than the initially drawn emitter width (A_{Ed}, cf. Fig. 3.6) [70].

A local doping concentration variation under the emitter window is assumed to have a significant impact on the uniformity of the transfer current (i_T). In consequence a model for the collector current when the doping is nonuniform has been presented [71]. The assumed uniform box-shaped junction hence needs to be corrected electrically in a last step (cf. Fig. 3.2).

For the HICUM model the effect is accounted for by a correction of the emitter window area through the current spreading factor γ_C to transform the emitter window area into an effective electrical dimension. The effective emitter area (A_{Eeff}) that forms the junction with the intrinsic base (junction area disposed in the emitter window, separated from the extrinsic base spacers) is hence different from the previously calculated actual emitter window A_{E0}.

The effective rectangular electrical emitter window is defined by scaling the collector current using the two current components as

$$\gamma_C = I_{CP}/I_{CA}, \tag{3.4}$$

where I_{CP} is the perimetric contribution and I_{CA} represents the bottom-area related part.

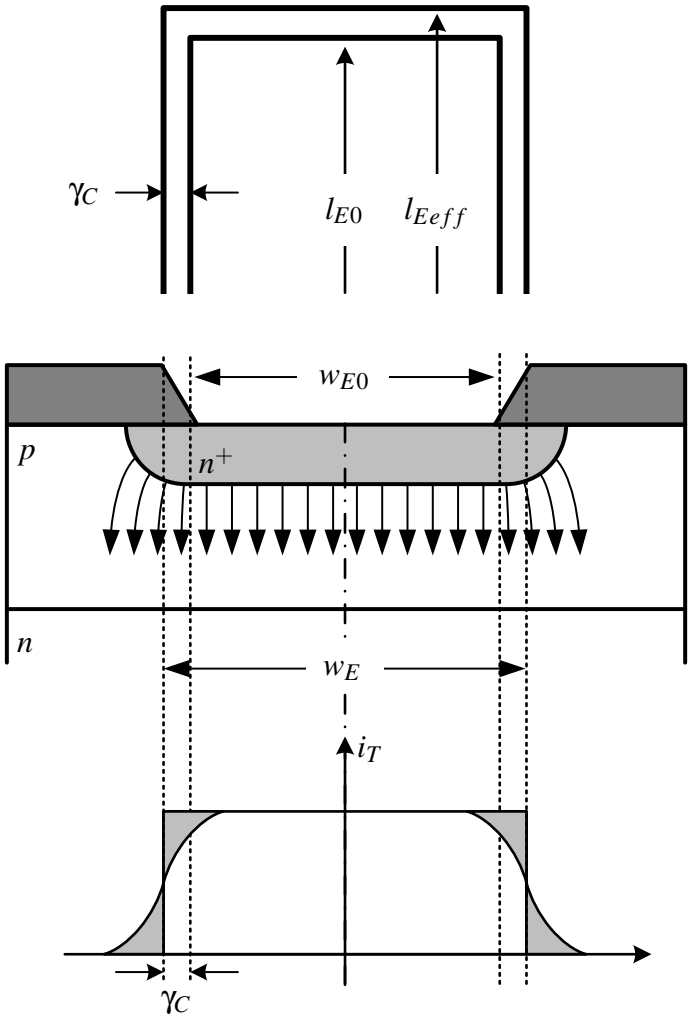


Figure 3.7: Difference of the effective electrical emitter area A_{Eeff} and the actual emitter window opening area A_{E0}

To arrive at a unique representation of the transfer current it is assumed that i_T (composed of surfasic-, perimetric- and corner-components) flows through an extended internal transistor

$$I_T = I_{TA} + I_{TP} + 4I_{TC} = \underbrace{J_{TA}A_{E0}}_{\text{Area}} + \underbrace{J_{TP}P_{E0}}_{\text{Perimeter}} + \underbrace{4I_{TC}}_{\text{Corner}} = J_{TA}A_{Eeff}, \tag{3.5}$$

where the effective emitter area A_{Eeff} is defined using a current spreading factor

$$A_{Eeff} = w_{Eeff} \cdot l_{Eeff} = (w_{E0} + 2 \cdot \gamma_C) \cdot (l_{E0} + 2 \cdot \gamma_C). \tag{3.6}$$

Especially for small devices further improvements in scalability may be achieved taking into

account three-dimensional effects through a modified corner radius r_{ceff} giving

$$A_{Eeff} = w_{Eeff} \cdot l_{Eeff} - (4-\pi) \cdot r_{ceff}^2 \quad \text{Effective emitter area with } r_{ceff} = r_c \cdot \gamma_C \quad (3.7)$$

$$P_{Eeff} = 2 \cdot (w_{Eeff} + l_{Eeff}) - 2 \cdot (4-\pi) \cdot r_{ceff} \quad \text{Effective emitter perimeter.} \quad (3.8)$$

By merging the internal and peripheral transistor into single transistor, this concept allows to model the peripheral transfer current without additional component or current source respectively.

In addition to the collector current this concept of geometry modification is used for the avalanche breakdown modeling as well. The corresponding geometry model parameter in the library is G_{AMAVL}. Likewise the effective area and perimeter taken for the base-collector junction (avalanche effect) is calculated

$$A_{avl} = (w_{E0} + 2 \cdot G_{AMAVL}) \cdot (l_{E0} + 2 \cdot G_{AMAVL}) - (4-\pi) \cdot (r_0 + G_{AMAVL}))^2, \quad (3.9)$$

$$P_{avl} = 2 \cdot (w_{E0} + l_{E0} + 4 \cdot G_{AMAVL}) - 2 \cdot (4-\pi) \cdot (r_0 + G_{AMAVL}). \quad (3.10)$$

Summary of Substantial Emitter Dimensions

As mentioned before the differentiation between the available geometries and configurations is done by individual specification of the emitter dimensions (with drawn dimensions w_{Ed} and l_{Ed}), and the number of emitter, base, and collector fingers. In order to properly reference the different dimensions used in the calculations behind the geometry scalable physics-based model, the most important dimensions, symbols and abbreviations for emitter-size effects are summarized below. The indication of the three indexes ($_{Ed}$, $_{E0}$ and $_{Eeff}$) has the following geometrical interpretation (cf. Fig. 3.8)

- A_{Ed}, P_{Ed}: drawn emitter window area and perimeter from mask layout
- w_{Ed}, l_{Ed}: drawn emitter width and length
- A_{E0}, P_{E0}: (electrical) emitter window area and perimeter[1]
- w_{E0}, l_{E0}: (electrical) emitter window width and length[1]
- γ_C: ratio of periphery to area specific collector current [2]
- A_{Eeff}, P_{Eeff}: effective emitter area and perimeter specific to collector current
- w_{Eeff}, l_{Eeff}: effective emitter width and length specific to collector current

[1] taking into account additional spacer width and doping non-uniformity

[2] equal to emitter width increase due to periphery injection (e.g. $w_{Eeff} = w_{E0} + 2\gamma_C$)

3.2 Equations for Geometry-Scaling in Advanced Vertical NPN Bipolar Devices

Since in the HICUM EC each spatial region of the transistor is represented by a corresponding equivalent circuit element one may define different scaling equations per component.

Given the relation of model parameter and dimension as well as device configuration the development of scaling equations is difficult. The idea of scalable modeling is to have a device layout corresponding model that covers all usage scenarios (bias) and is fully scalable (geometry). Therefore a number of parameters are integrated using scaling equations. Yet undesired side-effects as for example line-end foreshortening along the vertical (long) axis[1] and corner-rounding effects in lithography have a significant impact on small devices and principal process scalability as well as existing scaling rules have to be verified for advancing process technologies.

The development of scaling equations starts with the extraction of local parameter sets for individual devices, where-after the dependence of the extracted parameter on the device length may be inspected in order to verify their validity or find adequate new scaling laws. In the process of parameter determination, a set of compact model parameters is then extracted based on a variety of TCAD simulated or measured device characteristics. In order to correctly take into account all inherent scaling effects, parameters are extracted for a number of multi-length and multi-width devices.

3.2.1 Definition of Global Geometry Parameters

In order to verify the transistor dimensions, imaging techniques such as SEM and TEM may provide cross-sections and top views to obtain the spatial dimensions in the most important transistor configurations. For this purpose an actual transistor structure can be characterized by means of imaging technologies to double-check the dimensions defined in the design rules after manufacturing in silicon. Especially for dimensions that cannot be extracted otherwise from electrical measurements this is an important method of parameter estimation and extraction.

The definitions of the relevant transistor dimensions used to calculate area- and perimeter-specific model parameters by means of the scaling equations defined in the model library are given in Fig. 3.8. A (simplified) schematic cross-section (cut along the length axis) provides lateral device dimensions parallel to the width of the transistor w_E. The corresponding top view with matching device dimensions is shown thereunder.

[1] the representation of a line or an object shorter than the actual length; in optical lithography for logic circuits, the issue of structure end foreshortening becomes more and more relevant with increasingly small dimension making photoresist line prints shorter than the corresponding line [72]

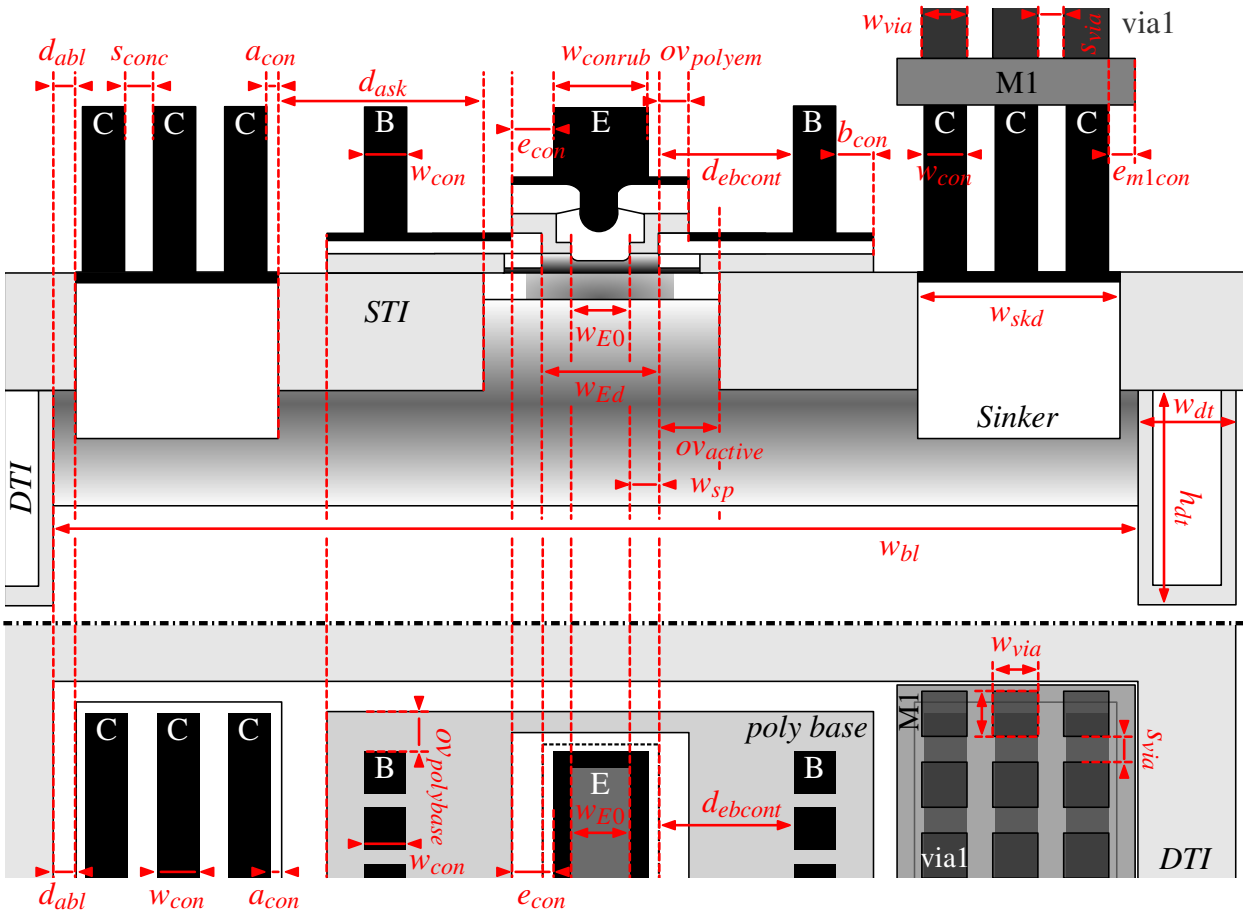


Figure 3.8: Cross section and top view of vertical SiGe HBT device in symmetrical CBEBC configuration with definition of global geometry parameters [*stripe contacts for collector and emitter, polysilicon base connected through via contacts, interconnects of first metal layer (*M1*) and first via level (*via1*) for collector shown*]

Table 3.2 summarizes the definitions of the relevant transistor dimensions and gives the corresponding textual definitions for spatial dimensions that apply to all devices (hence defined only once) in a given process.

Abbr.	Parameter Description	Abbr.	Parameter Description
w_{min}	minimum emitter width	$w_{polyemsp}$	*poly-emitter spacer width*, 0
w_{max}	maximum emitter width	s_{em}	space between two emitters
l_{min}	minimum emitter length	w_{con}	contact width
l_{max}	maximum emitter length	$w_{conruban}$	contact stripe width
w_{sp}	internal spacer width	s_{conb}	space between 2 base contacts
dw_e	difference between real and electrical emitter width*	s_{conc}	space between 2 collector contacts
dl_e	difference between real and electrical emitter length*	e_{con}	distance of emitter contact to poly emitter
dw_b	difference between drawn and real base	b_{con}	overlap of poly base exceeding the base contact
dw_c	difference between drawn and real collector	a_{con}	enclosure of collector contact by active
ov_{polyem}	distance between drawn emitter and poly emitter	d_{ask}	distance active to sinker
$ov_{polybase}$	distance of poly base overlap	d_{abl}	distance active (sinker) to buried layer
w_{skd}	collector sinker width	w_{dt}	deep trench width
ov_{active}	distance of active opening to drawn emitter	h_{dt}	deep trench depth
d_{ebcont}	distance of drawn emitter to base contact	e_{m1con}	overlap of first metal level
w_{via}	via width	s_{via}	space between 2 vias

Table 3.2: Definition of global geometry parameters that are constant for a given technology and independent of device geometry [**workaround in order to take non-ideality and non-uniform doping into account; not visible in cross section*]

3.2.2 Layout Independent Technology Parameters

In addition a second table (cf. Tab. 3.3) gives the values of specific process informations that scale with device geometry. These are for example sheet resistances obtained in extraction procedures, or doping and resistivity information obtained from the process specifications.

Param.	Description	Param.	Description	Param.	Description
$c1\_rsu$	scalable coefficient for RSU computation	r_{kci0}	resistance of the epitaxial-layer of the intrinsic collector	r_{sbi0}	zero bias internal base sheet resistance
ro_{sub}	specific substrate resistivity	lat	current spreading factor length	r_{sbx}	polybase link sheet resistance under internal spacer
n_{sub}	substrate doping	r_{ssi}	silicided polybase sheet resistance	r_{spo}	unsilicided polybase link sheet resistance
d_{bl}	buried layer geometry correction	r_{sbl}	buried layer sheet resistance	r_{skl}	lineic sinker resistance
r_{con}	back end contact resistance	r_{via}	back end via resistance	r_{ke}	specific areal emitter resistance

Table 3.3: Definition of parameters linked to the process that are used to recalculate parameters from device geometry using scaling equations

Another special part is the set of parameters used for back-end capacitance calculation and oxide capacitance calculation (c_{eoxl} and c_{coxl}). The back-end contributions (base-emitter capacitance *cbeconl_cbe* for example) is obtained from numerical electro-magnetic (EM) field-simulations.

3.2.3 Layout Dependent Parameters Related to Transistor Areas

In general there is a number of current, capacitance, transit time and other parameters that scale with the device geometry. For these parameters the general device layout of the respective junction or area has to be calculated. The geometry-specific parameter calculation is based on the geometry-independent parameters as well as the instance parameters provided at PCell instantiation.

Since several components of the device models are described and extracted from parameters (e.g. current) as a function of (electrical) perimeter to area ratio it is useful to define the corresponding transistor regions wrt. each junction. Hence this section summarizes the calculation of dimensions relative to each region and the linked scaling equations used in the template will be described.

Base-Emitter (BE) Junction Parameters

The most important part of the transistor, when it comes to scalable modeling, is the BE junction. The layout was already briefly discussed and special geometrical correction for the transfer current has to be taken into account (γ_C).

The calculation of the emitter area starts with the drawn dimensions (instance parameters provided as input to the model library). The corner rounding radius calculation is based on the information of the minimum allowed feature size of the emitter defined in the global parameter w_{Emin}.

The area calculation is the same for any configuration meaning that even for multi-emitter devices in a first step the area of one individual emitter finger is calculated.

$$A_{Ed} = w_{Ed} \cdot l_{Ed} \qquad \text{Drawn emitter area} \qquad (3.11)$$

$$P_{Ed} = 2 \cdot (w_{Ed} + l_{Ed}) \qquad \text{Drawn emitter perimeter} \qquad (3.12)$$

$$r_d = w_{Emin}/2 \qquad \text{Drawn emitter corner rounding radius} \qquad (3.13)$$

Correction of the spacer and possible non-uniformity in the process gives the actual emitter area and perimeter.

$$w_{E0} = w_{Ed} - (w_{sp} + dw_e) \qquad \text{Actual emitter opening width} \qquad (3.14)$$

$$l_{E0} = l_{Ed} - (w_{sp} + dl_e) \qquad \text{Actual emitter opening length} \qquad (3.15)$$

$$r_0 = r_d - \left(\frac{w_{sp}}{2} + \frac{dw_e}{2}\right) \qquad \text{Actual emitter corner rounding radius} \qquad (3.16)$$

$$A_{E0} = w_{E0} \cdot l_{E0} - (4 - \pi) \cdot r_0^2 \qquad \text{Actual emitter area} \qquad (3.17)$$

$$P_{E0} = 2 \cdot (w_{E0} + l_{E0} - 2 \cdot (4 - \pi) \cdot r_0) \qquad \text{Actual emitter perimeter} \qquad (3.18)$$

Correction for the transfer current gives the effective electrical area.

$$A_{Eeff} = w_{Eeff} \cdot l_{Eeff} = (w_{E0} + 2 \cdot \gamma_C) \cdot (l_{E0} + 2 \cdot \gamma_C) \qquad \text{Eff. el. emitter area} \qquad (3.19)$$

$$P_{Eeff} = 2 \cdot (w_{Eeff} + l_{Eeff}) = (w_{E0} + 2 \cdot \gamma_C) \cdot (l_{E0} + 2 \cdot \gamma_C) \qquad \text{Eff. el. emitter perimeter} \qquad (3.20)$$

Taking into account the (optional) correction factor d_{cbe} the final BE dimensions are calculated

$$A_{cbe} = (w_{E0} + 2 \cdot d_{cbe}) \cdot (l_{E0} + 2 \cdot d_{cbe}) - (4 - \pi) \cdot (r_0 + d_{cbe})^2 \qquad \text{BE junction area} \qquad (3.21)$$

$$P_{cbe} = 2 \cdot (w_{E0} + l_{E0} + 4 \cdot d_{cbe}) - 2 \cdot (4 - \pi) \cdot (r_0 + d_{cbe}) \qquad \text{BE junction perimeter} \qquad (3.22)$$

In a next step the remaining areas related to the BE junction such as the drawn polyemitter are

determined based on drawn emitter dimensions.

$$w_{polyem} = w_{Ed} + 2 \cdot ov_{polyem} \quad \text{Polyemitter width} \quad (3.23)$$

$$l_{polyem} = l_{Ed} + 2 \cdot ov_{polyem} \quad \text{Polyemitter length} \quad (3.24)$$

$$A_{Epolyem} = w_{polyem} \cdot l_{polyem} - (4-\pi) \cdot r_0^2 \quad \text{Polyemitter area} \quad (3.25)$$

$$P_{Epolyem} = 2 \cdot [w_{polyem} + l_{polyem}] - 2 \cdot (4-\pi) \cdot r_0 \quad \text{Polyemitter perimeter} \quad (3.26)$$

Base-Collector (BC) Junction Geometry Parameters

In principle, the BC junction as another important portion of the inner transistor structure follows the same rules for calculation. However the junction is considered to be situated in the whole transistor width between both bounding STI structures. The distance between the outer L spacers and the STI denominated ov_{active} is hence the basis for the calculation of the BC junction area (likewise based on the drawn emitter dimensions).

$$w_{bd} = w_{Ed} + 2 \cdot ov_{active} \quad \text{Drawn active/base width} \quad (3.27)$$

$$l_{bd} = l_{Ed} + 2 \cdot ov_{active} \quad \text{Drawn active/base length} \quad (3.28)$$

$$A_{bd} = w_{bd} \cdot l_{bd} \quad \text{Drawn active/base area} \quad (3.29)$$

$$P_{bd} = 2 \cdot (w_{bd} + l_{bd}) \quad \text{Drawn active/base perimeter} \quad (3.30)$$

However given the difficulties with increasingly small feature size the correction factor dw_B has been introduced to obtain a more realistic base-collector junction width and length.

$$w_{b0} = w_{bd} - dw_b \quad \text{Real base/collector width} \quad (3.31)$$

$$l_{b0} = l_{bd} - dw_b \quad \text{Real base/collector length} \quad (3.32)$$

$$A_{b0} = w_{b0} \cdot l_{b0} \quad \text{Real base/collector area} \quad (3.33)$$

$$P_{b0} = 2 \cdot (w_{b0} + l_{b0}) \quad \text{Real base/collector perimeter} \quad (3.34)$$

In case there is a need to correct the BC junction area and perimeter for a difference between drawn emitter and effective BC junction dimensions, the value of d_{cbc} may be adjusted. However this factor has not been needed in recent technologies and therefore is not referenced in Tab. 3.2. Using the default value of 0 the CB area and perimeter correspond to the dimensions of the actual BE junction (cf. eqn. 3.17 and 3.18).

$$A_{cbc} = (w_{E0} + 2 \cdot d_{cbc}) \cdot (l_{E0} + 2 \cdot d_{cbc}) - (4-\pi) \cdot (r_0 + d_{cbc})^2 \quad \text{BC area} \quad (3.35)$$

$$P_{cbc} = 2 \cdot (w_{E0} + 2 \cdot d_{cbc} + l_{E0} + 2 \cdot d_{cbc}) - 2 \cdot (4-\pi) \cdot (r_0 + dcbc) \quad \text{BC perimeter} \quad (3.36)$$

A last parameter strongly related to the BC junction is the area and perimeter related to the avalanche breakdown. In accordance with the concept of a unified transfer current, an avalanche correction factor γ_{avl} is defined.

$$A_{avl} = (w_{E0} + 2\cdot\gamma_{avl})\cdot(l_{E0} + 2\cdot\gamma_{avl}) - (4-\pi)\cdot(r_o - \gamma_{avl})^2 \quad \text{Eff. el. aval. area} \quad (3.37)$$

$$P_{avl} = 2\cdot[(w_{E0} + 2\cdot\gamma_{avl}) + (l_{E0} + 2\cdot\gamma_{avl})] - 2\cdot(4-\pi)\cdot(r_o - \gamma_{avl}) \quad \text{Eff. el. aval. perim.} \quad (3.38)$$

Collector Geometry Parameters

For the calculation of the collector periphery as well as the junction between collector and substrate a more complex distinction of cases is required. Depending on the transistor configuration the distance between the bounding DTI structures that limit the buried layer varies. Flags based on the number of contacts (nbe, nbc, nbe) are used in order to calculate the effective dimensions.

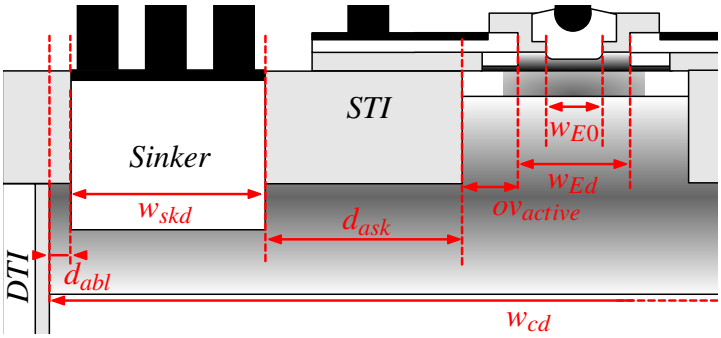


Figure 3.9: Cross section view of collector link region for calculation of the buried layer area A_{c0} from given technology information

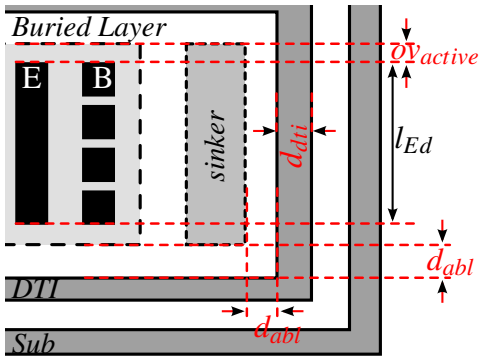


Figure 3.10: Top view of collector link region for calculation of the buried layer area A_{c0}

First the drawn buried layer dimensions are calculated. Note that due to the usually large spatial dimension of the entire buried layer the effect of corner rounding is considered to be negligible. The flag F_C is used in subsequent calculations. It is zero for a single C contact, otherwise one.

$$w_{cd} = 2\cdot d_{abl} + w_{skd} + d_{ask} + w_{bd} + \ldots$$
$$\ldots [N_{BE} - 1]\cdot[s_{em} + w_{Ed}] + F_C\cdot[d_{ask} + w_{skd}] \quad \text{Drawn buried layer width} \quad (3.39)$$

$$l_{cd} = l_{Bd} + 2\cdot d_{abl} \quad \text{Drawn buried layer length} \quad (3.40)$$

$$A_{cd} = w_{Cd}\cdot l_{cd} \quad \text{Drawn buried layer area} \quad (3.41)$$

$$P_{cd} = 2\cdot(w_{cd} + l_{cd}) \quad \text{Drawn buried layer perimeter} \quad (3.42)$$

In analogy to the BC junction a geometry correction offset dw_b may be used to obtain the real collector dimensions from the drawn ones. The related length of the sinker (SK) based on the drawn emitter window length is defined along side

$$w_{c0} = w_{cd} - dw_b \quad \text{Real buried layer width} \tag{3.43}$$

$$l_{c0} = l_{cd} - dw_b \quad \text{Real buried layer length} \tag{3.44}$$

$$l_{skd} = l_{Ed} + 2 \cdot ov_{active} \quad \text{Drawn sinker length} \tag{3.45}$$

During the extraction of the external collector resistance (R_{Cx}) an additional parameter d_{bl} is determined along with the sheet resistances. This value is used to correct the real buried layer length (l_{c0}) by an electrical offset for an effective buried layer length reduction due to lateral current spreading effects between collector and emitter. However this value has no impact on other scalable model parameter of the substrate network.

$$l_{cf} = l_{c0} - d_{bl} \quad \text{Effective buried layer length for } R_{Cx} \tag{3.46}$$

$$A_{c0} = w_{c0} \cdot l_{cf} \quad \text{Real buried layer area} \tag{3.47}$$

$$P_{c0} = 2 \cdot (w_{c0} + l_{cf}) \quad \text{Real buried layer perimeter} \tag{3.48}$$

3.3 Scaling of Transistor Equivalent Circuit Elements

The vertical HBT device can principally be understood as one-dimensional (1-D), vertical device under the emitter, yet all the periphery (related to lateral scaling) affect the device characteristics and parameters. Within a very general cross section the principal elements related to the individual space charge regions can be defined (cf. Fig. 3.11).

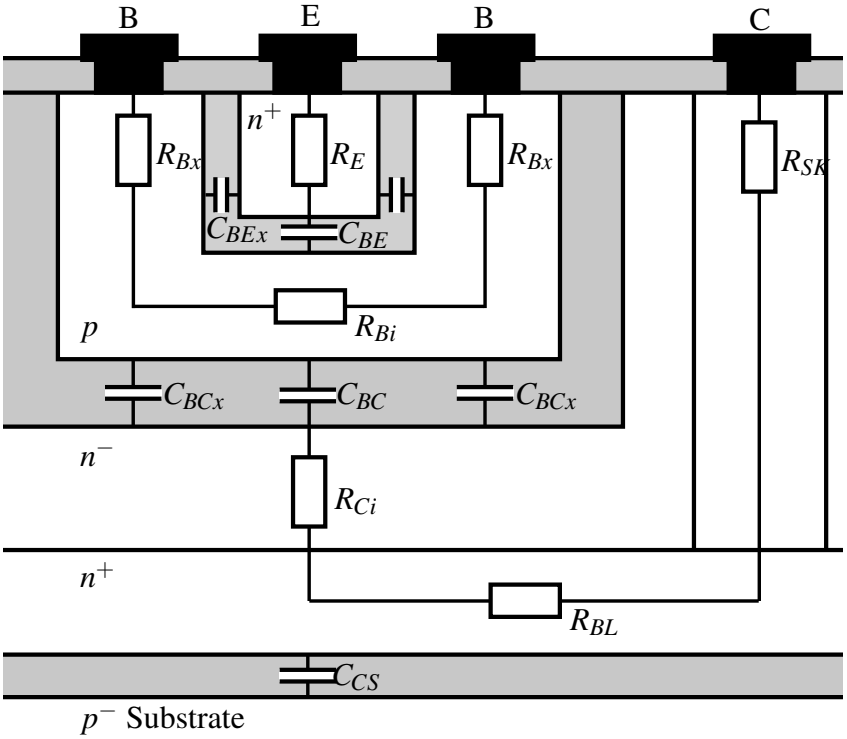


Figure 3.11: Equivalent circuit elements of the BJT related to junctions (*SCRs shown as gray regions*) including extrinsic (parasitic) elements (indexed $_x$)

All parameters contributing to the device model need to be analyzed for their dependence on device layout. Due to the strong impact on device characteristics, one of the first steps in determining the equivalent circuit elements for HBT modeling is the accurate extraction of extrinsic elements such as link resistances and capacitance values. The (constant) pad inductances, and contact resistances are relatively small but may in sum have a significant influence on the extraction of the intrinsic parts as well.

By analysis of the progression of the total parameter value in the P/A approach, it is possible to separate individual contributions: A general scaling approach for device modeling is to divide a given parameter into its sidewall and bottom fractions. For most of the scaling equations there are hence two model parameters defined. The first being defined per length along the perimeter of the respective junction and the second one defined per unit area dedicated to the bottom surface of

the respective region. In turn, a basic scaling equation contains two variables (instance variables l and w defining the surface and perimeter) and two parameter values (representing the individual contributions to the total model variable). Following eqn. (3.1), subsequent calculations use the area scaling parameter, indicated by Par_A and the perimeter scaling parameter through Par_P. In case of a unitary parameter that is exclusively scaled by a specific (area-)ratio, the designation P_u is used.

3.3.1 BE Capacitance

Strictly following the P/A separation approach, capacitances are generally scaled with the junction area and perimeter as shown in the equation below as well as the corresponding schematic (cf. Fig. 3.12).

$$C_j = C_{jA} + C_{jP} \tag{3.49}$$

A special case are oxide capacitances between the external transistor regions. These are modeled as dependent exclusively on the perimeter relative to the respective region as well as the number of parallel emitter fingers

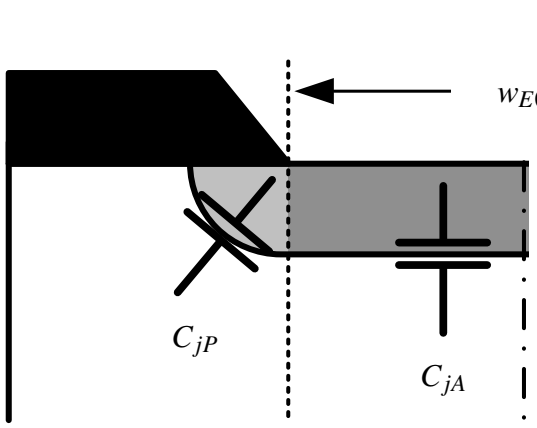


Figure 3.12: Capacitance partitioning in perimeter and area for junction capacitances of the intrinsic device (C_j)

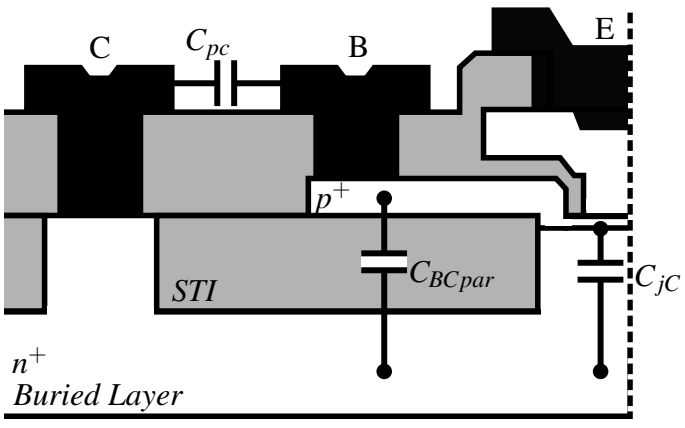


Figure 3.13: Cross section of BC region in vertical SiGe HBT device with related capacitance contributions of junction (C_{jC}) and oxide (C_{BCpar}) as well as back-end (C_{pc})

It has to be considered that the intrinsic and extrinsic junctions do not have the same characteristics or junction parameters respectively. Hence a capacitance splitting approach is used, calling for separate equations with their corresponding junction parameter set (C_{j0}, V_D, z) for the intrinsic and extrinsic capacitances. It is understood that the zero-bias value of the junction depletion

capacitance C_{JE0} scales linearly with the base-emitter junction area A_E:

$$C_{JEI0} = nbe \cdot c_{jea} \cdot A_{cbe} \quad \text{intrinsic zero-bias BE capacitance} \tag{3.50}$$

$$C_{JEP0} = nbe \cdot c_{jep} \cdot P_{cbe} \quad \text{extrinsic zero-bias BE capacitance} \tag{3.51}$$

$$H_{JEI} = h_{jeiu} \cdot \frac{A_{Eeff}}{A_{cbe}} \quad \text{hole charge weighting factor} \tag{3.52}$$

In addition, the emitter-base isolation capacitance (C_{BEpar}) is taken into account. This is modeled as a partitioned capacitance between the base perimeter (perimeter connection to the base node B*, $C_{BEpar,2}$) and external base node which are separated by the external base resistance. This concept makes the model more flexible and the distributed parasitic isolation capacitance may for example include metallization etc. Both parasitic oxide capacitances (BE and BC junction) are considered to scale linearly with the perimeter of their respective areas.

$$C_{BEpar} = nbe \cdot c_{eoxl} \cdot P_{cbe} \quad \text{parasitic BE oxide capacitance} \tag{3.53}$$

The linked parameter f_{bepar} (defined as the ratio of the inner to the total parasitic capacitance) is used for the proper capacitance partitioning. It is a unitary parameter without need for further scaling with geometry.

$$f_{BEpar} = \frac{C_{BEpar,2}}{C_{BEpar}} \quad \text{BE partitioning factor} \tag{3.54}$$

3.3.2 BC Capacitance

Likewise the intrinsic and extrinsic BC capacitances are considered to be proportional to an effective junction area and an effective perimeter, respectively. On the basis of the BC area and perimeter the model parameters are calculated from:

$$C_{JCI0} = nbe \cdot c_{jca} \cdot A_{cbc} \quad \text{intrinsic zero-bias BC capacitance} \tag{3.55}$$

$$C_{JCP0} = nbe \cdot c_{jcp} \cdot P_{cbc} \quad \text{extrinsic zero-bias BC capacitance} \tag{3.56}$$

$$H_{JCI} = h_{jciu} \cdot \frac{A_{Eeff}}{A_{cbc}} \quad \text{hole charge weighting factor} \tag{3.57}$$

$$C_{BCpar} = nbe \cdot c_{coxl} \cdot P_{bd} \quad \text{parasitic BC oxide capacitance} \tag{3.58}$$

$$f_{BCpar} = \frac{C_{BCx,2}}{C_{BCx}} \quad \text{parasitic BC partitioning factor} \tag{3.59}$$

Here the base-collector oxide capacitance C_{BCpar} is likewise added with the linked parameter

f_{BCpar} used without any scaling. However the total external BC capacitance is considered as a bias-dependent external depletion capacitance C_{jCx} and a bias-independent parasitic capacitance resulting from the STI and metallization.

The overlap capacitance is bias independent and can be extracted from C-V curves by de-embedding the pad capacitance and separating the bias dependent capacitance. Periphery contributions are usually calculated using TCAD simulations with a field-simulation approach since determining the parasitic elements by test structures requires additional extraction effort, consumes silicon area and reliability is limited through accuracy constraints of the measurement equipment.

The used fitting parameters a and b are determined individually for each technology. Given values are for the BiCMOS55 technology.

3.3.3 Substrate Network

Accurate junction capacitance modeling is required for accurate substrate crosstalk calculation between circuit components. The ohmic resistance and the permittivity of the substrate become increasingly important towards high frequencies due to their influence on the isolation between the substrate-collector SCR and the substrate contact.

CS junction capacitance

In a vertical HBT the entire buried layer structure rests atop a lightly doped p-type substrate to which electrical contact is provided outside the DTI through a ring-shaped $p^{++} - substrate$ contact at the surface isolating the NPN device from adjacent transistors.

In the HICUM EC the connection between the substrate contact (S) and the internal collector (C') is provided by a dedicated substrate network. The S contact is either far away from the CS junction or separated by a DTI. The network is hence comprised of a link resistance r_{Su} in series to the CS depletion capacitance (C_{CS}) caused by a high substrate resistivity ρ_{SU}. In addition, the high bulk permittivity (ε_{Si}) leads to a capacitance C_{SU} in parallel (becoming dominant towards high frequency). Physically the contributions can be scaled using a bottom surface A_{CS} and a periphery RC network.

From classical capacitor theory it is known that the capacitance depends linearly on material parameters ε, the effective area of the corresponding layer as well as the thickness d

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d}. \qquad \text{capacitance of a plate capacitor} \tag{3.60}$$

The depletion-layer capacitor of the CS junction is determined under the assumptions that the doping concentration does not change over the junction and the face area of the buried layer is entirely between the limits of the buried layer. A complicated sidewall calculation as for the case

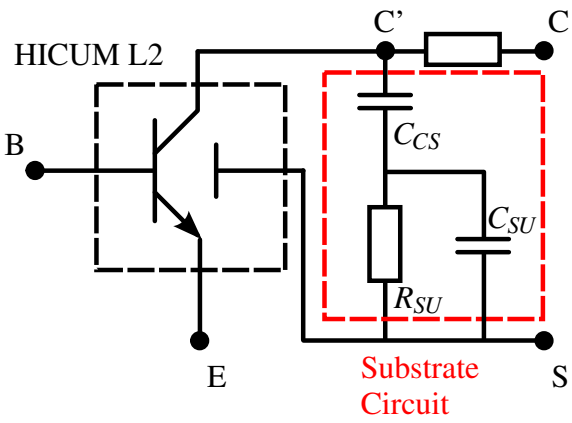


Figure 3.14: Substrate network equivalent circuit used in HICUM compact model

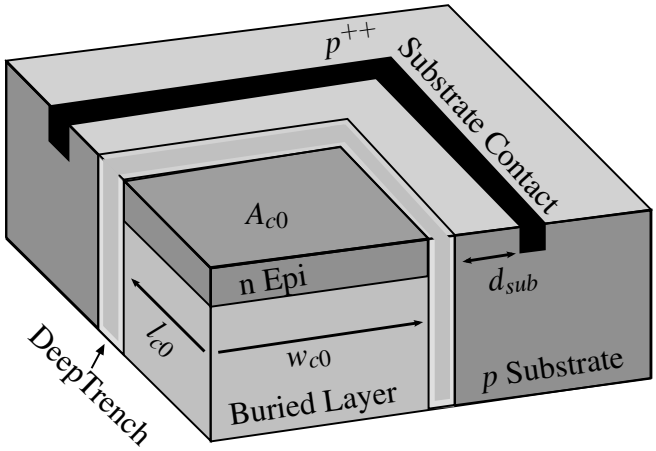


Figure 3.15: Three dimensional cut cross section of the substrate and buried layer bounded by the DTI with external substrate ring connection (p^{++})

of a diffused region is hence not required.

The junction capacitance C_{SU} is hence simply proportional to the region enclosed by the DTI

$$C_{SU} = C_{sua} \cdot A_{c0} + C_{sup} \cdot P_{c0} \qquad \text{substrate capacitance} \qquad (3.61)$$

The buried layer area is calculated using the dimensions of the inner transistor, the drawn buried layer width w_{cd} and the device configuration using the following constant external dimensions:

- w_{dt}: deep trench width
- d_{ask}: distance between active and sinker
- d_{abl}: distance of active opening (sinker sided) and DTI

The real buried layer area is obtained by correction through the difference between drawn and real collector dw_c:

$$w_{c0} = w_{cd} - dw_c \qquad \text{real buried layer width} \qquad (3.62)$$

$$l_{c0} = l_{cd} - dw_c \qquad \text{real buried layer length} \qquad (3.63)$$

$$A_{c0} = w_{c0} \cdot l_{c0} \qquad \text{real buried layer area} \qquad (3.64)$$

$$P_{c0} = 2 \cdot (w_{c0} + l_{c0}) \qquad \text{real buried layer perimeter} \qquad (3.65)$$

The capacitance between collector and substrate is scaled using the real buried layer dimen-

sions A_{c0} and P_{c0}:

$$C_{jsa0} = c_{jsa} \cdot A_{c0} \quad \text{intrinsic zero-bias CS capacitance} \quad (3.66)$$

$$C_{jsp0} = c_{jsp} \cdot P_{c0} \quad \text{extrinsic zero-bias CS capacitance} \quad (3.67)$$

$$C_{JS0} = C_{jsa0} + C_{jsp0} \quad \text{total CS capacitance} \quad (3.68)$$

Substrate Coupling Capacitance and Resistance

The resistance and capacitance of the substrate are scaled by the number of transistors. Only one parameter (e.g. the resistive component) needs to be scaled according to the emitter dimensions since both quantities are linked through the material constants of the silicon substrate ε_{Si}.

$$R_{SU} = nbt \cdot \frac{\rho_{sub,si}}{C_{1rsu}\left(\frac{l_{Ed}}{w_{Ed}} + \frac{w_{Ed}}{l_{Ed}}\right)} \quad \text{substrate resistance} \quad (3.69)$$

$$C_{SU} = \frac{\varepsilon_{Si}}{R_{SU}} \quad \text{substrate capacitance} \quad (3.70)$$

Given these relations one can also estimate some simple dependences. A higher substrate doping in order to reduce series resistance for example increases parasitic capacitances.

3.3.4 Emitter Resistance R_E

The scaling of physical resistances that account for voltage drops incurred by currents flowing in the charge neutral regions of the base, emitter, and collector, respectively is strongly related to the architecture of the device.

As a first contribution the emitter resistance exhibits a very simple scaling rule being directly dependent on the inverse actual emitter area A_{E0} as well as the number of emitter fingers

$$R_E = R_{ke} \cdot \frac{1}{nbe \cdot A_{E0}} + R_{E,con} \quad \text{Emitter resistance} \quad (3.71)$$

$$R_{E,con} = \frac{R_{kvia}}{nbe \cdot A_{via}} \quad \text{Emitter contact related resistance} \quad (3.72)$$

3.3.5 Internal Base Resistance R_{Bi}

Opposed to the link resistances the internal base resistance R_{Bi} is divided into a geometry and a current dependent part. The current dependent modeling is a built-in feature of the HICUM model. However the dependence on device geometry is to be modeled externally and implemented

following the concept presented in [73].

$$w_{pm} = w_p + w_{ss} \cdot \frac{r_{spo}}{r_{ssi}} \tag{3.73}$$

$$y = w_{E0} + \frac{l_{E0}}{2} + 3 \cdot w_l \tag{3.74}$$

$$x = \sqrt{\frac{r_{spo}}{r_{sbx}}} \cdot \frac{y}{w_{pm} + w_l} \tag{3.75}$$

$$f_{i1} = 0.23 \cdot \left((x - 0.25) + \sqrt{[x - 0.25]^2 + 10^{-4}} \right) \tag{3.76}$$

$$f_i = \min[f_{i1}, 1] \tag{3.77}$$

$$g = \frac{1 + 3f_i}{12} - \left(\frac{1}{12} - \frac{1}{28.6} \right) \cdot w_{E0} \cdot \frac{1 - f_i}{l_{E0}} \tag{3.78}$$

$$r_{bi0} = r_{sbp} \cdot \frac{w_{E0} \cdot \frac{g}{l_{E0}}}{nbe} \quad \text{with } r_{sbp} = r_{sbi0}. \tag{3.79}$$

Within the HICUM model itself the intrinsic base resistance dependence on base bias is calculated by a ratio of the base charge Q_{p0}/Q_p obtained from evaluation of the GICCR.

3.3.6 Extrinsic Base Link Resistance R_{Bx}

The extrinsic base link resistance takes into account the three contributions of the silicided region and both link regions surrounding the emitter and under the emitter spacer:

- R_{sil}: silicided region with process-defined sheet resistance r_{ssi} (typically around $15\Omega/\square$)
- R_{link}: poly-silicon to mono-silicon interface region in emitter periphery
- R_{sbx}: link region surrounding the emitter window under the spacer

Some additional dimensions need to be calculated using the global values defined for the overlap distances of the poly-emitter and -base.

$$w_l = \frac{w_{sp}}{2} \tag{3.80}$$

$$w_p = ov_{polyem} \tag{3.81}$$

$$w_{ls} = \frac{w_{sp}}{2} + ov_{polyem} \tag{3.82}$$

$$w_{ss} = ov_{polybase} - ov_{polyem} \tag{3.83}$$

$$d = d_{ebcont} - ov_{polyem} \tag{3.84}$$

Due to the sound physical basis all individual contributions can be attributed to the device cross section and modeled by the according equivalent circuit

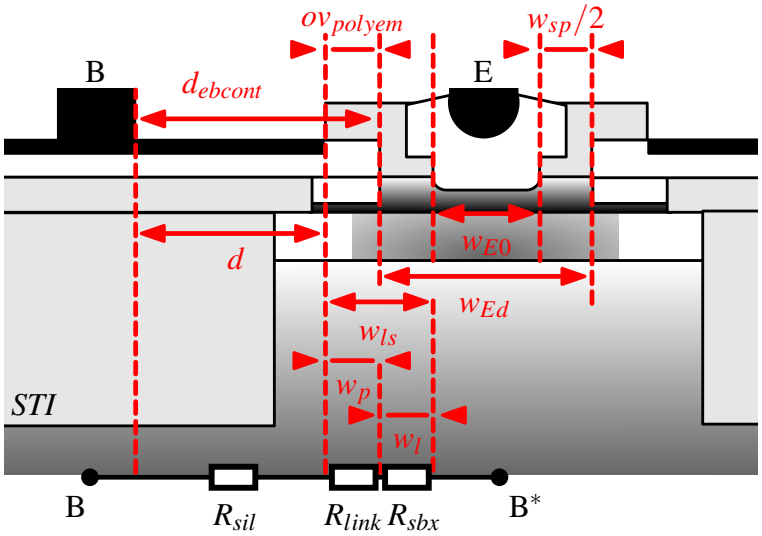


Figure 3.16: Cross section of BE region for extrinsic base resistance R_{Bx} calculation

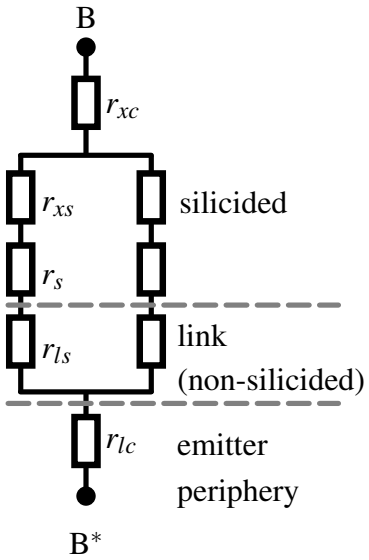


Figure 3.17: Equivalent circuit of external base resistance contributions

The according resistances are hence calculated using these dimensions as well as information about the device configuration. The three sheet resistances of each region are defined by the process specification or determined from poly resistor structures (in case of r_{ssi}) or obtained in a direct extraction procedure (cf. sect. 4.4.2 for r_{slink}, r_{sbx}).

For the calculation the flag for a single-base configuration makes the essential difference (depending on the value of *nbb*) through case-by-case analysis shown below [74].

$$r_{bx} = \frac{1}{2 \cdot g_{bx*}} \quad \text{for } nbb > 1 \text{ or} \qquad \text{Extrinsic base link resistance, general} \qquad (3.85)$$

$$r_{bx} = \frac{1}{g_{bx*}} \quad \text{for } nbb = 1 \text{ with} \qquad \text{Extrinsic base link resistance, single base} \qquad (3.86)$$

$$g_{bx*} = nbe \cdot \left[\frac{2}{r_{xs} + r_s + r_{ls}} + \frac{1}{r_{xc} + r_{lc}} \right] \qquad \text{Inverse link resistance} \qquad (3.87)$$

$$r_s = r_{s1b} = r_{ssi} \cdot \frac{w_{E0} + 2 \cdot w_{ls} + \dfrac{l_{E0} + w_{ss}}{2}}{3 \cdot w_{ss}} \qquad \text{Distinction of cases: single base contact} \qquad (3.88)$$

$$r_s = r_{s2b} = r_{ssi} \cdot \frac{\frac{w_{E0}}{2} + w_{ls}}{6 \cdot w_{ss}} \qquad \text{Case of symmetrical transistor layout} \qquad (3.89)$$

$$r_{ls} = r_{ls1b} = r_{slink} \cdot \frac{w_{ls}}{\frac{l_{E0}}{2} + w_{ls} + w_{E0}} \qquad \text{Case of single base contact} \qquad (3.90)$$

$$r_{ls} = r_{ls2b} = r_{slink} \cdot \frac{w_{ls}}{\frac{w_{E0}}{2}} \qquad \text{Cases of symmetrical transistor layout} \qquad (3.91)$$

$$r_{slink} = \frac{r_{spo} \cdot w_p + r_{sbx} \cdot w_l}{w_{ls}} \qquad \text{modified link sheet resistance} \qquad (3.92)$$

$$r_{xs} = r_{ssi} \cdot \frac{d}{w_{ss}} \qquad \text{silicided contribution} \qquad (3.93)$$

$$r_{xc} = r_{ssi} \cdot \frac{d}{l_{E0} + 2 \cdot w_{ls}} \qquad (3.94)$$

$$r_{lc} = r_{slink} \cdot \frac{w_{ls}}{l_{E0} + 2 \cdot w_{ls}} \qquad \text{link contribution} \qquad (3.95)$$

Another approach for external base resistance scalability was presented in [73] and refined in [75] using the theoretical current distribution calculation of [76] together with numerical device simulation results. The corresponding top view as well as the slightly changed equivalent circuit are shown in Fig. 3.18 and 3.19.

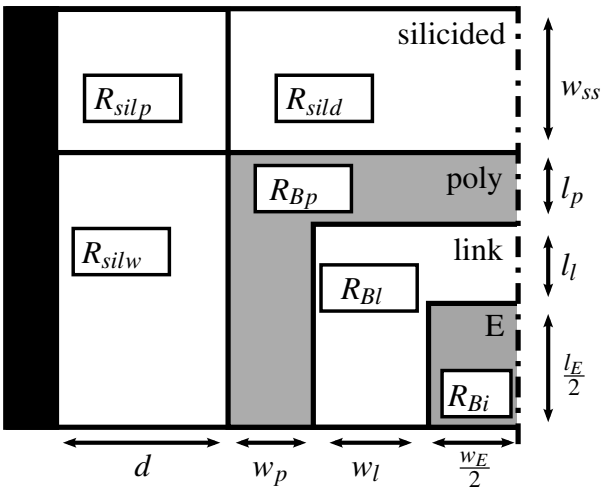


Figure 3.18: Top view of BE region with spatial dimensions for extrinsic base resistance (R_{Bx}) calculation

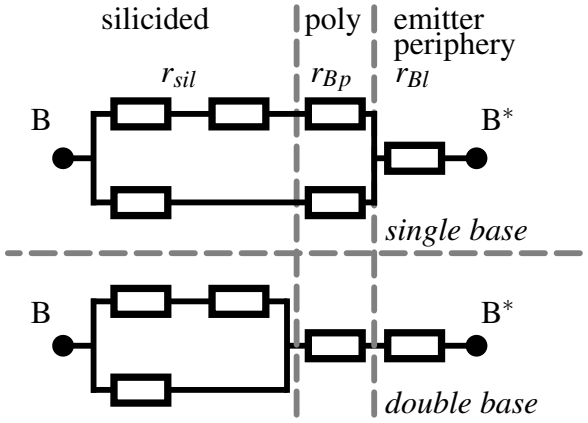


Figure 3.19: Equivalent circuit of external base resistance contributions

Verification by Numerical Device Simulation

In order to verify the suitability of the used formulation a numerical device simulation study was carried out. The three resistances R_{sil} (silicided region), R_{link} (poly-silicon to mono-silicon interface) and R_{sbx} (emitter periphery) were defined according to their values used for modeling. Using a constant doping profile the hole-mobilities for the respective sections (cf. Fig. 3.18) were adjusted to reproduce the sheet resistance (r_S) of each region following the formula

$$r_s = \frac{1}{q\mu_p N l_z} \qquad \text{general dependence of sheet resistance on material parameter} \qquad (3.96)$$

Even though for this simple structure the discretization is not very critical for this simple structure, the mesh grids in the top part of the emitter periphery were kept constant throughout the trials in order to ensure comparability between the results.

R_{Bx} Region	Width	Length	Sheet resistance
emitter periphery link	$w_l = 40nm = \frac{w_{sp}}{2}$	$l_l = 40nm = \frac{w_{sp}}{2}$	$r_{slink} = 6000\frac{\Omega}{\square}$
poly-mono interface link	$w_p = 27nm$	$l_p = 27nm$	$r_{sbx} = 400\frac{\Omega}{\square}$
silicided region	$w_{sil} = d = 234nm$	$l_{sil} = w_{ss} = 67nm$	$r_{ssil} = 15\frac{\Omega}{\square}$
internal base	$w_E = 90nm$	$l_E * variable$	r_{Bi}

Table 3.4: Set of input parameter for numerical device simulation

For two-dimensional simulation the different base regions were modeled as resistive sheets (cf. Fig. 3.20) in parallel to the BE junction (direction of lateral hole current flow). The current is

injected through pseudo-three dimensional generation in vertical direction in the inner base region (internal transistor below the emitter window).

As shown in Tab. 3.4 each region is defined through spatial dimensions (w and l) corresponding to the definition in the template as well as a sheet resistance r_S. In addition to the inner base one separates the high-resistive base link region under the spacers, (r_{Sl}), the non-silicided poly-mono interface region (r_{Sp}) ans the silicided low-resistive polysilicon region (r_{Ssil}). A sample input file is found in the appendix (cf. App. M).

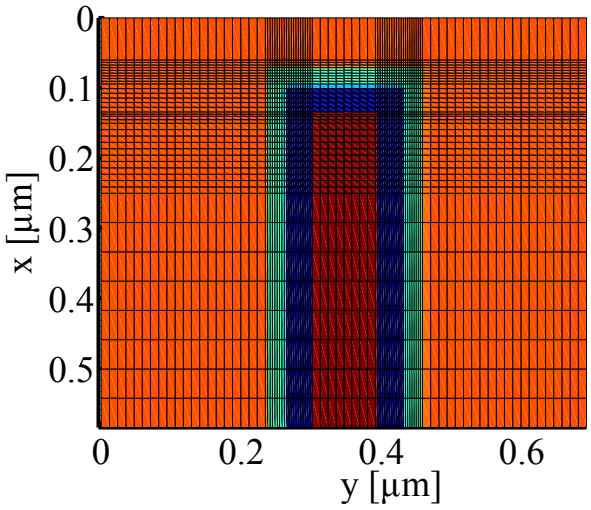


Figure 3.20: Top view of simulated symmetrical (DBC) 2D R_{Bx} structure with discretization grid and distribution of hole mobility μ_p [*top half of* 1μ*m device structure*]

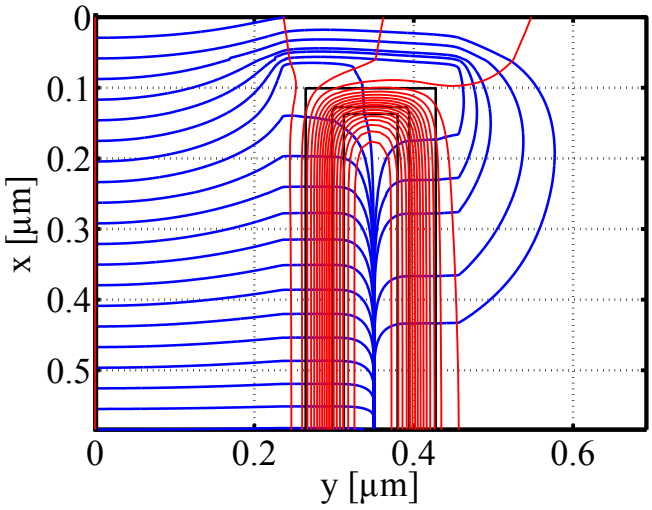


Figure 3.21: Streamline plot of simulated structure with single sided base contact (SBC) [*equipotential- and streamlines between inner base and external base contact, current injection into 2D plane*]

The device simulation was carried out for a variation of emitter window lengths (l_E) in accordance with the geometry information of the 55*nm* BiCMOS technology. The obtained results were compared with analytical formulations found in the according geometry-scalable SPICE library (denoted *SPICE library*) and the formulation given in [75] (denoted *Schroter 2008*).

In order to eliminate the influence of variation of the inner base resistance R_{SBi} the respective area was defined extremely low-ohmic with a sheet resistance of $0.1\Omega/\square$. This way the dependence on the transfer current (as in measured devices) was effectively canceled out.

The relative error of the analytical calculus is significant for both cases of single and double base configuration. Especially for very long devices the deviation was found to increase from about 0.2 to 0.8. Predictions for single base devices are more accurate with a maximum deviation around 0.4 for long devices. One can conclude that the formulation presented in (3.85) through (3.95) is equally suited for re-calculation without a need for adaptation.

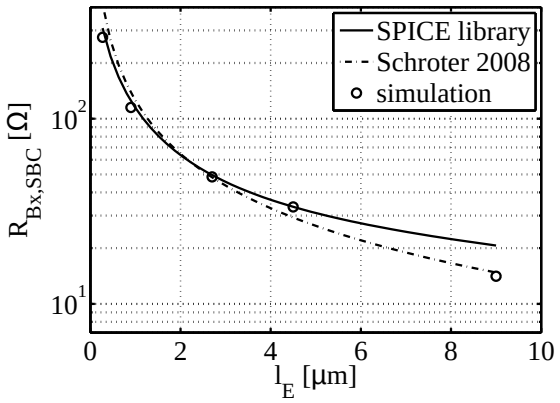


Figure 3.22: Resistance scaling for single base configuration (SBC) [*obtained resistance for simulation (points) compared with analytical formulations*]

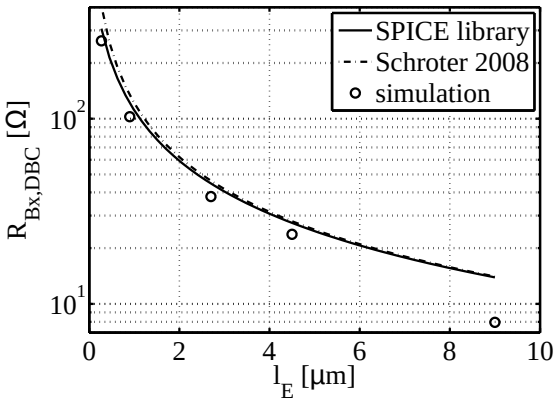


Figure 3.23: Resistance scaling for double base configuration (DBC) [*obtained resistance for simulation (points) compared with analytical formulations*]

3.3.7 Collector Series Resistance R_{Cx}

The following calculation is tailored for calculation of the collector resistance for the case of collector contact stripes in parallel with the emitter window length. For the case of symmetrical devices the flag indicating that there is more than one collector contact ($nbc > 1$) is defined as F_C.

$$F_C = 1 \text{ if } nbc > 1, \text{ else } F_C = 0 \qquad \text{collector contact flag} \qquad (3.97)$$

$$w_i = w_{sp} + s_{em} \qquad \text{dist. betw. two emitter stripes} \qquad (3.98)$$

$$l_1 = w_{sp}/2 + ov_{active} + d_{abl} + \frac{d_{bl}}{2} \qquad \text{length emitter stripe to BL limit} \qquad (3.99)$$

$$w_x = \frac{w_{sp}}{2} + ov_{active} + d_{ask} \qquad \text{case of single collector contact} \qquad (3.100)$$

$$w_{x2} = \frac{w_{sp}}{2} + ov_{active} + d_{abl} \qquad \text{case of double collector contact} \qquad (3.101)$$

$$w_{blsk} = w_x + nbe \cdot w_{E0} + (nbe-1) \cdot w_i + w_{x2} \qquad \text{internal distance between sinkers} \qquad (3.102)$$

$$r_{01c} = \frac{w_x + nbe \cdot \frac{w_{E0}}{3} + (2 \cdot nbe - 1) \cdot (nbe - 1) \cdot \frac{w_i}{6 \cdot nbe}}{l_{E0}} \qquad \text{for } nbc = 1 \qquad (3.103)$$

$$r_{02c} = \frac{\frac{w_x}{2} + nbe \cdot \frac{w_{E0}}{12} + (nbe-1) \cdot (nbe-2) \cdot \frac{w_i}{12 \cdot nbe}}{l_{E0}} \qquad \text{for } nbc > 1 \qquad (3.104)$$

$$r_c = (1 - F_C) \cdot r_{01c} + F_C \cdot r_{02c} \qquad (3.105)$$

Once the sheet resistance contributions of BL and sinker are separated in the extraction procedure (cf. sect. 4.4.3) using dedicated test structures, a sophisticated calculation approach for more complicated device structures can be applied to recalculate the resistance contributions [77, 78].

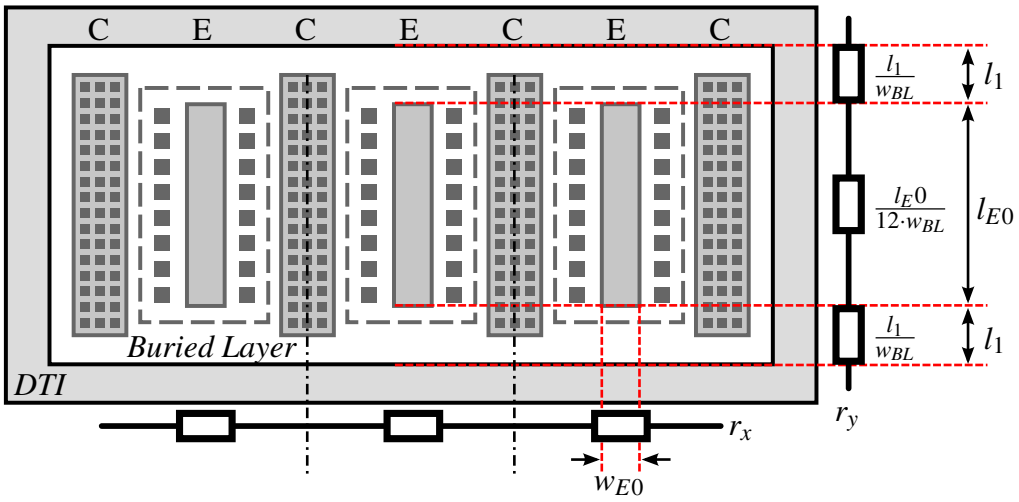


Figure 3.24: Top view of multi-finger transistor configuration for extrinsic collector series resistance (R_{Cx}) calculation

The presented formulations are based on the fundamental concept of a power dissipation approach ($R_{BL} = P_C / {I_C}^2$) solving the Poisson equation in the Fourier space. The buried layer sheet resistance is assumed to be constant throughout the transistor structure so that an analytically derived scalable formula based on Fourier series development is applicable.

In order to calculate the specific contribution of each area extensive calculations using a representation with hyperbolic functions is required. Auxiliary variables are defined as

$$m(x) = 2 \cdot x - 1 \tag{3.106}$$

$$k_1(x) = m(x) \cdot \frac{w_{E0} + w_i}{4 \cdot w_{blsk}} \tag{3.107}$$

$$k_2(x) = m(x) \cdot \frac{w_{E0} + w_i}{2 \cdot w_{blsk}} \tag{3.108}$$

$$k(x) = (1 - F_C) \cdot k_1(x) + F_C \cdot k_2(x) \tag{3.109}$$

The total buried layer resistance r_{bln} is defined as

$$r_{bln} = r_{sbl} \cdot \left[r - \sum_{n=1}^{6} S(n) \right] \tag{3.110}$$

With r being a geometry constant the variable $S(x)$ has to be calculated depending on its input value of x

$$S(x) = l_{cf} \cdot \left[\frac{w_{blsk}}{(m(x) \cdot \pi \cdot nbe \cdot l_{E0} \cdot w_{E0})} \right]^2 \cdot \frac{gx(x)}{(1+F_C)} \left[px\left(A(x), \frac{l_{E0}}{2 \cdot l_{cf}}\right) - \ldots \right.$$
$$\left. \cdots - px\left(A(x), \frac{l_{E0}+2l_1}{2 \cdot l_{cf}}\right) + \frac{1}{2} \cdot px\left(A(x), \frac{l_{E0}+l_1}{l_{cf}}\right) + \frac{1}{2} \cdot px\left(A(x), \frac{l_1}{l_{cf}}\right) \right] \quad (3.111)$$

The parameters defined in the equation $S(x)$ are further sub-divided according to the specific case

$$gx(x) = (1-F_C) \cdot g_{x1a}(x) + F_C \cdot g_{x2a}(x) \qquad \text{if } k(x) \text{ is an integer number} \quad (3.112)$$
$$gx(x) = (1-F_C) \cdot g_{x1b}(x) + F_C \cdot g_{x2b}(x) \qquad \text{if } k(x) \text{ is no integer number} \quad (3.113)$$
$$A(x) = (1+F_C) \cdot m(x) \cdot \left[\frac{\pi}{2} \cdot \frac{l_{cf}}{w_{blsk}} \right] \quad (3.114)$$

with the auto-convolutions g_x defined individually for different cases through integration

$$g_{x1a}(x) = 2 \cdot w_{blsk} \cdot \left[\frac{8 \cdot nbe}{m(x) \cdot \pi} \cdot \sin\left[\frac{m(x) \cdot \pi \cdot w_{E0}}{4 \cdot w_{blsk}} \right] \cdot \cos\left[\left(w_{x2} + \frac{w_{E0}}{2} \right) \cdot \frac{m(x) \cdot \pi}{2 \cdot w_{blsk}} \right] \right]^2 \quad (3.115)$$

$$g_{x1b}(x) = 2 \cdot w_{blsk} \cdot \left[\frac{8}{m(x) \cdot \pi} \cdot \sin\left[\frac{m(x) \cdot \pi \cdot w_{E0}}{4 \cdot w_{blsk}} \right] \cdot \ldots \right.$$
$$\left. \ldots \cdot \cos\left[\left(w_{x2} - \frac{w_i}{2} + nbe \cdot \frac{w_{E0}+w_i}{2} \right) \frac{m(x) \cdot \pi}{2 \cdot w_{blsk}} \right] \cdot \frac{\sin(k(x) \cdot \pi \cdot nbe)}{\sin(k(x) \cdot \pi)} \right]^2 \quad (3.116)$$

$$g_{x2a}(x) = w_{blsk} \cdot \left[\frac{4 \cdot nbe}{m(x) \cdot \pi} \cdot \sin\left(\frac{m(x) \cdot \pi \cdot w_{E0}}{2 \cdot w_{blsk}} \right) \right]^2 \quad (3.117)$$

$$g_{x2b}(x) = w_{blsk} \cdot \left[\frac{4}{m(x) \cdot \pi} \cdot \sin\left(\frac{m(x) \cdot \pi \cdot w_{E0}}{2 \cdot w_{blsk}} \right) \cdot \frac{\sin(k(x) \cdot \pi \cdot nbe)}{\sin(k(x) \cdot \pi)} \right]^2 \quad (3.118)$$

As defined in the r_{bln} formula with the sum $\sum S(n)$, the variable x takes values from [1 … 6]. One hence needs a case-by case analysis for the variable $px(d,x)$. Here the instance variable d represents the function $A(x)$ upon function call.

Depending on the input parameters the conditional statement for $px(d,x)$ is defined:

$$px(d,x) = 0 \qquad \text{if } d > 100 \text{ and } x < 2 \tag{3.119}$$

$$px(d,x) = \frac{1}{d} \qquad \text{if } d > 100 \text{ and } x \geq 2 \tag{3.120}$$

$$px(d,x) = \frac{\cosh(d) - \cosh(d \cdot (1-2x))}{d \cdot \sinh(d)} \qquad \text{if } d \leq 100 \tag{3.121}$$

Likewise the conditional statement for $qx(d,x)$ is defined:

$$qx(d,x) = 0 \qquad \text{if } d > 100 \text{ and } x = 0 \tag{3.122}$$

$$qx(d,x) = \frac{2}{d} \qquad \text{if } d > 100 \text{ and } x = 1 \tag{3.123}$$

$$qx(d,x) = \frac{1}{d} \qquad \text{if } d > 100 \text{ and } x > 1 \tag{3.124}$$

$$qx(d,x) = \frac{\sinh(d) - \sinh(d \cdot (1-2x))}{d \cdot \cosh(d)} \qquad \text{if } d \leq 100 \tag{3.125}$$

$$r_{sk} = \frac{r_{skl}}{l_{skd}} \tag{3.126}$$

$$g_{cside} = \frac{1}{r_{sk} + 2 \cdot r_{bln}} + \frac{1}{2 \cdot r_{sk} + 2 \cdot r_{bln}} \tag{3.127}$$

$$g_{cmiddle} = \frac{1}{r_{sk} + r_{bln}} \tag{3.128}$$

$$R_{Cx} = \frac{r_{sk}}{nbc} + r_{bln} \qquad \text{two or less collector contacts} \tag{3.129}$$

$$R_{Cx} = \frac{1}{2 \cdot g_{cside} + (nbc-3) \cdot g_{cmiddle}} \qquad \text{three or more collector contacts} \tag{3.130}$$

3.3.8 Back-End Resistances and Capacitances

The back end elements relate to all effects associated with the metal stack on top of the transistor (cf. Fig. 3.26 and 3.27) including parasitic coupling. For circuit simulation one however has to differentiate cases of RF and DC simulation when accounting for the parasitics from the bias T and RF test structure:

- for DC simulation all series contact resistances (R_{PjDC}) have to be taken into account
- for RF simulation the deembedding takes care of all parasitics up to the first metal layer (M1) and only a subset of parasitics (e.g. C_p) within the device need to be simulated

A simple extension around the simulated transistor (intrinsic model) needs to be defined using an external sub-circuit (.SUBCKT). A sub-circuit consists of a control sequence followed by device and model description. It presents a circuit that can be called from another circuit by reference to its name. The sub-circuit embedding comprises the parasitic elements shown in Fig. 3.25.

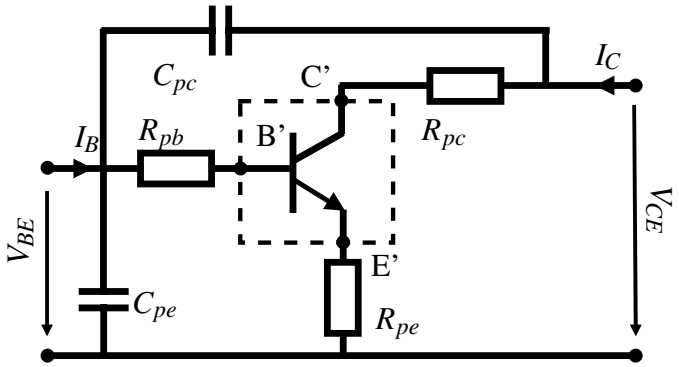


Figure 3.25: Equivalent circuit for device simulation with back-end parasitics

For RF simulation the contributions from coupling capacitances between the first metal stack ($C_{(j)m1m1}$) as well as the contacts ($C_{(j)con}$) have to be taken into account as shown exemplarily for the BE junction in Fig. 3.26.

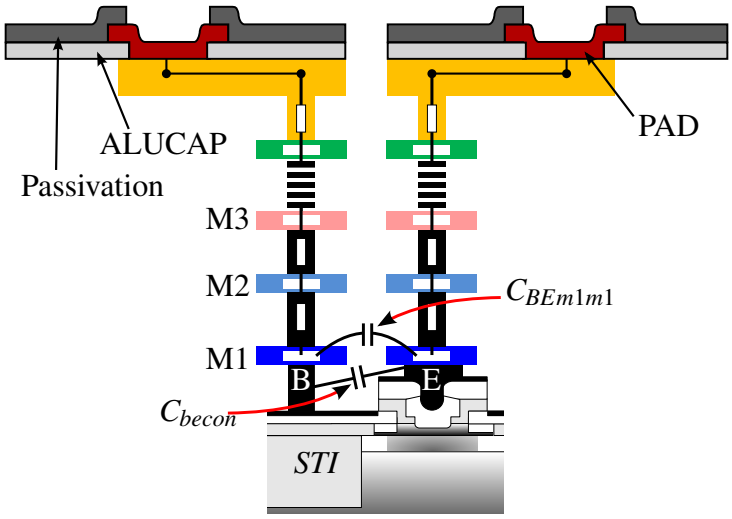


Figure 3.26: Schematic cross section of the metal stack to provide connection of the buried device to the wafer surface; [*indication of important parasitics of BE node*]

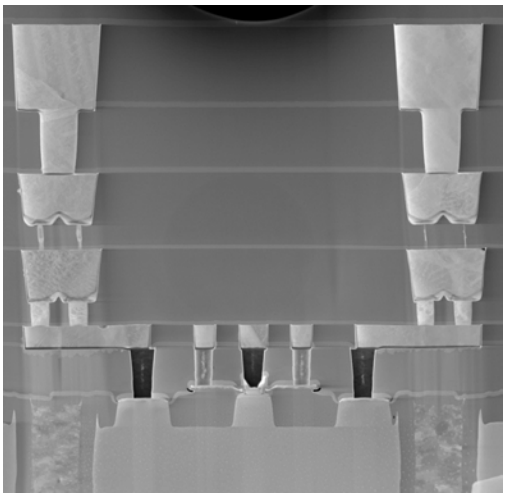


Figure 3.27: TEM cross section of first three metal layers of B5T transistor generation

For each pair of neighboring terminals (or junctions respectively) the lumped capacitance element caused by metallization and contact ($C_{p(j)}$) is hence calculated for the first metal layer for the respective nodes making use of the device dimensions and contact configuration.

BE Coupling Capacitance

The BE coupling capacitance C_{pe} is composed of two components: the fringing capacitance between the first metal layer (C_{bem1m1}) and the inter-contact capacitance (C_{becon}).

$$c_{beconl} = l_{Ed} \cdot c_{beconl\_cbe} + c_{becon\_lemin\_cbe} \text{ , for single collector } (nbc = 1) \tag{3.131}$$

$$c_{beconl} = l_{Ed} \cdot c_{beconl\_cbebc} + c_{becon\_lemin\_cbebc} \text{ , for multiple collectors } (nbc > 1) \tag{3.132}$$

$$C_{becon} = c_{beconl} \cdot \left(\frac{w_{Emin}}{w_{Ed}} \right) \tag{3.133}$$

$$C_{bem1m1} = (1 - l_{pec}) \cdot (l_{Ed} \cdot c_{bem1m1l\_cbe} + c_{bem1m1\_lemin\_cbe}) \text{ , for } nbc = 1 \tag{3.134}$$

$$C_{bem1m1} = (1 - l_{pec}) \cdot (l_{Ed} \cdot c_{bem1m1l\_cbebc} + c_{bem1m1\_lemin\_cbebc})) \text{ , for } nbc > 1 \tag{3.135}$$

$$C_{pe} = nbe \cdot (C_{becon} + C_{bem1m1}) \tag{3.136}$$

BC Coupling Capacitance

Back end capacitances related to the collector contacts are calculated depending on the drawn emitter window length l_{Ed}. The scalable evolution of the collector related coupling capacitance is obtained from:

$$c_{bccon} = l_{Ed} \cdot c_{bcconl} + c_{bccon\_lemin} \tag{3.137}$$

$$c_{bcm1m1} = (1 - l_{pec}) \cdot (l_{Ed} \cdot c_{bcm1m1l} + c_{bcm1m1\_lemin}) \tag{3.138}$$

$$C_{pc} = nbb \cdot (c_{bccon} + c_{bcm1m1}) \tag{3.139}$$

Base Resistance

The contact resistance related to the back end metallization on the extrinsic base contact is calculated in a straight forward way using the layout information. The number of via contacts is calculated as a function of the device geometry. With the known specific resistance of a single via the total contact resistance can be calculated:

$$n_{conBl} = \frac{l_{Ed} + 2 \cdot ov_{active} - 2 \cdot b_{con} + s_{conB}}{w_{con} + s_{con,B}} \text{ and } n_{conBw} = 1 \tag{3.140}$$

$$n_{conB} = nbb \cdot n_{conBl} \cdot n_{conBw} \tag{3.141}$$

$$r_{pb} = r_{Bcon} = \frac{r_{con}}{n_{conB}} \tag{3.142}$$

Emitter Resistance

The contact resistance related to the back end on top of the emitter stripe is obtained from:

$$n_{conEl} = \frac{l_{polyem} - 2 \cdot e_{con}}{w_{conrub}} \text{ and } n_{conEw} = 1 \tag{3.143}$$

$$n_{conE} = nbb \cdot n_{conEl} \cdot n_{conEw} \tag{3.144}$$

$$A_{via} = nbe \cdot [(l_{polyem} - 2e_{con}) \cdot w_{conruban}] \tag{3.145}$$

$$r_{Econ} = \frac{r_{kvia}}{A_{via}} \tag{3.146}$$

$$n_{viaEl} = \frac{l_{Ed} + 2 \cdot ov_{active} - 2 \cdot e_{m1con} + s_{via}}{w_{via} + s_{via}} \text{ and } n_{viaEw} = 2 \tag{3.147}$$

$$n_{viaE} = nbe \cdot n_{viaEl} \cdot n_{viaEw} \tag{3.148}$$

$$r_{Evia} = \frac{r_{via}}{n_{viaE}} \cdot (metals\_e - 1) \tag{3.149}$$

$$r_{pe} = r_{Econ} + (1 - l_{per}) r_{Evia} \tag{3.150}$$

Collector Resistance

The parasitic series resistance at the external collector contacts is obtained using the following formulas:

$$n_{conCl} = \frac{l_{skd} - 2 \cdot a_{con} + s_{con\_collector}}{w_{con} + s_{con\_collector}} \text{ and } n_{conCw} = 3 \tag{3.151}$$

$$n_{conC} = nbc \cdot n_{conCl} \cdot n_{conCw} \text{ if } nbe = 1 \tag{3.152}$$

$$n_{conC} = (nbc + 1) \cdot n_{conCl} \cdot n_{conCw} \text{ if } nbe > 1 \tag{3.153}$$

$$r_{Ccon} = \frac{r_{con}}{n_{conC}} \tag{3.154}$$

$$n_{viaCl} = \frac{l_{skd} - 2 \cdot e_{m1con} + s_{via}}{w_{via} + s_{via}} \tag{3.155}$$

$$n_{viaCw} = 7 \text{ for single collector } (nbc = 1), \text{ else } n_{viaCw} = 3 \tag{3.156}$$

$$n_{viaE} = nbc \cdot n_{viaCl} \cdot n_{viaCw} \tag{3.157}$$

$$r_{Cvia} = \frac{r_{via}}{n_{viaC}} \cdot (metals\_c - 1) \tag{3.158}$$

$$r_{pc} = r_{Ccon} + (1 - l_{per}) r_{Cvia} \tag{3.159}$$

3.3.9 Current Parameters

The diodes for ideal and non-ideal current components modeled in the HICUM equivalent circuit use the following designations: [BEi, BCi, BEp, BCx, SC] corresponding to the table for currents and components in HICUM (cf. Tab. 2.1).

The parameters obtained in each of the extraction steps are representing normalized saturation currents for the respective perimetric (index $_p$) and surfasic (index $_a$) contribution. Hence, the set of scaling equations of low current parameters are relatively simple functions of the transistors effective and real junction width and length.

Static Base-Emitter Junction Current Components

In the native HICUM model the quasi-static internal base current with its corresponding recombination part flowing into the emitter is split into a bottom and a peripheral component. By definition the bottom portion models the current injected across the emitter area while the peripheral component models the current injected across the peripheral BE junction.

All base current components are scaled using the number of emitter fingers multiplied with the respective real emitter area (A_{E0}) and perimeter (P_{E0}) respectively. In turn the first group of equations for base-emitter related components reads as follows:

$$I_{BEIS} = nbe \cdot j_{bea} \cdot A_{E0} \tag{3.160}$$

$$I_{BEPS} = nbe \cdot j_{bep} \cdot P_{E0} \tag{3.161}$$

$$I_{REIS} = nbe \cdot j_{rea} \cdot A_{E0} \tag{3.162}$$

$$I_{REPS} = nbe \cdot j_{rep} \cdot P_{E0} \tag{3.163}$$

The BE-tunneling effect is modeled with the parameter I_{BETS} of the BE tunneling current equation that is linearly scaled with the effective emitter area.

$$I_{BETS} = j_{bets} \cdot A_{Eeff} \tag{3.164}$$

Independent of device geometry the thermal scaling equations built into the model code as well as the non-ideality factors (m) are used for all geometries with their respective (uniform) parameters.

Static Base-Collector Junction Current Components

The current relative to the internal and external BC junction are scaled with the real base-collector area and perimeter

$$I_{BCi} = nbe \cdot jbca \cdot A_{b0} \tag{3.165}$$

$$I_{BCx} = nbe \cdot jbcp \cdot P_{b0} \tag{3.166}$$

The corresponding non-ideality factors m_{BC} are again independent of geometry.

Substrate Transistor

The static transfer current of the substrate transistor scales with the buried layer area and perimeter bounded by the trench isolation. However a difference is made between the two contributions using the real dimensions of the buried layer for CS junction current:

$$I_{TSS} = nbe \cdot (j_{tssa} \cdot A_{bd} + j_{tssp} \cdot P_{bd}) \tag{3.167}$$

$$I_{SCS} = j_{csa} \cdot A_{c0} + j_{csp} \cdot P_{c0}. \tag{3.168}$$

Avalanche current

From the equation below one can tell that the avalanche effect scales with the transfer current. In order to take avalanche multiplication into account no additional saturation current is defined.

$$I_{AVL} = I_{Tf} F_{AVL} (V_{DCi} - v_{B'C'}) \exp\left(-\frac{Q_{AVl}}{C_{JCi}(V_{DCi} - v_{B'C'})}\right) \tag{3.169}$$

Yet according to the definition in the model code, the parameters used for weighting depend directly on emitter area (A_E), physical constants (a_n and b_n) and temperature

$$F_{AVL} = 2 \cdot a_n / b_n \tag{3.170}$$

$$Q_{AVL} = b_n \varepsilon \cdot A_E / 2. \tag{3.171}$$

Therefore the avalanche current factor F and the exponent factor Q for avalanche current are scaled with the device geometry. Through γ_{avl} the effective avalanche area (A_{avl}) as defined in eqn. (3.37) is used. The relation to the BC junction is then taken into account within the scaling equation of the exponent factor as

$$Q_{AVL} = nbe \cdot q_{avlu} \cdot A_{cbc} \tag{3.172}$$

$$F_{AVL} = f_{avlu} \cdot \frac{A_{avl}}{A_{Eeff}}. \tag{3.173}$$

Quasi-static transfer current

In order to compute the model parameters for two- (and three-dimensional respectively) scaled transistors, all area specific elements (such as currents and charges) used in the Generalized Integral Charge-Control Relation (GICCR) need to be be multiplied by the corresponding effective emitter area A_{Eeff} as defined in eqn. (3.19).

As pointed out before the transfer current at low injection is calculated using two model pa-

rameters, which are namely the hole charge Q as well as the GICCR factor c_{10}

$$i_T = \frac{c_{10}}{Q_{p,T}} \left[\exp\left(\frac{V_{B'E'}}{V_T}\right) - \exp\left(\frac{V_{B'C'}}{V_T}\right) \right], \text{ with} \tag{3.174}$$

$$c_{10} = (qA_E)^2 V_T \overline{\mu_{nB} n_{iB}{}^2}. \tag{3.175}$$

In the model definition I_{Ch} (onset of high injection phenomena) accounting for high injection effects is specified as a model parameter that is (roughly) proportional to the emitter area

$$c_1 = c_{10}\left(1 + \frac{I_{TF1}}{I_{Ch}}\right). \tag{3.176}$$

However the corresponding normalized scalable parameter j_{ch} is passed to the simulator without any further manipulation.

Accordingly the GICCR model parameters are scaled using the effective electrical emitter using the scaling equations as shown below. The modified hole charge (Q_{PT}) is making use of the hole charge at thermal equilibrium Q_{P0}, which in itself scales with the emitter area

$$Q_{P0} = nbe \cdot Q_{P0U} \cdot A_{Eeff} \tag{3.177}$$

$$C_{10} = nbe^2 \cdot C_{10U} \cdot A_{Eeff}{}^2. \tag{3.178}$$

3.3.10 Transit Time Parameters

The intrinsic device speed of the SiGe HBT device is a function of the emitter (τ_E), base (τ_B) and collector (τ_C) transit times (with the base transit time τ_B being the limiting term). In a simple approximation by the drift diffusion approach the base transit time of a compositional-graded base depends on the base width (w_B) and the electron diffusivity (D_e) as well as carrier velocity (v_T) and the band-gap difference (ΔE_g)

$$\tau_B = \frac{L_g w_B}{D_e} - \left(\frac{L_g^2}{D_e} - \frac{L_g}{v_T}\right)\left(1 - \exp\left(-\frac{w_B}{L_g}\right)\right), \text{ with } L_g = \frac{w_B \cdot kT}{\Delta E_g} [79]. \tag{3.179}$$

Hence, to a large extent the transit time scaling is linked to the vertical profile design which is fixed for a given technology. The transit time shall therefore be associated to the intrinsic transistor. However the variation of external elements and parasitic effects (e.g. parasitic capacitances, series resistances and self-heating) call for a geometry-dependent scaling of the transit time complex.

As shown in eq. (3.180) the transit time at low injection (τ_0) scales with the relation of real emitter area (A_{E0}) relative to the effective emitter area (A_{Eeff}) [71]. Furthermore the inner series resistance in the collector plays an important role. The area-specific internal collector resistance is

defined in the model as inversely proportional to the emitter area.

$$T_0 = T_{0a} \cdot \frac{A_{E0}}{A_{Eeff}} + T_{0p} \cdot \left(1 - \frac{A_{E0}}{A_{Eeff}}\right) \tag{3.180}$$

$$L_{ATB} = \frac{lat}{w_{Eeff}} \tag{3.181}$$

$$L_{ATL} = \frac{lat}{l_{Eeff}} \tag{3.182}$$

$$F_{CS} = 1 + L_{ATB} \text{ , for identical lat parameter } L_{ATB} = L_{ATL} \tag{3.183}$$

$$F_{CS} = \frac{L_{ATB} - L_{ATL}}{\ln\left(\frac{1+L_{ATB}}{1+L_{ATL}}\right)} \text{ , for other cases} \tag{3.184}$$

$$R_{CI0} = \frac{r_{kci0}}{A_{Eeff} \cdot F_{CS} \cdot nbe} \tag{3.185}$$

Given the device architecture and the current flow through the device the effective lateral extension of the SIC may vary. Another built in feature of the HICUM model (that is closely related to the transit time scaling) is hence the collector current spreading in the intrinsic transistor or lateral geometry scaling (at high injection) respectively [71].

Therethrough, in addition to vertical scaling by means of the injection width (w_{Ci}) another degree of freedom for lateral scaling is given. This is taken into account by the parameters L_{ATB} and L_{ATL} as well as the collector current spreading factor fcs used to scale the critical current I_{CK}. L_{ATB} here is the charge scaling factor for Q_{fC} in w_E direction while is L_{ATL} accounts for scaling in l_E direction. In order to simplify the extraction procedure the model library uses a unique parameter lat that is normalized by the effective device dimension (c.f. eqn. (3.181) and (3.182)).

3.3.11 Self Heating and Thermal Resistance

Device heating and the estimation of thermal runaway and a safe operating area (SOA) are important for todays transistors with high current densities in small volumes. Transistor inherent effects which lead to thermal instability and breakdown at high current densities need to be taken into account. For a full analysis of thermal effects, predictions are usually calculated using computational fluid dynamics (CFD) and finite element analysis (FEA) methods. Yet these are not practical for device modeling. Therefore, parameters related to self heating and the thermal network need to be consciously scaled with the set of equations given below.

It is relatively easy, to add a electrothermal model to the electric model. By adding an extra thermal node to the electrical compact models, this node will provide information about junction temperature of the active device and represents a connection between the individual device and rest of the simulated circuit network.

Proper self-heating effect characterization is based on steady-state extraction of the thermal resistances (R_{TH}) and transient conditions to extract the thermal capacitance (C_{TH}). In HICUM the temperature increase of a device is modeled as a function of the dissipated power (P_D) and the thermal resistance (R_{TH}) of each transistor. The dissipated power is calculated using the terminal voltages and currents inside the model yet the computation of thermal resistances is strongly dependent on the specific device layout. The thermal circuit shown in Fig. 2.10 is a popular way to represent temperature effects for AC and DC characteristics of transistors. To begin there is a number of predefined parameters defined in the model library relative to the back-end resistance calculation:

Abbr.	Parameter Description	Abbr.	Parameter Description
$roth_w$	Thermal resistance of tungsten (W)	$roth_{cu}$	Thermal resistance of copper (Cu)
h_{con}	Contact height	h_{m1}	Height of first metal layer (M1)
h_{via1_m2}	Height of VIA1 and M2	h_{via2_m3}	Height of VIA2 and M3
h_{via3_m4}	Height of VIA3 and M4	h_{via4_m5}	Height of VIA4 and M5
$roth_{si}$	Thermal resistance of bulk Si	t_{wafer}	Wafer thickness
$toth_0$	Total stack height		

Table 3.5: Definition of global parameters related to the back-end

The modeling and extraction of thermal effects was part of the work presented in [80]. One defines the thermal capacitance (C_{TH}) and the resistance (R_{TH}) as the ratio of device temperature increase (ΔT) to dissipated power (P_{diss}). For the electrical equivalent circuit implemented in HICUM the controlling 'voltage' is replaced by ΔT, whereas the 'current' is represented by the power dissipated in the device under test. The derivative of the junction temperature (T_j) is the thermal resistance

$$R_{TH} = \frac{\partial T_j}{\partial P_{th}} \text{ with the simplification } \Delta T = R_{TH} \cdot P_{diss}. \tag{3.186}$$

By default intra-device mutual thermal coupling is not taken into account. For simulation in large circuits the separate thermal node for thermal coupling gives more options. However for single devices the standard PCell layout with five metal layers (M5) is used as shown in the simplified schematic view of Fig. 3.28.

The resistance calculation in the substrate uses information about wafer thickness and other material specific parameters as defined in Tab. 3.5. The implementation of parameters is based on a rather complex set of scaling equations.The different implementations of rth_i are separated for calculation of square-shaped (index $_{sqr}$) as well as rectangular (index $_{rect}$) devices whereas for

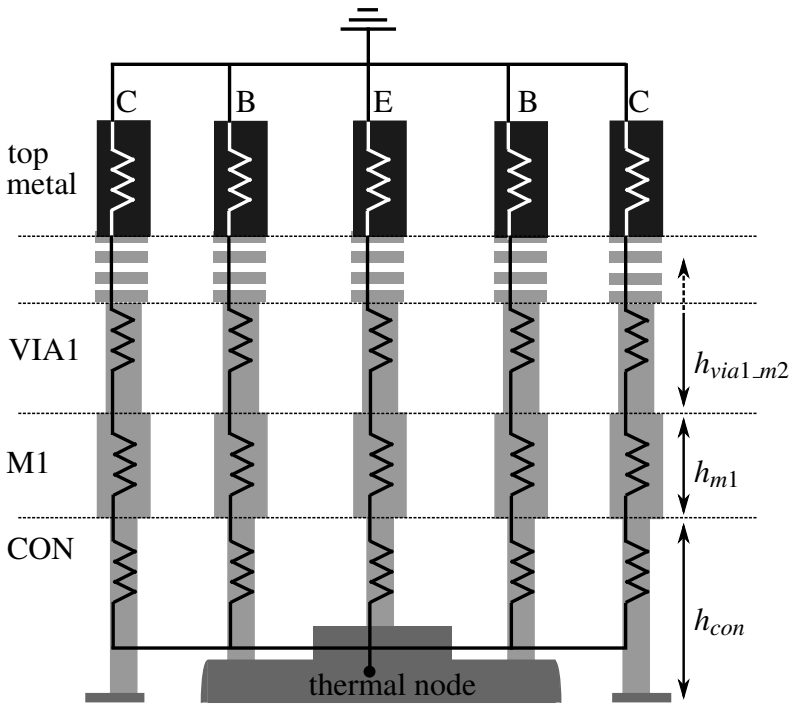


Figure 3.28: Schematic view of back-end for thermal simulation of symmetrical CBEBC device (for a full cross-sectional view cf. [80])

the most part of supported devices the rectangular case applies. Once more the flag F_C is used, depending on the device configuration regarding the collector contact configuration.

In this approach, the DTI is regarded as thermal isolator. For proper thermal resistances modeling the heat flow through the respective areas of the device are evaluated. From material properties (such as the thermal conductivity of silicon $\kappa_{si}(T)$) and dimensions the general equation for DTI based technologies is derived:

$$r_{th} = r_{th\_dt} = \frac{1}{\frac{1}{roth_{si}\ cdot\left[\frac{rth_0}{nbe} + \sum_{i=1}^{4} rth_i\right]} + \frac{1}{rth_{be}}}, \text{ for deep trench technology} \tag{3.187}$$

$$rth_i = \frac{1}{\kappa_{si}(T)} \int \frac{di}{A_i(x,y,z)}, i = x,y,z \tag{3.188}$$

Back-end process layers and metallic interconnects play an important role in the calculation of thermal resistance. A non-negligible heat flow is absorbed through the back-end-of-line. Therefore the influence of the back-end layers and metallic interconnects on the total resistance R_{TH} needs to be evaluated. For accurate evaluation the contact number as well as the number of vias (*nviac*) defined during the back-end resistance calculation are used to obtain the total resistance rth_{be} from

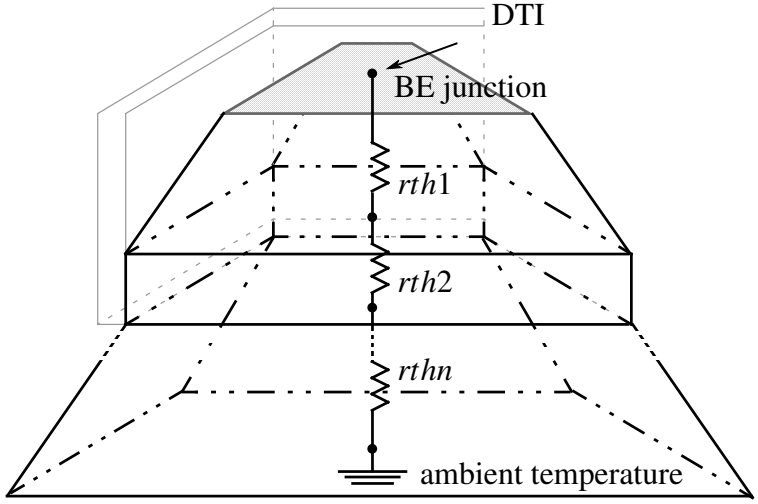


Figure 3.29: Schematic view of the distributed thermal network between device and the ambient temperature (cf. [80])

the device (acting as a heat source) to external node. A simplified graphical representation of the thermal network between the source (BE junction) and the ambient temperature as reference node is shown in Fig. 3.29.

$$rth\_con\_m1 = \frac{roth_w \cdot hcon + roth_{cu} \cdot h_{m1}}{(nconc + ncone + nconb) \cdot w_{con}^2} \tag{3.189}$$

$$nvia_{(k)} = nviac \tag{3.190}$$

$$rth\_via(k)\_m(k+1) = roth_{cu} \cdot \frac{hvia(k)\_m(k+1)}{nvia_{(k)} \cdot w_{via}^2} \tag{3.191}$$

$$rth_{be} = rth\_con\_m1 + \sum_{k=1}^{4} rth\_via(k)\_m(k+1) \tag{3.192}$$

The components (rth_i) can be split up depending on their respective transistor area. The intermediate resistance associated with the emitter window opening is defined

$$h0 = w_{sp}/2 + ov_{active} + d_{abl}\ ,\ \ w0 = w_{E0}\ ,\ \ l0 = l_{E0} \tag{3.193}$$

$$a0 = \frac{w_{E0} + 2 \cdot h0}{w_{E0}} \ \text{ and } \ b0 = \frac{l_{E0} + 2 \cdot h0}{l_{E0}} \tag{3.194}$$

$$rth0_{sqr} = \frac{h0}{w0 \cdot (w0 + 2 \cdot h0)} \tag{3.195}$$

$$rth0_{rect} = \frac{h0 \cdot \ln\left(\frac{a0}{b0}\right)}{w0 \cdot l0 \cdot (a0 - b0)} \tag{3.196}$$

$$rth_0 = rth0_{sqr}\ \text{, if } |w_{Ed} - l_{Ed}| \leq 5 \cdot 10^{-9}\text{, and } rth_0 = rth0_{rect}\ \text{ else} \tag{3.197}$$

The rest of the components (first through fourth component [$rth_1 ... rth_4$]) account for the buried layer scaling with the associated drawn dimensions (w_{cd} and l_{cd}).

$$h1 = s_{em}/2 - ov_{active} - d_{abl} \text{ and } l1 = l_{cd} \tag{3.198}$$

$$w1 = d_{abl} + ov_{active} + w_{Ed} + ov_{active} + d_{abl} \tag{3.199}$$

$$rth1\_1c = \ln(\frac{w1+h1}{w1})/l1 \tag{3.200}$$

$$rth1\_2c = \frac{\ln(\frac{w1+2\cdot h1}{w1})}{2\cdot l1} \tag{3.201}$$

$$rth_1 = \frac{1}{F_C\cdot(nbe/rth1\_2c) + (1-F_C)\cdot[(nbe-1)/rth1\_2c + 1/rth1\_1c]} \tag{3.202}$$

$$h2 = ov_{active} + d_{ask} + w_{skd} + d_{abl} - s_{em}/2 \text{ and } l2 = l_{cd} \tag{3.203}$$

$$a2\_1c = w_{cd}/((nbe-1)\cdot(s_{em}/2 + w_{Ed} + s_{em}/2) + (d_{abl} + ov_{active} + w_{Ed} + s_{em}/2)) \tag{3.204}$$

$$a2\_2c = w_{cd}/(nbe\cdot(s_{em}/2 + w_{Ed} + s_{em}/2)) \tag{3.205}$$

$$rth2\_1c = \ln(a2\_1c)/l2 \tag{3.206}$$

$$rth2\_2c = \ln(a2\_2c)/(2\cdot l2) \tag{3.207}$$

$$rth2 = F_C\cdot rth2\_2c + (1-F_C)\cdot rth2\_1c \tag{3.208}$$

$$rth_2 = rth2 \text{ , if } \frac{s_{em}}{2} < (ov_{active} + d_{ask} + w_{skd} + d_{abl}) \text{ , else } rth_2 = 0 \tag{3.209}$$

The third component is dedicated to the area enclosed deep trench up to the DTI depth h_{dt}.

$$h3 = h_{dt} - (\frac{w_{sp}}{2} + ov_{active} + d_{ask} + w_{skd} + d_{abl}) \tag{3.210}$$

$$w3 = w_{cd} \text{ and } l3 = l_{cd} \tag{3.211}$$

$$rth3 = h3/(w3\cdot l3) \tag{3.212}$$

$$rth_3 = 0 \text{ , if } h_{dt} < (\frac{w_{sp}}{2} + ov_{active} + d_{ask} + w_{skd} + d_{abl}) \tag{3.213}$$

$$rth_3 = rth3 \text{ else} \tag{3.214}$$

The fourth component accounts for the full bulk silicon and uses the associated dimension of

the wafer t_{wafer}.

$$h4 = t_{wafer} - (\frac{w_{sp}}{2} + ov_{active} + d_{ask} + w_{skd} + d_{abl}) \tag{3.215}$$

$$\text{if } \; h_{dt} < (\frac{w_{sp}}{2} + ov_{active} + d_{ask} + w_{skd} + d_{abl}) \tag{3.216}$$

$$h4 = t_{wafer} - h_{dt} \text{ else} \tag{3.217}$$

$$w4 = w_{cd} \;\text{ and }\; l4 = l_{cd} \tag{3.218}$$

$$a4 = (w4 + 2 \cdot h4)/w4 \;\text{ and }\; b4 = (l4 + 2 \cdot h4)/l4 \tag{3.219}$$

$$rth4_{sqr} = h4/(w4 \cdot (w4 + 2 \cdot h4)) \tag{3.220}$$

$$rth4_{rect} = h4 \cdot \ln(a4/b4)/(w4 \cdot l4 \cdot (a4 - b4)) \tag{3.221}$$

$$rth_4 = rth4_{sqr} \text{ , if } abs(w_{cd} - l_{cd}) \leq 5 \cdot 10^{-9}, \text{ and } rth_4 = rth4_{rect} \;\text{ else} \tag{3.222}$$

In the end the two governing parameters for self-heating are calculated when the flag sh_{eff} is set to 1. Otherwise this flag allows to simulate without taking into account the SH network.

$$c_{th} = \frac{toth}{R_{TH}} \text{ , with } toth = \frac{toth_0}{A_{cd}^{3/2}} \tag{3.223}$$

$$R_{TH} = sh_{eff} \cdot r_{th} \;\text{ and }\; C_{TH} = sh_{eff} \cdot c_{th} \tag{3.224}$$

3.3.12 Noise

1/f Noise

The technological scaling of flicker (1/f) noise in SiGe HBTs and their minimization is an important factor for amplification applications. In the model the 1/f noise scaling is strongly linked to the resistance scaling as well as to the base current I_B. Explicitly added in the HICUM model version L2.31, the flicker noise of the emitter resistance has a significant influence and must be taken into account.

Transistor size dependent variation of the low frequency noise amplitude is scaling with the inverse number of carriers in the noise generating elements. Hence, it is implemented using the emitter geometry scaling for both flicker noise factors (KF factor) using the number of emitter fingers as well as the drawn emitter dimensions (A_{Ed}). The two corresponding exponential factors (model parameters a_F and a_{FrE}) are considered independent of device geometry.

$$K_F = \frac{k_{fu}}{nbe \cdot A_{Ed}} \tag{3.225}$$

$$K_{FRE} = \frac{k_{fure}}{nbe \cdot A_{Ed}} \tag{3.226}$$

Thermal and shot noise

Generally, the shot noise is calculated based on the amount currents across the respective junctions. Since all these parameters (transfer current I_T, base current components as well as avalanche current) are already taken into account by the scalable model library the Shot noise has no additional scaling parameter.

Likewise thermal noise generated in ohmic resistances (r_E, r_{Cx}, r_{Bx}, or r_{Bi}) is modeled using the respective resistance value as well as the device temperature T. Both these quantities already follow the device geometry, hence the thermal noise has no additional scaling parameter.

3.3.13 Unitary Parameters

As mentioned before, there is a number of parameters, that are either strongly linked to the vertical profile or the process and hence don't scale with the lateral device geometry. However these parameters might be equally significant for the accurate modeling and the description of device operation. The set of unitary parameters comprises closely physics related quantities such as transit times (τ) and grading- (z) as well as non-ideality- (m) factors. But additional decisive parameters such as flags are influencing the model behavior in a similar way and are hence presented as well.

One particularly important parameter for example is the flag FL_{SH} used to turn on or off the simulation of device self-heating effects. This instance parameter, directly passed to the simulator, is especially useful to avoid calculation of SH equations and thus improve convergence. Another one is the $TUNODE$ that attributes the tunneling current either to the internal base node or to the perimeter base.

Junction Capacitances

Due to the strong interaction with the device charges the junction capacitances present an important part of the device model. Each one of the zero-bias junction capacitances is a strong function of the device dimensions. However, the bias dependence of the junction capacitances is assumed to be identical for any given device dimension. Both the grading coefficients (z) as well as the built-in potential (V_D) are therefore not scaled as a function of device geometry. Table 3.6 below gives an overview which of the junction-related parameters are independent of lateral scaling.

BE junction	BC junction	CS junction
V_{DEI}; Internal BE built-in potential	V_{DCI}; Internal BC built-in potential	V_{DS}; CS built-in potential
Z_{EI}; Internal BE grading coefficient	Z_{CI}; Internal BC grading coefficient	Z_S; CS grading coefficient
V_{DEP}; Peripheral BE built-in potential	V_{DCX}; External BC built-in potential	V_{DSP}; peripheral CS built-in potential
A_{JEI}, Limit of maximum to zero-bias value of internal BE capacitance[1]	Z_{CX}; External BC grading coefficient	Z_{SP}; peripheral CS grading coefficient
F_{BEPAR}, BE partitioning option for partial capacitance components	F_{BCPAR}, BC partitioning option for partial capacitance components	

Table 3.6: Unitary model parameter for junction capacitances

Vertical non-quasi-static effects

There are three parameters describing non-quasi-static effects in the device. None of them is implemented in a way that scales with device geometry.

FL_{NQS}; Flag for vertical NQS effects	A_{LIT}; Factor for additional delay time of transfer current	A_{LQF}; Factor for additional delay time of minority charge

Table 3.7: Unitary model parameter for non-quasi-static effects

Static currents

Most of the currents in the devices scale with the bottom surface of the emitter window. The saturation currents of the respective equations are therefore individually calculated depending on the PCell instance parameters. However, the non-ideality factors (m) are found to be independent of the lateral device scaling.

M_{REI}; Internal BE recombination current ideality factor	M_{REP}; Peripheral BE recombination current ideality factor
M_{BEI}; Internal BE current ideality factor	M_{BCI}; Peripheral BE recombination current ideality factor
M_{BEP}; Peripheral BE current ideality factor	M_{BCX}; Peripheral BE recombination current ideality factor
A_{BET}; Exponent factor for tunneling current	

Table 3.8: Unitary model parameter of static currents

Noise parameters

Emitter geometry scaling is having an impact on the noise given the direct dependence on the current flowing through the junctions as described in section 3.3.12. Based on noise measurements, one can conclude that the noise exponent factors are constant for all geometries in SiGe HBTs.

A_F; Flicker noise exponent factor	A_{FRE}; Emitter resistance flicker noise exponent factor

Table 3.9: Unitary model parameter of noise model

Transfer current and charge weighting factors

The principal charges in the transistor are scaled with device geometry as shown in eqn. (3.177) and (3.178). However the terms describing bias dependence as well as the charge weighting factors (h) are properties of the vertical device profile and thus assumed to be constant for a given device technology.

A_{HJEI}; Slope factor of $H_{JEI}(V_{BE})$	R_{HJEI}; Smoothing factor for H_{JEI} at high forward bias
H_{FE}; Emitter minority charge weighting factor (for HBTs)	H_{FC}; Collector minority charge weighting factor (for HBTs)

Table 3.10: Unitary model parameter related to the transfer current and charge weighting factors

Transit time and high current

A significant part of model parameters in the HICUM model is attributed to effects under high injection condition. Therefore the set of parameters defining the onset of the critical current (I_{CK}) for example are very important for the accurate description of the device characteristics at high frequency.

The low-current forward transit time (τ_0) is scaled with the device geometry to take into account variation with the lateral dimension. However most of the remaining transit time parameters are linked to the vertical profile and thus implemented as unitary parameters.

The definition of critical voltages and currents is likewise important for a sound physical link between model and reality. Due to the limited interest for accurate description of the reverse operation of high-speed HBT devices and hence the reverse transit time (or storage time for inverse operation τ_r, model parameter: T_R) is modeled completely geometry-independent.

V_{PTCI}; Internal BC punch-through voltage	V_{PTCX}; punch-through voltage of the external collector region	V_{PTS}; CS punch-through effect
F_{THC}; Partitioning factor for base and collector transit time	A_{HC}; Smoothing factor for current dependence of B and C transit time	H_{F0}; Weight factor for low current minority charge
T_{BHREC}; Base current recombination time constant at the BC barrier	V_{LIM}; Separation voltage between ohmic and saturation velocity regime	V_{CES}; Internal CE saturation voltage
D_{ELCK}; Field dependence factor for I_{CK}	D_{T0H}; Time constant for base and BE SCR width modulation	T_{BVL}; Time constant for carrier jam at low V_{CE}
T_{EF0}; Neutral emitter storage time	G_{TFE}; Exponent factor for current dependence of neutral emitter storage time	T_{HCS}; Saturation time constant at high current

Table 3.11: Unitary model parameter related to the transfer current and charge weighting factors

Resistances

Given the extraction procedure of the different resistance contributions all sheet resistance values are scaled with device dimension as well as configuration to accurately model the parasitic access resistance as well as the inner base resistance. Only a set of correction factors and ratios taken into account inside the model are hence assumed to be constant.

F_{QI}; Ratio of internal to total minority charge	F_{DQR0}; Correction factor for modulation by BE and BC SCR
F_{GEO}; Factor for geometry dependence of emitter current crowding	F_{CRBI}; Ratio of HF shunt to total internal capacitance (lateral NQS effect)

Table 3.12: Unitary model parameter related to resistance calculation

Substrate transistor

The substrate transistor saturation current is modeled using the bottom area of the transistor. However as for the static currents the non-ideality factor (m) of the current components is constant throughout the various device geometries. Likewise the scaling of the according transit time is linked to the vertical CS junction profile and hence assumed to be unaffected by device scaling.

M_{SF}; Forward ideality factor of substrate transfer current	M_{SC}; Ideality factor of CS diode	T_{SF}; Forward transit time of the substrate transistor

Table 3.13: Unitary model parameter for the substrate transistor

Barrier Effect

The heterojunction barrier effect is a newly implemented feature of model version L2.30. In theory the modeled barrier effect is a function of the vertical Ge profile layout and thus not scaling with device geometry. The corresponding parameters are therefore taken as unitary parameters.

V_{CBAR}; BC barrier voltage	I_{CBAR}; Current normalization parameter	A_{CBAR}; Smoothing parameter for bias dependence of V_{CBAR}

Table 3.14: Unitary model parameter for heterojunction barrier effect

3.4 Temperature Scaling

All parameters so far have been scaled for variations of the process through apparent geometry changes. However, the model not only has to provide good agreement of simulation and device behavior over a large bias range at nominal temperature. Varying ambient conditions ($T_{amb} \neq T_0$) need to be taken into account by the model as well.

To cover a wide range of operating conditions scalability is extended wrt. temperature. This however is a native feature of the model so the inherent temperature scaling parameters (cf. Tab. 3.15) are given in the model definition. Yet the addition of scaling laws to reproduce the temperature dependence of parameters adds another dimension. In consequence (for standard circuits not operating at cryogenic temperatures) the amount of measurements and validation steps need to be extended within a given range (typically 233K up to 423K or -40C to 125C respectively).

The .TEMP operator is passed upon initialization of the device simulation. This control sets the default simulation temperature. Usually, one references T_0 as the nominal temperature of the device where model extraction is performed. Historically (due to the simple calculus with a value of $T_K = 300K$) the reference ambient temperature for simulation and parameter extraction is defined $T_0 = 27C$ with the thermodynamic temperature (absolute temperature) $T_{abs} = 273.16K$. However this value is not fixed and has been changed to $T_0 = 25C$ for the model parameter extraction for BiCMOS55 technology. Since parameters often-times are scaled using the deviation from the reference temperature (T_0 or T_{ref}), where model parameters are extracted, the definition of a normalized temperature (T_n) is useful

$$T_K = T_{ref} + 273.16K \qquad \text{temperature in K scale} \tag{3.227}$$

$$T_n = \frac{TEMP}{T_{ref} + 273.16K} = \frac{TEMP}{T_K}. \qquad \text{normalized temperature} \tag{3.228}$$

Through the temperature dependence of the thermal voltage (V_T), the electron and hole mobilities (μ) as well as carrier densities (N), the junction temperature of the device influences a multitude of parameters and device properties (such as intrinsic carrier concentration (n_i), carrier lifetime etc.). Quantities affected by temperature variation are therefore included in an electro-thermal model description that reproduces their temperature dependencies. In most cases electro-thermal effects are modeled using an exponent coefficient (Z_{ETA} or ζ), that is introduced specifically for the modeling of temperature dependence (cf. Tab. 3.15).

Furthermore band-gap voltages (V_g) are defined since the junction characteristics are influenced by energy gap temperature equations. In the operating range of interest a linear scaling of the band-

Param.	Description	Param.	Description	Param.	Description
Z_{ETACI}	Temperature coefficient for R_{CI0}	Z_{ETARBI}	Temperature exponent of internal base resistance	V_{GB}	Band-gap-voltage extrapolated to $0K$
A_{LB}	Relative temperature coefficient of current gain	Z_{ETARBX}	Temperature exponent of external base resistance	Z_{ETACT}	Exponent coefficient in transfer current
A_{LVS}	Relative temperature coefficient of saturation drift velocity	Z_{ETARCX}	Temperature exponent of external collector resistance	V_{GE}	Effective emitter band-gap voltage
A_{LT0}	First-order relative temperature coefficient of parameter T_0	Z_{ETARE}	Temperature exponent of emitter resistance	Z_{ETABET}	Exponent coefficient in BE junction
K_{T0}	Second-order relative temperature coefficient of parameter T_0	A_{LFAV}	Relative temperature coefficient for avalanche breakdown Favl	V_{GC}	Eff. collector band-gap voltage V_{GCeff}
A_{LCES}	Relative temperature coefficient of V_{CES}	A_{LQAV}	Relative temperature coefficient for avalanche breakdown Q_{AVL}	V_{GS}	Eff. substrate band-gap voltage V_{GSeff}
D_{VGBE}	Band-gap diff. betw. B and BE SCR	$Z_{ETAHJEI}$	Temperature coefficient for A_{HJEI}	$Z_{ETAVGBE}$	Temperature coefficient for H_{JEI0}
Z_{ETARTH}	Temperature exponent of the thermal	A_{LRTH}	Temperature exponent of the thermal	D_{VGBE}	Band-gap difference for H_{F0} weight factor

Table 3.15: Definition of temperature scaling parameters independent of device geometry

gap energy (V_g) may be sufficient [44, 81], yet a more complicated empirical model is used:

$$V_g(T) = \frac{E_g(T)}{q} \approx V_g|_{T=0} + K_1 \cdot T \cdot \ln(T) + K_2 \cdot T \tag{3.229}$$

$$V_g(T) = V_g T_0 + k_1 \cdot \frac{T}{T_0} \cdot \ln\left(\frac{T}{T_0}\right) + k_2 \cdot \left(\frac{T}{T_0} - 1\right) \tag{3.230}$$

In the model this relation is represented using the parameters f_{1vg} and f_{2vg} (coefficient K_1 and K_2 respectively) which are general simulator parameters.

Given the temperature dependence of the intrinsic carrier density n_i, the transfer current depends strongly on the junction temperature. Hence the ICCR weighting factor c_{10} is modeled as

follows

$$c_{10}(T) \propto V_T(T)\overline{\mu_{nB}(T)n_{iB}(T)^2} \text{ implemented with the scaling equation:} \tag{3.231}$$

$$c_{10}(T) = c_{10}(T_0)\left(\frac{T}{T_0}\right)^3 \exp\left[\frac{V_{Gb}}{V_T(T)}\left(\frac{T}{T_0}-1\right)\right], \tag{3.232}$$

where the band-gap value V_{Gb} is linearly temperature dependent in itself.

Saturation Current Temperature Equations

The saturation currents (I_S) of the quasi-static base current and other contributions are modeled in a very similar manner (with the relative temperature coefficient for the current gain B_f defined α_{Bf} or A_{LB}). The absolute temperature difference used in the scaling of the current gain B_f is defined as $\Delta T = T - T_0$.

$$B_f(T) = B_f(T_0)[1 + A_{LB}\Delta T] \tag{3.233}$$

$$I_{BS}(T) = I_{BS}(T_0)\left(\frac{T}{T_0}\right)^3 \exp\left[\frac{V_{Gb}}{m_B V_T(T)}\left(\frac{T}{T_0}-1\right) - A_{LB}\Delta T\right] \tag{3.234}$$

Capacitance Temperature Equations

The scaling law linking the junction capacitance parameters with the temperature elevation of the device is given through

$$V_D(T) = V_D(T_0)\left(\frac{T}{T_0}\right) - V_{Gj}\left(\frac{T}{T_0}-1\right) - 3V_T \ln\left(\frac{T}{T_0}\right) \tag{3.235}$$

$$C_{J0}(T) = C_{J0}(T_0)\left(\frac{V_D(T_0)}{V_D(T)}\right)^z, \tag{3.236}$$

with V_{Gj} being the effective band-gap voltage V_{geff} for the respective junction.

Resistor Temperature Equations

All resistances are modeled using an exponential temperature dependence as exemplarily defined below for the zero-bias value of the inner base resistance

$$R_{BI0}(T) = R_{BI0}(T_0)\left(\frac{T}{T_0}\right)^{Z_{ETARBI}}. \tag{3.237}$$

The remaining resistances (external base resistance R_{BX}, external collector resistance R_{CX}, and emitter series resistance R_E) follow the same principal scaling rule as R_{BI0}, with linked set of model parameters Z_{ETARBX}, Z_{ETARCX} and Z_{ETARE} attributed to the corresponding transistor regions.

Transit Time and Minority Charge

Since several physical properties of silicon (e.g. carrier mobility μ) scale with temperature, expressions for temperature effects of the forward transit time are needed.

$$R_{CI}(T) = r_{CI}(T_0)\left[\frac{T}{T_0}\right]^{Z_{ETACI}} \tag{3.238}$$

$$V_{LIM}(T) = V_{LIM}(T_0)[1 - A_{LVS}\Delta T]\left(\frac{T}{T_0}\right)^{Z_{ETACI}} \tag{3.239}$$

$$V_{CES}(T) = V_{CES}(T_0)[1 - A_{LCES}\Delta T] \tag{3.240}$$

$$\tau_0(T) = \tau_0(T_0)[1 - A_{LT0}\Delta T + K_{T0}\Delta T^2] \tag{3.241}$$

$$T_{HCS}(T) = T_{HCS}(T_0)\left(\frac{T}{T_0}\right)^{(Z_{ETACI}-1)} \tag{3.242}$$

$$T_{EF0}(T) = T_{EF0}(T_0)\left(\frac{T/T_0}{1 + A_{LB}\Delta T}\right) \tag{3.243}$$

The temperature dependence of the zero-bias hole charge (Q_{P0}) is approximated by

$$Q_{P0}(T) = Q_{P0}(T_0)\left([2 - \left(\frac{V_{DEi}(T)}{V_{DEi}(T_0)}\right)^{z_{Ei}}\right], \tag{3.244}$$

along with the temperature dependence of the principal charge weighting factors

$$H(T) = H(T_0)\exp\left(\frac{\Delta V_G}{V_T}\left(\frac{T}{T_0} - 1\right)\right]. \tag{3.245}$$

The according band-gap voltages describing the temperature dependences of the minority charge weight factors for other transistor regions are defined

- $\Delta V_{GBE}(D_{VGBE})$ for H_{F0}
- $V_{GB} - V_{GE}$ for H_{FE}
- $V_{GB} - V_{GC}$ for H_{FC}

Special attention is given to the temperature scaling of the new model variables used in the charge weighting for the reverse early effect defined through H_{JEI0} and A_{HJEI}

$$H_{JEI0}(T) = H_{JEI0}(T_0)\exp\left[-\frac{\Delta V_{GBE}}{V_T}\left(\left(\frac{T}{T_0}\right)^{Z_{ETAVGBE}} - 1\right)\right] \tag{3.246}$$

$$\text{and } A_{HJEI}(T) = A_{HJEI}(T_0)\left(\frac{T}{T_0}\right)^{Z_{ETAHJEI}}. \tag{3.247}$$

Breakdown

For the collector to base junction the avalanche process limits the maximum voltage the transistor can sustain. The avalanche breakdown related electrical parameters of the modeled bipolar transistor scale with an exponential dependence on the junction temperature as the energy gap in silicon decreases with temperature. Hence the breakdown voltage due to tunneling has a negative temperature coefficient. This can be used to differentiate the tunneling effect from the avalanche mechanism which exhibits a positive temperature coefficient.

The temperature dependence of breakdown and avalanche multiplication is natively taken into account in the HICUM model (cf. eqn. (3.248) and (3.249)). The two model parameters used for description of the avalanche multiplication within the HICUM model each have their own temperature exponent factor A_L [82].

$$F_{AVL}(T) = F_{AVL}(T_0)\exp(A_{LFAV}\Delta T) \tag{3.248}$$

$$Q_{AVL}(T) = Q_{AVL}(T_0)\exp(A_{LQAV}\Delta T) \tag{3.249}$$

3.5 Test Structures for Model Parameter Extraction

The most part of the extraction of model parameters for bipolar junction transistors (BJT) is performed using test structures that are basically identical to a real device that is available to the circuit designer of a selected technology. Yet proper modeling of high-performance bipolar logic arrays with a multi-layer metal back-end is complex and many contributions have to be taken into account.

Generally there are three ways to obtain information about the device:

Firstly a numerical technology computer aided design (TCAD) device simulation may be conducted by use of drawn device dimensions defined in the process route or the device layout. Yet even though these simulations have matured over the years and by now are capable of reproducing the influence of silicon processing, the models behind the simulations are almost always estimations (used models are simplified) and the device dimensions are subject to large variations. Hence TCAD simulation provides a good source for first estimates but on-wafer measurements are preferred. However proper modeling parasitic elements like interconnections, isolations and the substrate is important for determining circuit operation and performance in large designs and for some of these contributions practical test structures do not exist.

A second source of information are dedicated DC or CV test structures (with contact pads for on-wafer characterization), which generally provide very accurate current readings or high capacitance values (by manufacturing parallel arrays of device structures). Especially for the extraction of series resistances (e.g. R_{Bx} and R_{Cx}) as well as accurate measurement of the (small) base current

(I_B) under low bias, these structures are preferred. Compared to the generally used RF structures the DC scribes allow measurement below the typical leakage current of RF setups (bias tees). However, series resistances in wiring have to be respected and due to the connection through DC pads, no information is obtained about dynamic device characteristics.

Therefore a third category of test structures (RF devices) are processed for small-signal characterization. In contrast to DC and CV structures (manufactured in space-economic stripes) these structures are having standardized RF or so called GSG pads. In order to reach good signal separation and shielding RF structures therefore consume significantly more silicon area. Since the used test fixtures have significant influence on the measured parameters of the devices, it is mandatory to deembed them. On-wafer de-embedding methods with open-short dummy DUTs are the most frequently used solution [83]. Several steps of calibration, on-wafer probing (measurement of DUT and dummys) and post processing (de-embedding) have to be executed before a parameter extraction can be performed. Yet the deembedding structures (dummys) corresponding to each device require the same amount of silicon area as the actual device and thus a careful selection of geometries and configurations is required.

3.5.1 Dedicated Series Resistance Test Structures

In order to obtain the external series resistances of transistors several approaches exist. Even though process information (e.g. through PCM data from 4-point probe structures and contact chains) give some good estimates, the most accurate results are obtained from dedicated resistance structures. These allow direct electrical DC measurement of the respective transistor area with high precision SMUs and optimized wiring.

Neither for the emitter nor for the intrinsic collector resistance there is a structure available at time. However there are two different concepts existing for the base and collector resistance contributions. In order to increase accuracy, all these structures use a Kelvin type connection scheme for two of the respective device terminals. This technique allows a contact resistance (R_{cont}[1]) elimination up to the first metal layer (M1) as shown in Fig. 3.30 and 3.31.

Together with other test structures (for DC and CV measurement) they are manufactured in a contact line configuration with 22 pins[2]. For the resistance structures used in modeling 6 individual access pads are required for each device. However to optimize the connection a common substrate terminal may be used for all devices manufactured within one line of DC pads, reducing the number of required external contacts.

[1] leading to voltage drops in access resistances

[2] earlier technologies used layouts with 12 pins per scribe line

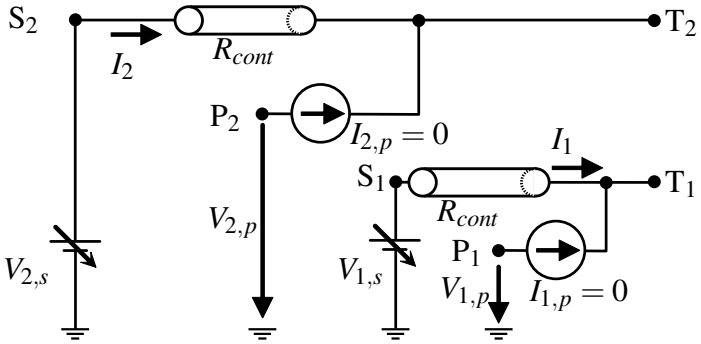


Figure 3.30: Equivalent circuit of Kelvin type measurement with individual force and sense probe per terminal; [*P: probe terminal for accurate voltage reading ($I_P = 0$), S: force port for current supply (I_{meas}), T: device terminal*]

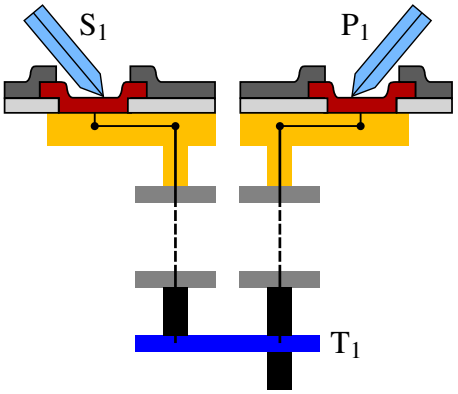


Figure 3.31: Schematic cross section of Kelvin connection in metal layers adapted for two pins (P and S) per terminal (T) shorted in M1

The Dual Base Tetrode Structure

The electron triode is defined as having three terminals (E, B and C). The addition of electrodes to a given diode transforms the normal BJT transistor into a so called tetrode device with a fourth terminal. For bipolar modeling the tetrode test structure, refers to a double base configuration with individual access to the two terminals (B_1 and B_2) and presents an improved approach over previous structures [84]. As seen in the cross section (cf. Fig. 3.33) the individual base terminals are separated by a ring-emitter structure entirely enclosing the inner base contact B_1. Since the device is symmetrical the measured resistances are connected in parallel (cf. Fig. 3.32).

The description of the structure advantages, the mode of operation and the dedicated extraction procedure, allowing to extract both base resistance contributions R_{Bx} and R_{Bi} is given in the respective section 4.4.2.

The Dual Collector Structure

The buried layer test structure was refined compared to previously used approaches as presented in [85]. Instead of a (comparably small) variation of the inner emitter width (w_E), a variation of the dimension of one BL side ($w_{BL,2}$) gives increased flexibility. Compared to the emitter window width scaling approach used so far [78] a higher accuracy can be achieved given the range of variation of the total buried layer dimension where injected current flows.

In addition to the standard extraction procedure, the structure might be used in forward active mode. While limiting current flow through only one of the collector contacts (the second used as sense probe with current forced to zero, $I_{sense} \stackrel{!}{=} 0$) an extraction (or verification) of the BL resistivity from measurements under active transistor operation is possible.

The top view of the device (cf. 3.34) shows the operation principle for resistance measurement.

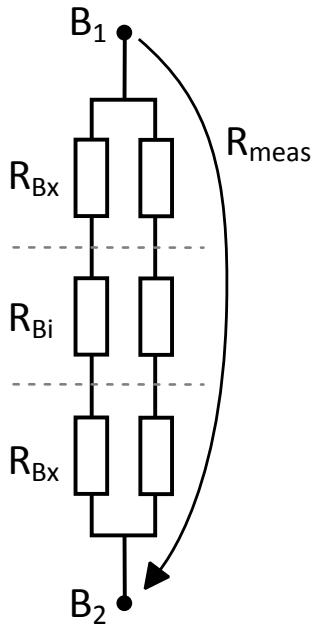


Figure 3.32: Dual base tetrode device equivalent circuit of resistance contributions

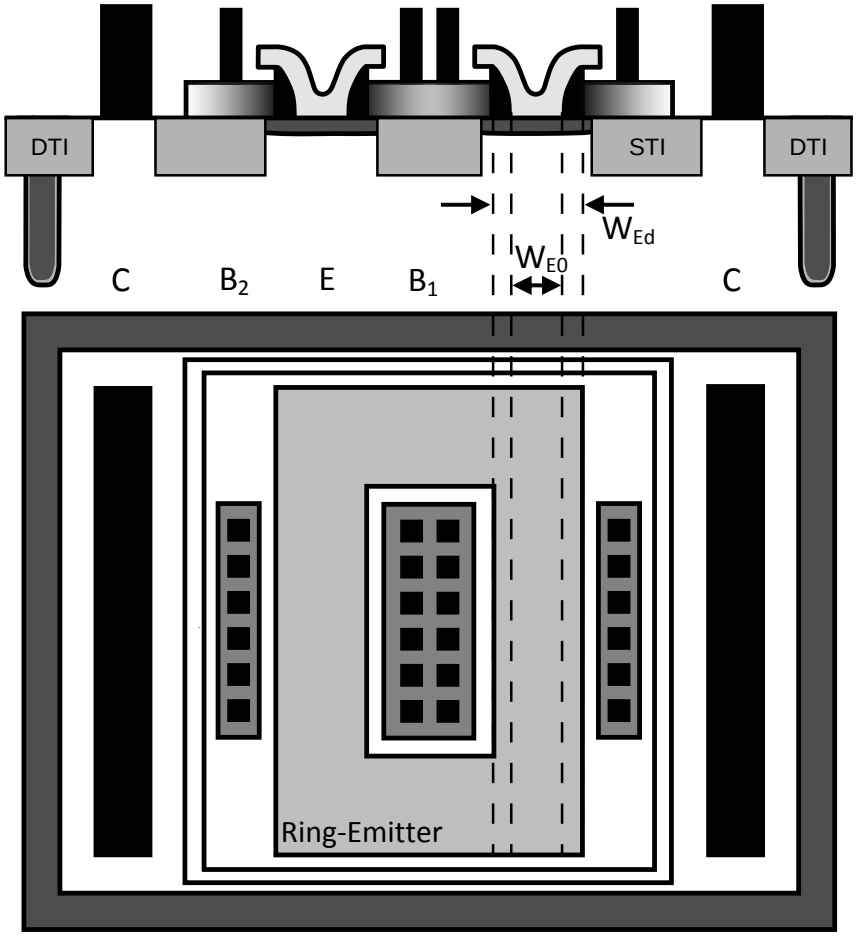


Figure 3.33: Top view and schematic cross section of the ring-emitter dual-base tetrode device with individual access to the separated base contacts B_1 and B_2

In off-mode the current injected into the first collector contact (C_1) flows through the resistive collector sinker (SK) under the contact and the buried layer (BL) and is monitored on the second collector terminal (C_2). The according extraction procedure is described in section 4.4.3.

DC and CV devices

Besides structures for resistance evaluation devices for capacitance-voltage (CV) profiling are manufactured. Groups or matrices of devices are manufactured in parallel in order to yield sufficiently large capacitances[1]. Alongside there are OPEN structures serving for measurement of the (constant) capacitance originating from test setup and pads and most importantly the backend.

The dedicated DC structures serve another purpose. DC SMUs are capable of delivering few

[1] in the order of few pF

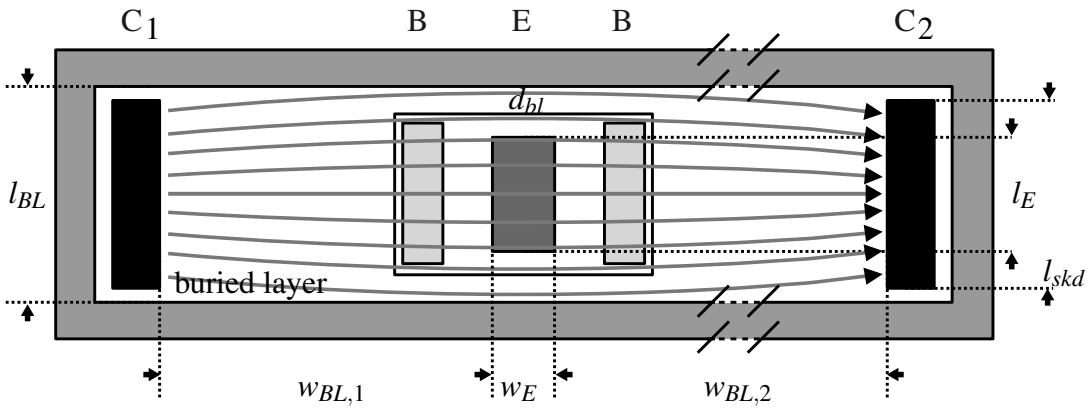


Figure 3.34: Top view of collector resistance (R_C) test structure; current flow in BL between terminals C_1 and C_2; two independently configured BL sides ($w_{BL,1} = w_{BL,nom}$ and $w_{BL,2} = n \cdot w_{BL,nom}, n = 2,3$)

pA measurement resolution[1], when leakage currents of the bias tees in RF setups are avoided. Hence to characterize the small base current contribution under low bias for modeling the non-ideal base current these structures are highly useful. Yet due to higher series contact resistances the use in the high injection range is limited.

RF devices

Standard devices, that can be well characterized, are essential to model parameter extraction. In order to evaluate the true performance of the device in a comprehensive way, RF structures are required. These provide most of the desired characteristics (DC current, capacitance from cold S parameter measurement, device performance in active mode).

As opposed to pure DC structures (with access to all four terminals) certain limitations exist: in oder to provide high resolution, two-port measurement instruments are commonly used. This limits the number of independently accessible device terminals and thus RF devices are manufactured in common emitter configuration, meaning that the emitter and substrate node are tied to the ground while base and collector can be biased independently. External voltage sources are used to establish any of the desired bias conditions for different modes of transistor operation. The substrate node however is not individually accessible and thus the substrate transistor can not be fully characterized individually.

A schematic top view of the used RF structures with corresponding contact configuration is shown in Fig. 3.35. A layout top view with a selection of available devices (different configurations) is given in Fig. 3.36. Due to the multitude of characterization possibilities the most part of geometries and configurations supported in the DK is available in form of RF devices.

[1] as compared to a *nA* range for RF devices

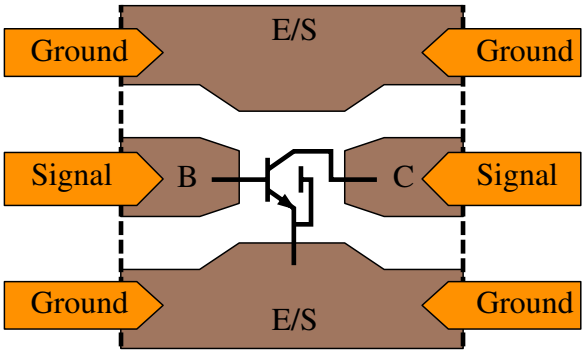


Figure 3.35: Top view of RF device with probe pinning and pad configuration for on-wafer measurement

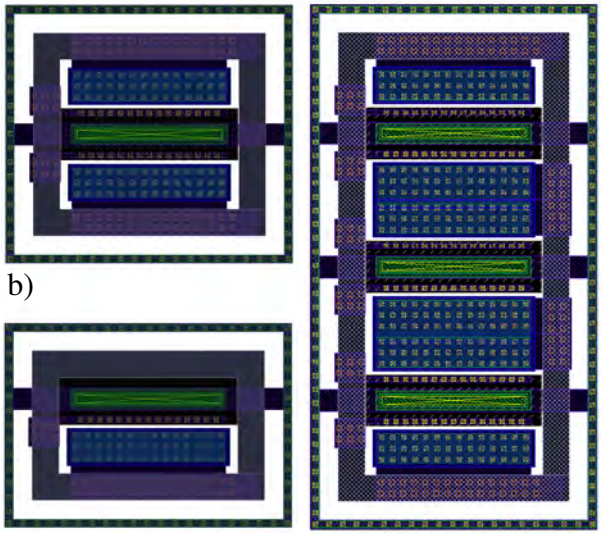


Figure 3.36: Top view of device configurations: a) CBE device, b) symmetrical CBEBC device, c) multi-emitter ($nbe = 3$) device

Area Consumption

The device size of a 0.2x10μm^2 standard symmetrical CBEBC transistor consumes only about 70μm^2 of silicon area[1]. However the required (standardized) RF pads with a minimum separation distance and size for probing call for a multiple of the space (active device area is less than 1% of the total footprint of the structure) making RF extraction very cost-intensive.

A brief comparison of the required space per device layout for the different aforementioned applications is given in Tab. 3.16.

[1] measured between the outside of the substrate contact ring, 12.9x5.2μm^2 (67μm^2)

Device Type	Pad Layout Description	Area Consumption
RF devices	GSG pad configuration with two (80x50μm^2), signal pads and two ground stripes (300x100μm^2)	0.961mm^2 per structure, OPEN and SHORT required
DC devices	DC pads with standard size (70x70μm^2), 5 pads per structure, plus joint substrate	0.229mm^2 for four DC devices $^{(1)}$
CV devices	DC pads with standard size (70x70μm^2), 5 pads per structure, plus joint substrate	0.458mm^2 for four DC devices $^{(2)}$

Table 3.16: Summary of silicon area consumption of elementary test structures for modeling, [$^{(1)}$: no dummies required, $^{(2)}$: one additional OPEN dummy per structure]

3.5.2 Measured Networks on Standard RF Devices

The set of measurements carried out in a characterization campaign is variable and strongly depends on the applied extraction strategies. The measured characteristics are standardized and indicated using a (company-)specific numbering scheme. This convention fixes the applied voltage and current sweeps and steps so that data is comparable between individual measurement campaigns. The NW designation stands for network and indicates the used measurement configuration.

DC Measurements

- NW1: Gummel characteristic (V_{BE} sweep at constant V_{BC})
- NW3: Fwd. output characteristic (V_{CE} sweep at constant I_B)
- NW5: Fwd. Early char. and breakdown (avalanche) msmt. (V_{CE} sweep at $V_{BE} = const.$)
- Thermal measurement of characteristics with reduced number of points at $T \neq T_0$

The corresponding equivalent circuits for the aforementioned three principal DC networks are shown in Fig. 3.37 through 3.39. Fig. 3.40 and Tab. 3.17 show most of the regions the transistor operates in. Generally the forward active mode is used if the transistor is to operate as an amplifier. Logic applications utilize both the cutoff and the saturation modes. The reverse active mode however has very limited application and breakdown has to be avoided by proper circuit design respecting the safe operating area (SOA).

RF Measurements [with given frequency list]

- NW12: Hot S-par. msmt.: [S]-param. in fwd. mode (V_{BE} sweep at $V_{BC} = const.$)

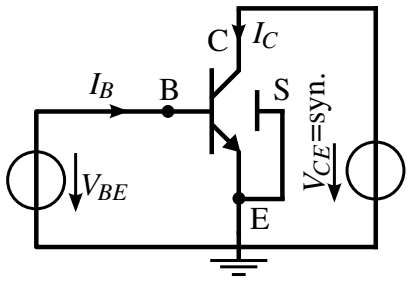


Figure 3.37: Measurement configuration for forward Gummel characteristic; $I_B, I_C = f(V_{BE})|_{V_{BC}}$, NW1

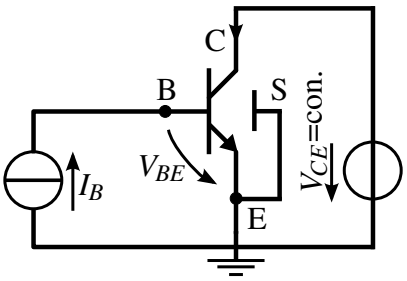


Figure 3.38: Measurement configuration for output characteristic; $I_C, V_{BE} = f(V_{CE})|_{I_B}$, NW3

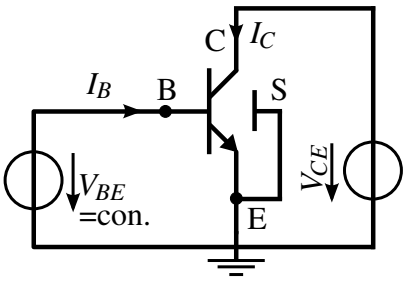


Figure 3.39: Measurement configuration for forward Early characteristic; $I_B, I_C = f(V_{CE})|_{V_{BE}}$, NW5

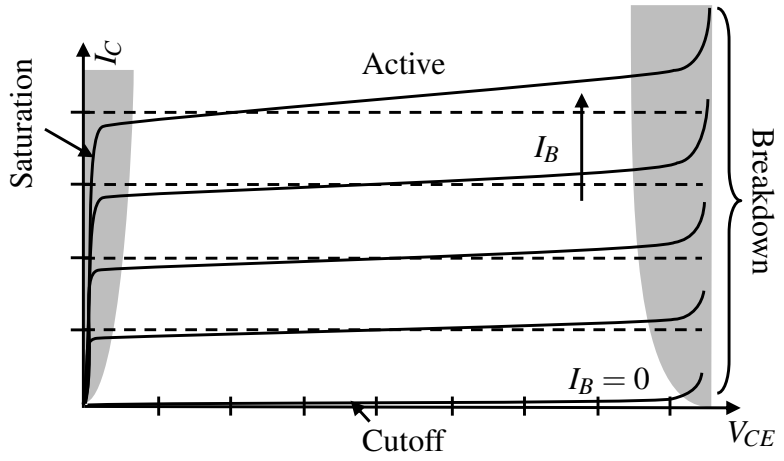


Figure 3.40: Output characteristic in common emitter configuration; *with characteristic transistor operating regions*, $I_C = f(V_{CE})|_{I_B}$

- NW14: Cold S-par. msmt.: for BE and BC junction capacitance (V_B sweep at $V_C = 0$)
- NW15: Cold S-par. msmt.: for BC and CS junction capacitance (V_C sweep at $V_B = 0$)
- NW24: Hot S-par. msmt.: [S]-param. in fwd. mode (V_{BE} sweep at $V_{CE} = const.$)

For small signal analysis of HBTs, one assumes that the device follows an applied AC signal (voltages and currents) quasi-statically. The direction of I_E is defined out of the emitter terminal, in the direction of the hole current and opposite to the direction of the electron current. The corresponding schematic circuit for RF measurement is shown in Fig. 3.41.

The list of measured frequencies is highly dependent on the used instrument. A standard VNA typically measures up to frequencies of $f_{meas} = 50$ or $67GHz$ with commonly used mixers or extensions up to $110GHz$. For most of the needs in device modeling this range is sufficient. The list or sweep of frequencies to measure is typically few points per decade (e.g. 32 points within 0.1 .. $110GHz$).

Operating Region	BE diode	BC diode	Characteristic
Active	Forward	Reverse	Amplifier, $\beta = I_C/I_B$
Cut-off	Reverse	Reverse	$I_C \simeq I_B \simeq I_E \simeq 0$
Saturation	Forward	Forward	$V_{CE} \simeq 0$
Reverse active	Reverse	Forward	limited use

Table 3.17: Operating regions of the BJT; *biasing of junctions and characteristic properties*

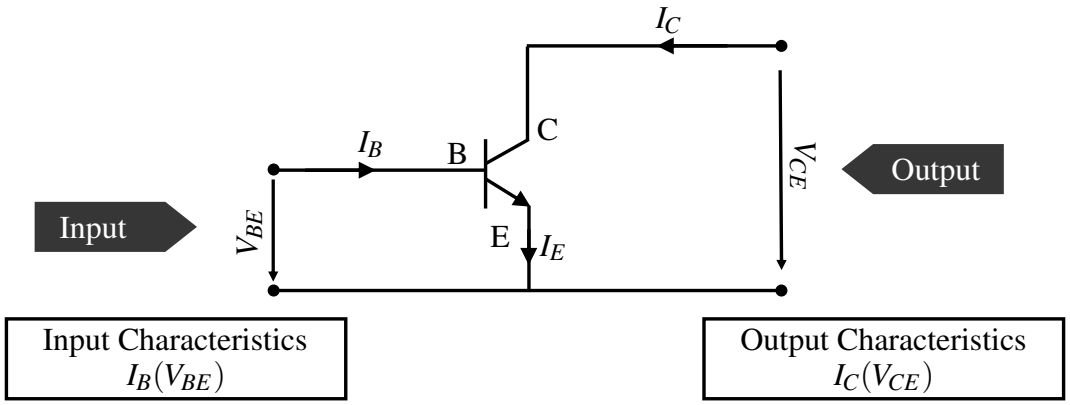


Figure 3.41: Common emitter configuration of BJT

A short summary of the main use of the measured network configurations is given in Tab. 3.18. However given the multitude of extraction steps employed in HBT modeling, that make use of specific portions of measured characteristics, this list is not fully comprehensive.

Network Configuration	Extracted Parameter Set
NW1	Base and collector DC parameters from positive V_{BE}; substrate current from negative V_{BE}
NW3	High current parameters, saturation, reverse mode, avalanche breakdown
NW5	Early voltage and breakdown parameters
NW12	Dynamic (transit time) parameters, emitter resistance
NW14	BE and BC junction capacitance
NW15	BC and CS junction capacitance

Table 3.18: Use of measured networks in parameter extraction procedure

Additional Measurements

Some additional measurements are typically required either using specific instruments. Thus oftentimes these measurements are done in dedicated (external) campaigns and therefore not comprised in the main routines for on-wafer characterization.

- low frequency S-Parameter measurement (for C_{TH})
- noise measurement for Base and Flicker noise
- pulsed measurement to avoid self-heating impact

Even though none of the standard measurements for modeling are destructive, careful sequencing of characterizations may be necessary. Especially for large transistors the device behavior under high current changes due to strong self-heating. For additional exploitation of the safe operating area a measurement of critical bias points may be required. In any case these measurement will likely stress (or break) the device and therefore are recommended to be performed at the end of the measurement campaign.

3.5.3 Measurement Versus Temperature

To gain information about the parameters that model the temperature effects and allow accurate extraction of the linked set of parameters, a precise characterization of the HBT device at various ambient temperatures (T_{amb}) is required. Usually a test station or prober with variable temperature control is used. Once data is acquired there are several options regarding the determination of temperature parameters implicating a more or less complicated extraction procedure:

A commonly used solution builds on a comprehensive first extraction of the reference model parameters (extracted and optimized at T_0), based on which a set of temperature parameters is extracted in another step. This option is an isolated approach, where temperature model parameters are determined without iterative loops nor an adjustment of the nominal parameters.

A second approach is based on the optimization of both parameter sets at the same time. Here both the temperature model and the reference model parameters are included for the optimization. The degree of freedom is highly increased yet due to the multitude of options unphysical values may be obtained.

For extraction of a multi-geometry model, the first option is hence preferred (only temperature model parameters are optimized at $T \neq T_0$). Since characteristics measured over temperature are not exploited in the same detail as those at reference temperature one can reduce the characterization effort. Only a reduced subset of measurements and measured geometries is required as compared to the measurement performed at nominal temperature. As a best practice, the subset of devices measured with a reduced number of bias points comprises one asymmetric device, fully symmetrical (CBEBC) length and width scalable structures (two lengths and two widths) as well as a large multi-emitter device (five finger) for verification purposes.

The selection of the measured temperatures and the measurement range is highly dependent on the intended application of the devices. As a very special case devices operating at cryogenic temperatures for example will be measured at room temperature (300K), as well as the boiling

point of cooling liquids at atmospheric pressure (e.g. nitrogen (77K) and helium (4K)) using cryogenic probe stations. Yet even though the small signal and noise performance of SiGe HBTs under cryogenic temperatures delivers good results, this is a very complicated procedure and only few applications exist.

One of todays typical applications of fast SiGe driven circuits is in the automotive industry. Temperatures may range over a wide spectrum depending on the exact placement of the final circuit in a vehicle [86]. A common range for model parameter extraction and verification is from -40C up to 125C (suitable for underhood application in cars). The according spread of temperatures measured for compact modeling together with the normalized temperature (T_n, cf. eqn. (3.228)) is shown in Fig. 3.42.

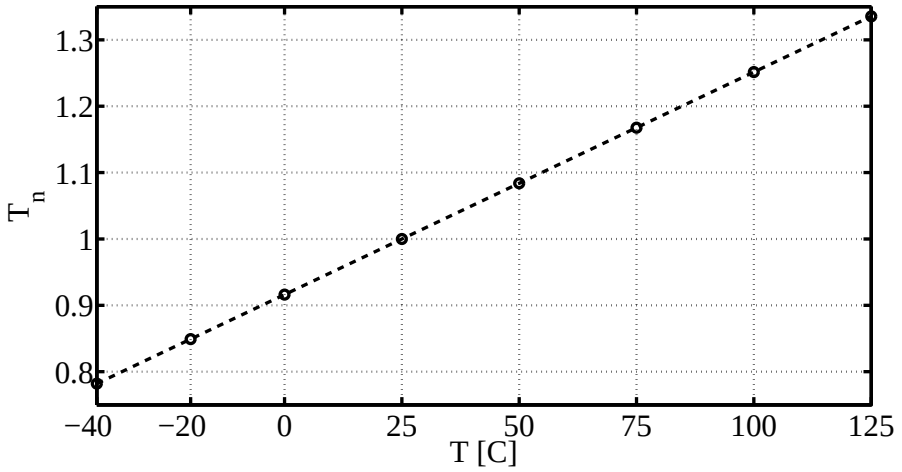


Figure 3.42: Normalized temperature (T_n) and typical temperature measurement steps for modeling

3.6 Conclusion

Today, all existing compact models suitable for advanced HTB devices are defined for a single device geometry (single geometry transistor model). The numerical formulation and implementation in circuit simulators thus lacks native geometry scaling. In order to cover multiple device geometries as well as layout variations, a custom scalable approach has to be added to the model by the foundry or manufacturer. This is done using model-specific scaling equations added as preprocessor in circuit simulators.

This approach leaves flexibility for technology- and foundry-independent, continuous development and improvement of model formulations on the side of the model developers. However the derivation of scaling routines, library implementation as well as appropriate methodologies and techniques for parameter extraction are to be covered by the manufacturer.

The basis of a scalable library for building in-house infrastructure is a general common guideline, applicable to most modern bipolar compact models as well as custom findings for parameter scaling, that is individual to the manufacturer. Using the presented set of equations and together with the model parameters extracted for a given BiCMOS technology, a gap-less geometry offer can be provided to circuit designers.

In this chapter the physical device dimensions as input parameter to the model library have been demonstrated as the origin of parameter recalculation for the native HICUM model equations. It has been shown that process-specific properties such as emitter periphery effects and the effective emitter area are covered by the used implementation. Often-times the possibility of a clear separation into internal and peripheral transistor has been shown, which relates effects closely to device physics, making parameter extraction easier.

Given the strong physical background of the HICUM model, most scaling equations have been categorized by the physical effect they are attributed to or a specific transistor region where they are effective. It has been shown, how the relevant process specific parameters such as dimensions of the extrinsic device architecture, sheet resistances, capacitances etc. are combined with design rules and information about the transistor configuration (e.g. emitter window dimensions as well as location and number of E, B and C contacts) to obtain the model parameters for circuit simulation. The re-calculated model parameters may then be passed to the EDA tool for simulation of a single transistor with an effective electrical emitter.

In the next chapter therefore the linked extraction strategies and other related procedures for model parameter determination, suitable for SiGe HBT devices are presented. However, due to the technology-specific individuality of the scaling laws, most of the presented methods are specifically tailored for the scalable model equations defined in this chapter.

Chapter 4

Geometry Scalable Model Parameter Extraction

4.1 General Model Parameter Extraction Flow

Regardless of the device model, the principal extraction flow for bipolar model for parameter determination is following some common, general steps. These can usually be sub-divided in three basic procedures as shown in Fig. 4.1: in a first step the junction capacitances, intrinsic and series resistances as well as other parasitic elements are determined. In a second step, the fundamental low-current parameters are extracted from DC measurement based IV characteristics, where after the high-frequency measurements are used in a third step to extract parameters required for the proper modeling of dynamic characteristics. The extraction procedure overview provides an idea and a best-practice guideline tailored to the specific needs within a company (STMicroelectronics). Although this list is not meant to be fully comprehensive, it gives a clear indication which steps need to be performed. Most of the steps shown in the schematic flow diagram (cf. Fig. 4.1) can be sub-divided further and will be discussed in detail in the subsequent sections.

Model Extraction Strategies

A desired approach for determining model parameters is a mathematical simplification of the model equations as far as possible. The so called direct parameter extraction therefore presents a closely physics related parameter extraction strategy and is the most sophisticated form. In contrast, optimization-based strategies use the full (e.g. non-linear) model equations to perform numerical optimization. However, they may find local optima. Over the years and with advancing device technology, many approaches for an accurate and physically oriented parameter extraction suited for HBTs have been suggested [87–91].

Optimization is used when there is no proper analytical expression available or if a direct

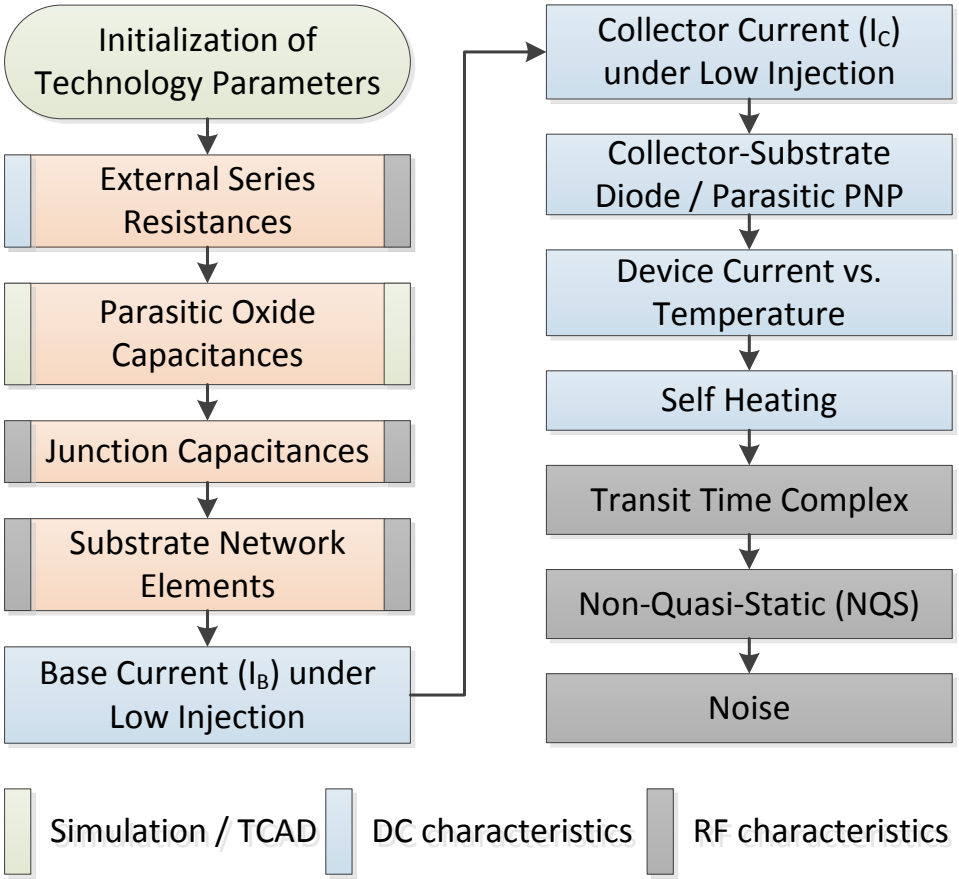


Figure 4.1: Principal flow diagram for bipolar transistor modeling highlighting the sequences of important extraction steps; *indication of origin of data for extraction*

extraction method is too difficult. In general, one differentiates two optimization strategies for parameter extraction referred to as *global* and *local* optimization.

For *global optimization* computer algorithms for error minimization are used to find one set of model parameters that will best fit the measured data. This methodology will hence give the minimum average error between measured and simulated data regardless of bias ranges or importance of certain effects since it treats each parameter as fitting parameter. A clear disadvantage of this method is that the extracted set of physically-based parameters in the model might not be consistent with their physical intent.

In contrast the *local optimization* approach is a more straightforward approach. As far as possible parameters are extracted independently of one another. With the knowledge of device physics and operating principles, bias ranges corresponding to dominant physical mechanisms are specifically selected. In turn, the set of parameters extracted with a local approach might not fit experimental data in all the bias conditions. However this method allows to develop extraction methodologies that are specifically tailored to a given physical meaning of a model parameter. If

properly executed, this approach should predict device operation quite well and model parameter values extracted in this manner will exhibit a strong physical relevance.

In a final phase of the model extraction procedure, the extracted parameters may be optimized using the properly initialized (extracted) parameters. Optimizing a nonlinear function by means of an iterative algorithm starts from some initial value of the argument and repeatedly calculates the next value according to rules and determined sensitivities[1] until an optimum is reached. The optimization for device modeling is usually performed using either a Newton-Raphson method for standard problems in optimization or a nonlinear least squares fitting approach.

With efficiency and physical meaningfulness for increasingly complex model equations in mind, optimization as well as direct extraction can only determine a few parameters at a time. Therefore model parameters are usually divided into subsets, so that each subset corresponds to a transistor part or electrical characteristic, from which the parameter can be obtained.

The global parameter extraction process can usually be divided in two stages: a first part is generally devoted to obtaining extrinsic elements of the equivalent circuit, followed by a second extraction for the intrinsic transistor. In turn before performing the extraction of the intrinsic equivalent circuit model, the values of the parasitics must be determined.

In general the model extraction flow comprises the following elements:

- Extraction of resistances (base link and intrinsic base, emitter, external collector)
- Extraction of junction capacitances and substrate network using cold S-Parameters
- Base- and transfer-current characteristics using DC measurement of symmetrical transistors
- Extraction of transit time parameters using hot (active) S-Parameters

4.1.1 Initialization, Extrinsic Elements and Junction Capacitances

The correct assessment of a first group of model parameters is essential for all the following extraction steps. Technology parameters and parameters of extrinsic elements are extracted in a first step. Since the HICUM model formulation is based on an extended and generalized ICCR approach the depletion charge calculation from applied bias voltage is a fundamental basis of the model. Thus junction capacitances are extracted to build a reliable basis for charge calculation.

The initialization step comprises the definition of technology parameters influencing both geometry (spatial dimensions from process) as well as electrical parameters. Either measured data (TEM pictures for dimensions, resistance test structures for sheet resistances) or results from numerical simulation (TCAD) is entered and provided to the model library independent of device geometry. The required information is summarized in Tab. 3.2 and 3.3.

Additionally, geometry correction factors are determined. These account for non-ideal process

[1] for optimization of multiple parameters at once

scaling and became increasingly important in recent technology generations. For the scalable model used in this work one precisely adjusts the two independent offset values dw_e and dl_e based on the DC current characteristics in from the forward Gummel characteristic ($V_{BC} = 0$).

Thereafter the first direct extraction procedures are executed. Depending on the actual procedure the extraction of parasitic circuit elements based on dedicated test structures or measurement techniques are performed. A number of different ways exist to obtain these essential information such as sheet resistances (r_S) and normalized junction capacitances (C_j).

A specialty of these first steps summarized in Fig. 4.2 is their independence on other extraction steps. After initialization of the basic dimensions and technology information, each one can be performed ind ent data becomes available.

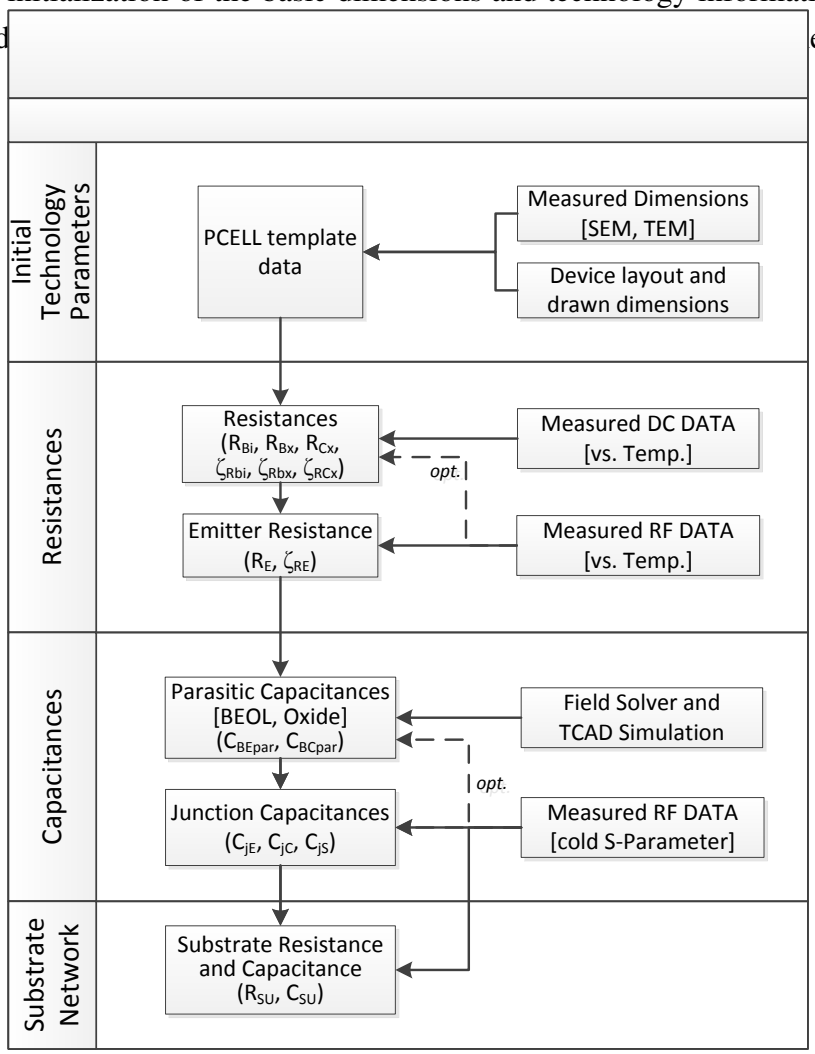


Figure 4.2: Initial extraction steps for bipolar transistor modeling for assessment of parasitic elements and capacitances

4.1.2 DC Parameter Extraction

After extraction of junction capacitances and parasitic elements the extraction of low-current parameters from measured IV curves is started. The current in base and collector is analyzed in different extraction steps dedicated to device operation (cf. Fig. 4.3). DC parameter extraction is linked to the data obtained from direct current electrical measurement steady-state. In this part of the extraction flow, all required quantities (or model parameters) can be obtained from various bias conditions under DC operation.

In general for all active mode measurements the positive V_{BE} voltage causes the p-type base to be higher in potential than the n-type emitter, thus forward-biasing the emitter-base junction. The (negative) collector-base voltage V_{CB} causes the n-type collector to be higher in potential than the p-type base, thus reverse-biasing the collector-base junction. The forward bias on the emitter-base junction will cause current to flow in the BE junction consisting of two components: electrons (e^-) injected from the emitter into the base, and holes (h^+) injected from the base into the emitter.

The DC extraction flow presented in Fig. 4.3 shows the individual extraction steps along with the according set of model parameter, that are determined from each measured characteristic.

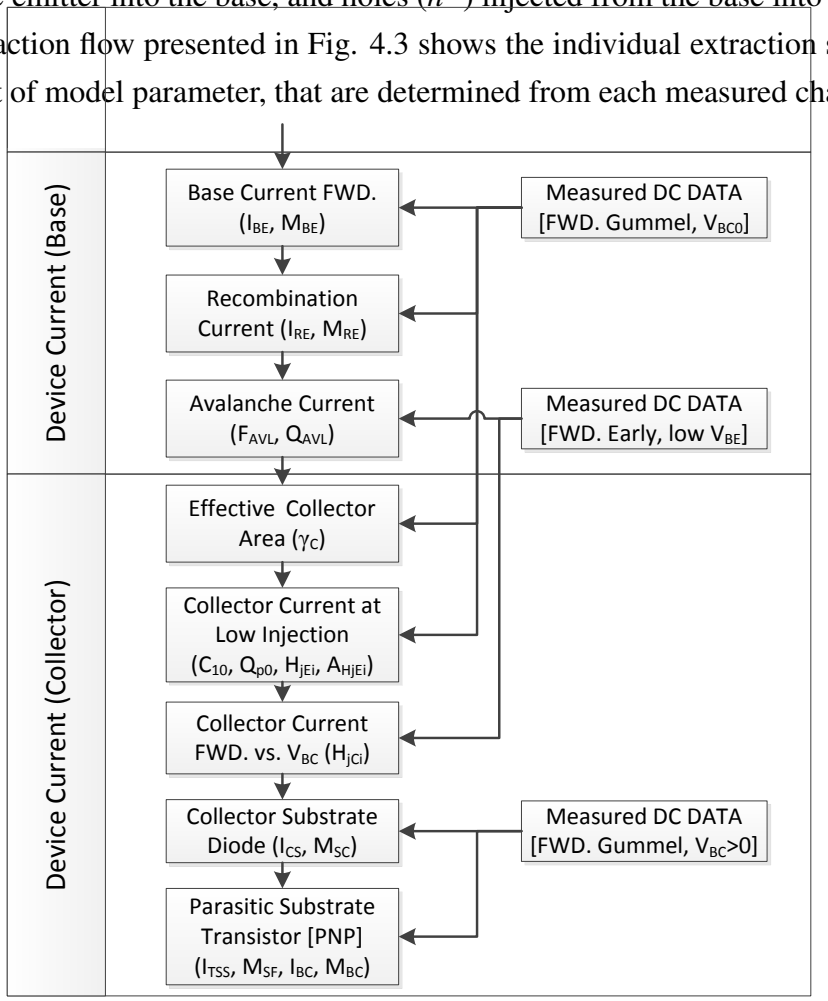


Figure 4.3: DC extraction flow for bipolar transistor in a recommended order

4.1.3 RF Parameter Extraction

The extraction of high-frequency parameters from measured data is very complex and the extraction flow shown in Fig. 4.4 only shows one example or a best practice. However, in these application-critical steps, frequent loops and involved parameter adjustments are required.

Even though most of the data used for assessment of high current characteristics and RF modeling is obtained from AC analysis by means of S-parameter measurement, the strong link of DC and RF characteristics as key feature of the HICUM model requires DC data for verification and extraction. In addition, some supplementary information is required for accurate noise modeling by means of direct noise measurement data. As for junction capacitances and resistances accounting for the variation of model parameters with ambient conditions requires measurement data taken at different device temperature values.

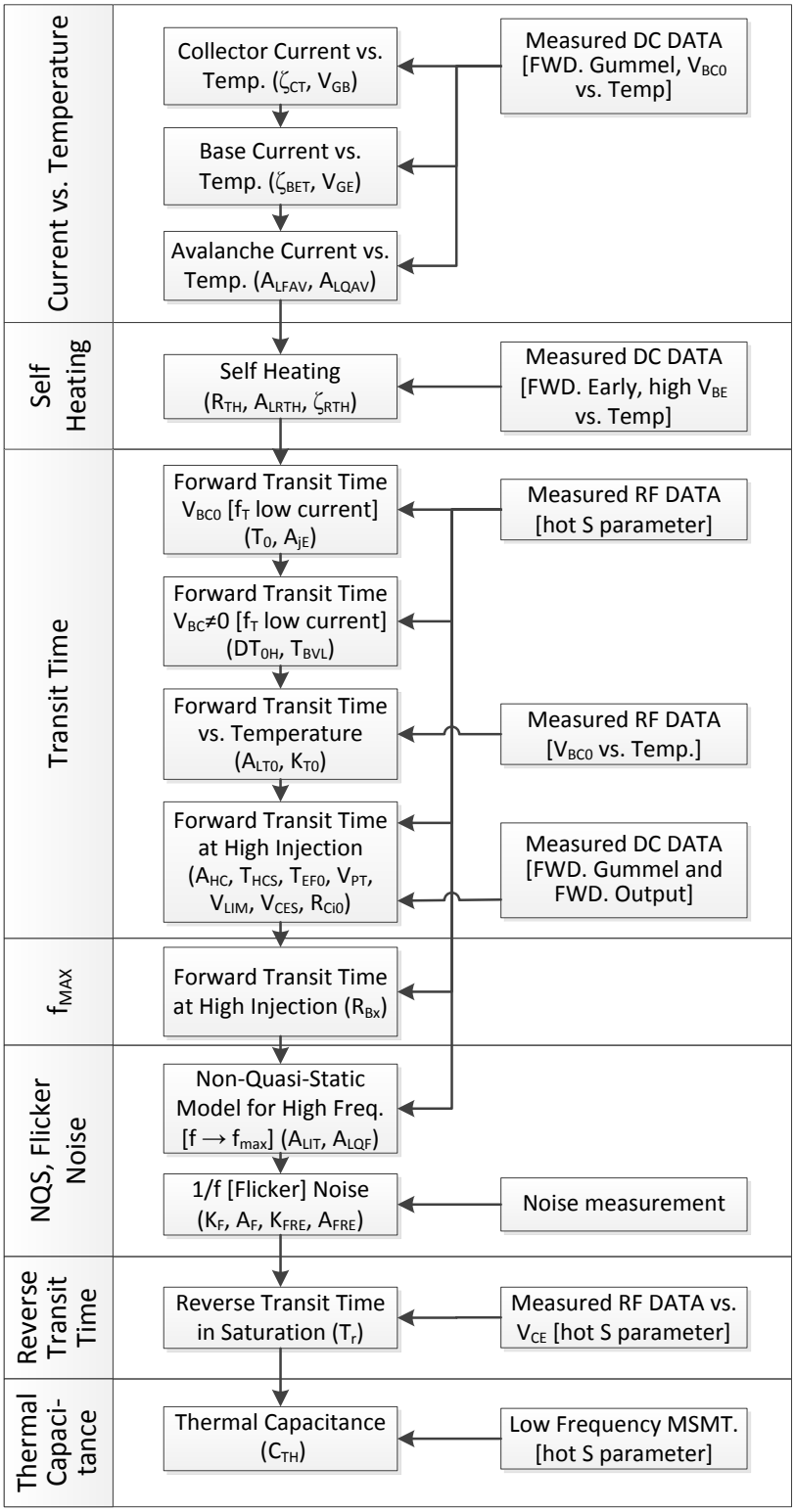


Figure 4.4: RF extraction flow for bipolar transistor

4.2 Scalable Model Extraction

Given the multitude of effects and rules covered by a physics-based geometry scalable model, there is a high number of different model parameters to be determined. In consequence the methodologies for parameter estimation are highly parameter-specific and complex. A general flow-chart is highly dependent on the personal preferences of the model engineer, the given working environment and may contain several loops where characteristics are revisited after the extraction (and change) of model parameters on a different characteristic or operation region respectively. Yet in this section the recommended (general) parameter extraction procedure used for the HICUM model is defined.

Even though the model version update to HICUM L2.3x comes along with a number of improved formulations (cf. section 2.5) that require modified extraction routines, a wide range of existing strategies can still be applied. Yet some new extraction methods were proposed in order to solve new interrelations of parameters.

In a first initialization step the model engineer takes care of all process-inherent information that do not require a dedicated extraction from (electrical) measured data (design rules and process specific parameters such as sheet resistances, doping, back-end capacitances etc.).

Before an extraction campaign based on experimental or measured data can be started, proper design and the measurement of test structures is required. Constraints such as consumed wafer real estate and measurement effort and complexity have to be considered. Upon silicon availability a variety of different electrical measurements are performed on devices to be modeled, in order to get a fully comprehensive view of their behavior under most of the operation modes used in chip design. Generally measurement instruments are connected to a workstation where the results are stored. With a large set of characterizing data files the modeling engineer will typically start the parameter extraction and data fitting process tailored to the model formulations and eventually obtain an individual set of parameters.

In turn the required steps can be synthesized as follows:

- Complete characterization of a multitude of devices (DC to RF and verification of the results)
- Parameter extraction and fitting using the measurement data
- Device description (design kit) using extracted parameters and model verification

4.2.1 Device Geometry Selection for Extraction

Different types of device layout are available and most of the electrical characteristics are a function of both geometry and bias. Thus in a first step suitable geometries for parameter extraction have to be selected from the available data-set.

The device scaling regarding geometry is twofold: for one the device can vary wrt. the lat-

eral geometrical dimensions while the second option is a variation of the connection scheme or configuration respectively. Bipolar transistors connected from one side exclusively (single-sided base and collector) are denoted using the order of their contacts in the device cross section. An Emitter followed by the Base connection and the Collector therefore results in a CBE abbreviation. This configuration provides the smallest footprint possible and thus is advantageous when high integration is desired.

For most RF applications though a single-sided configuration is not favorable due to the highly asymmetric current flow resulting in high series resistances causing increased voltage drops and current crowding in the intrinsic device. Hence for single emitter transistors it is very common to use a two-sided base connection (wrapped around the emitter). This structure is commonly denoted BEBC.

Additionally a third option is a two-sided collector design (wrapped collector) is beneficial for symmetrical current distribution and a low collector series resistance due to the doubled contact denoted CBEBC. In addition there are dedicated multi-emitter devices. Those are favorable if high currents are required or to reduce switching time and power dissipation in logic arrays. In the used process two or more symmetric CBEBC devices are simply joined by a BEBC extension giving the simple formula C-$n\times$(BEBC) for their denomination with n being the number of desired parallel emitter fingers.

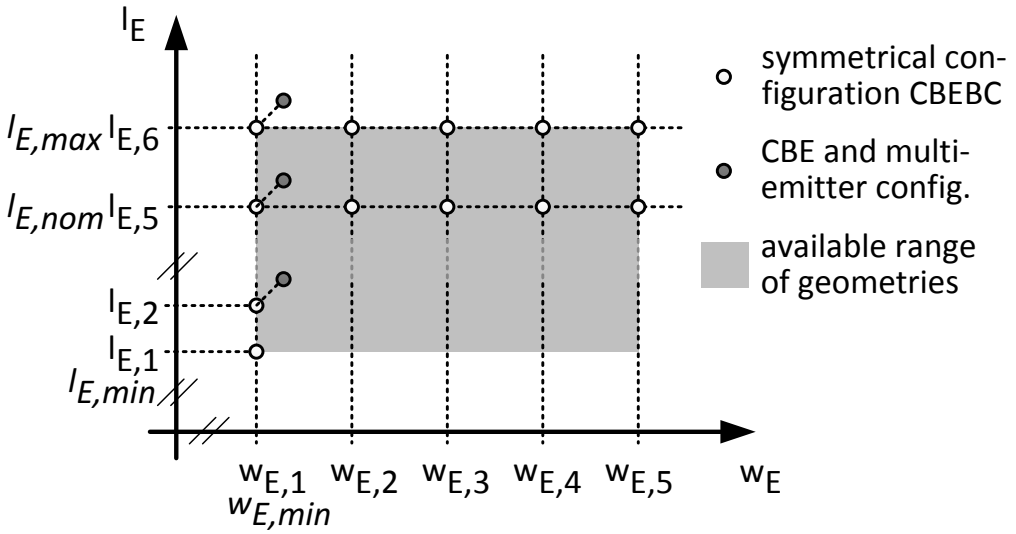


Figure 4.5: Principal device geometries available for model parameter extraction, [*discrete points: RF devices on the test mask for model parameter extraction, grayed area: device geometry offer to designers supported in the PDK*]

The summarizing schematic view (cf. Fig. 4.5) highlights the device offer that can be used for geometry scaling. In addition to a scalable width and length of the devices, to vary the emitter

area, multiple finger geometries (multi-emitter transistors) are available and taken into account in the model library.

As can be seen in Fig 4.5 there are several axes for geometry scaling. This principle has already been used for several device generations and thus presents a best practice approach. The main device type used for extraction is the symmetrical CBEBC device. As mask cost raises, the necessity for a minimum set of devices is given, without sacrificing accuracy.

For the minimal emitter window width (w_{min}) the full range of length-scalable devices is available in the RF block. Furthermore for the nominal device length ($l_{E,nom}$) and the largest emitter stripe structures ($l_{E,max}$), a variety of different device widths ($w_{E,min}$ .. $w_{E,max}$) is available. In total this covers a wide range of the full offer provided to designers in the PDK in order to have a sufficient number of devices for parameter extraction as well as model validation.

Being the ideal transistor with low access resistances (wrt. B and C terminal) and acceptably increased parasitics, the fully symmetrical HBT device is the predominant choice for device modeling and model parameter extraction. The CBEBC devices available on the test mask feature a broad range of variation of the lateral device dimensions (cf. Tab. 4.1).

W / L	0.20μm	0.25μm	0.30μm	0.35μm	0.42μm
0.45μm	NS122A045				
0.60μm	NS122A06				
1.00μm	NS122A10				
3.00μm	NS122A30				
5.00μm	NS122A50	NS122B50	NS122C50	NS122D50	NS122E50
10.0μm	NS122A100	NS122B100	NS122C100	NS122D100	NS122E100

Table 4.1: Set of single-emitter high-frequency transistors available for model extraction in symmetrical CBEBC configuration with their drawn dimensions (w and l) in lateral direction;
Nomenclature: N: *N*PN transistor; S: high *S*peed transistor; *Numbering*: N° of emitter [1^{st} pos.], base [2^{nd} pos.] and collector [3^{rd} pos.] contacts defining the device configuration, letters designate the emitter window width (w_E), numbers indicate emitter window length (l_E, [in 0.1nm])

Yet, for some applications, other structures (different from the CBEBC configuration, cf. Tab. 4.2) are required and thus provided to designers in the DK. In turn configurations different from the single-emitter fully symmetrical CBEBC structure need to be covered by the device model library.

However since these structures are not predominantly used for extraction but rather for model verification there is only a reduced sub-set of geometries available[1].

[1] e.g. for the BiCMOS55 technology a variation of the device configuration is only available for the minimal emitter window width (w_{min})

cfg. / L	CBE	CBEBCBEBC	C+(3xBEBC)	C+(4xBEBC)	C+(5xBEBC)
0.45μm	NS111A045				
3.00μm	NS111A30	NS243A30	NS364A30	NS485A30	NS5106A30
5.00μm	NS111A50				
10.0μm	NS111A100	NS243A100	NS364A100	NS485A100	NS5106A100

Table 4.2: Set of high-frequency transistors at fixed lateral width $w_E = 0.20\mu$m available for model extraction and verification for various configurations with their respective drawn dimensions (w and l) in lateral direction; Nomenclature: N: *N*PN transistor; S: high *S*peed transistor; Numbering: number of emitter [*nbe*, 1^{st} pos.], base [*nbb*, 2^{nd} pos.] and collector [*nbc*, 3^{rd} pos.] contacts

4.2.2 Transistor Selection Using the Aspect Ratio

In total the set of RF devices comprises more than 20 different layouts. Yet for parameter extraction, taking all of them into account in each extraction step is confusing. In addition the large amount of measured data increases the computing resources for data handling and significantly increases simulation time. In order to select a meaningful set of devices for extraction the definition of the aspect ratio is beneficial.

The aspect ratio (AR) is defined as the relation of the peripheral distance P relative to the surface A. For general emitter window scaling approach results in the following equation

$$AR = P_E/A_E = \frac{2 \cdot (w_E + l_E)}{w_E \cdot l_E}. \tag{4.1}$$

Since many of the parameters in the transistor model scale with both these spatial definitions, a proper selection of devices suitable for geometry-scalable parameter extraction needs to be done. For obtaining reliable unity parameters of surface as well as perimeter component of the respective model parameter the AR is to be selected carefully. A large variation of the relation of A and P allows to perform a linear regression analysis over a wide range of the x-axis (P/A).

As a best practice modeling structures are manufactured with one 'row' of devices at constant lateral emitter width w_E and the full range of lateral lengths l_E. Additionally it is desirable to have a number of various lateral widths (w_E) for one (or two) selected length(s). Additional devices might be large structures (e.g. a square configuration with $w_E = l_E$) however those are not necessarily required and may even be undesired for extraction of high current parameters due to strong self heating and other effects.

When using several different device sizes, plotting the normalized quantity or model parameter (Par$/A_{E0}$) as a function of the aspect ratio (P_{E0}/A_{E0}) ideally gives a linear dependence for both spatial scaling dimensions (width w and length l) when a standard geometry scaling approach is

applicable for the selection of devices. The corresponding linear regression formula (cf. eqn. (3.1)) normalized by the area A (for plotting cf. Fig. 4.6) then reads

$$\frac{\mathrm{Par}}{A} = \mathrm{Par}_A + \mathrm{Par}_P \frac{P}{A}. \tag{4.2}$$

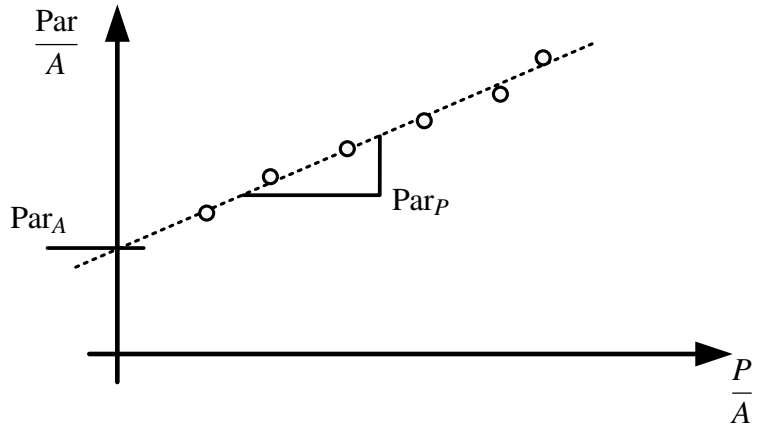


Figure 4.6: Principal extraction from P/A regression for scalable modeling

In order to extract the respective parameters from this regression formula one hence uses a simple direct extraction strategy with linear regression:

- Par_A - extraction from Y-axis intercept ($P_{E0}/A_{E0} = 0$)
- Par_P - extraction from curve slope

Depending on the selection of a sub-set of devices used for scalable parameter extraction the geometry variation results in a very favorable aspect ratio for the variability of one dimension. The preferred options are a variation of the device width at constant length and vice versa. Theoretical results (indicated by dashed lines) are computed along with discrete points representing the actual devices available for parameter extraction (cf. Fig. 4.7 through 4.9). Especially a variation of the lateral window width w_E provides the large range of AR variation suited for geometry scalable parameter extraction (cf. Fig. 4.7).

For both dimensions of $l_E = 5\mu m$ and $l_E = 10\mu m$ the variation of the AR is within a favorable range. A less optimal but still usable variability of the AR is observed for a scaling at fixed emitter window width (which is limited to a single value of $w_E = 0.2\mu m$, cf. Fig. 4.8).

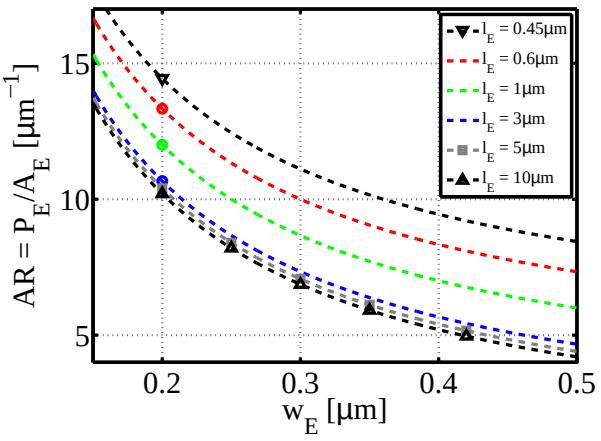


Figure 4.7: Aspect ratio (AR) vs. drawn emitter window width w_E for symmetrical devices

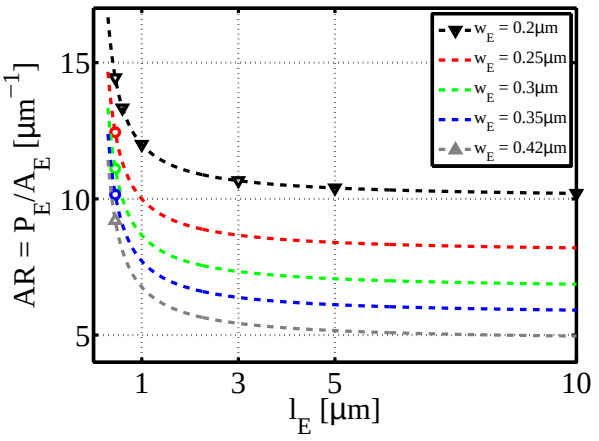


Figure 4.8: Aspect ratio (AR) vs. drawn emitter window length l_E for symmetrical devices

In contrast one can clearly see the multi-emitter devices available in the DK are not well suited for parameter extraction. The variation of the aspect ratio is negligible (cf. Fig. 4.9) hence these devices are less suitable for direct extraction. Yet even though multi-emitter devices are not suited for extraction purposes, where they present important references for model verification.

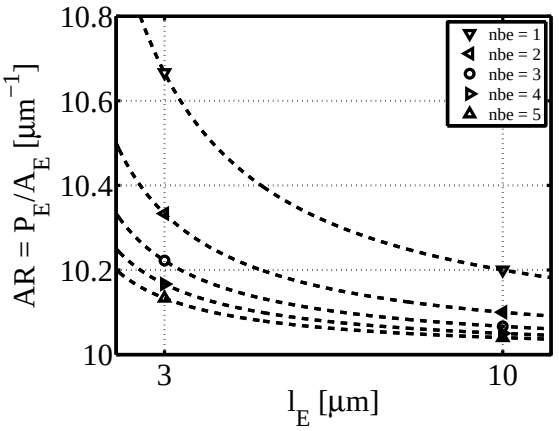


Figure 4.9: Aspect ratio (AR) vs. drawn emitter window length l_E for multiemitter devices (*nbe*)

4.2.3 Software Environment for Modeling and Parameter Extraction

The IC-CAP Device Modeling Software suite developed and maintained by Agilent Technologies presents an industry standard for DC and RF semiconductor device modeling throughout the device types (e.g. passives, MOS and BJT). Together with the Agilent Advanced Design System (ADS) it performs all essential tasks as data handling and storing as well as circuit simulation. IC-CAP thus has become one of the main frameworks for the development of a variety of extensions

dedicated to semiconductor device modeling.

The instrument drivers required for controlling measurement equipment and routines for characterization are part of the IC-CAP package. Thus it is equally suited for DC and RF measurement. Additional sub-program packages such as the IC-CAP Wafer Professional (WaferPro) toolkit [92, 93] provide the essential capabilities for automated measurements. In the IC-CAP WaferPro interface (cf. Fig. 4.10), test plans for automated measurement of semiconductor wafers can be defined ahead of time and executed once in the laboratory. Using user-defined test routines, test plans can be customized to specific characterization tasks and executed for unattended measurement campaigns. For example a set of DC characterization routines for measurement of dedicated resistance test structures can be defined so that all the required data for parameter extraction (including measurement data over temperature) is automatically measured at once.

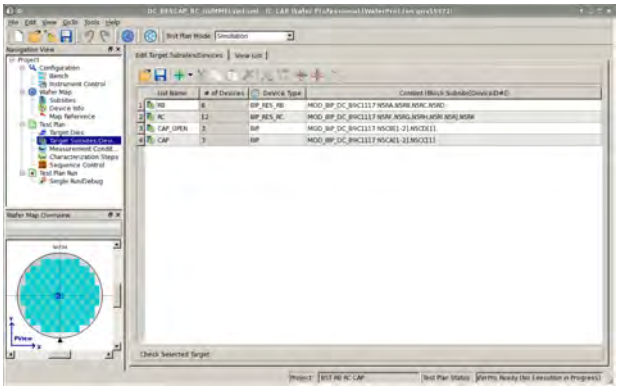


Figure 4.10: Setup screen in Agilent WaferPro software suite dedicated to (semi-)automated on-wafer characterization for device modeling

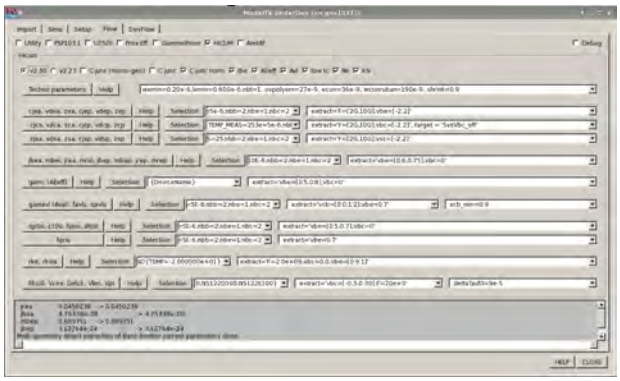


Figure 4.11: ModelToolKit (ModelTK) extraction software main screen for model parameter extraction with pre-defined (custom) routines

The measured data recorded with the WaferPro is fully compatible with and tailored for parameter extraction in the modelTK user interface (cf. Fig. 4.11). By default IC-CAP features parameter optimizers and flexibility to define custom extraction routines based on the imported data. Yet a powerful custom-designed environment (ModelToolKit / ModelTK) features automated data import, device and bias selection through a convenient user interface. This framework provides standardized extraction routines to the model engineer based on custom program code (based on PEL[1] and Python scripts) running in the background.

4.3 Basic Geometry Correction for Process Scalability

Advancing technology and ever-decreasing spatial device dimensions cause some significant problems in silicon-processing such as doping non-uniformities or differences between intended

[1] IC-CAP proprietary Parameter Extraction Language

(drawn) and real lateral dimensions etc. [94]. For this reason corrective terms taking into account geometry related deviations have been implemented for the calculation of w_{E0} and l_{E0} as shown in section 3.1.2 (cf. eqn. (3.14) and (3.15)). To determine their respective values, measured electrical characteristics of the collector current I_C are analyzed. Due to the influence of the lateral device dimension on subsequent extraction steps, this procedure shall be done at first since changes to the determined parameters will have a serious impact on any geometry scalable parameter.

In order to have a uniform geometry scaling, all symmetrical devices in CBEBC configuration are selected. Selecting the forward Gummel characteristic at V_{BC0} at nominal (room) temperature gives a large V_{BE} bias range to choose from. To cancel out high injection effects, one selects a bias range at low current. Keeping in mind the limitations of measurement equipment (such as instrument resolution and leakage currents) a bias of $V_{BE} = 0.6V$ and $V_{BE} = 0.65V$ is selected where the measured current is still accurately measured.

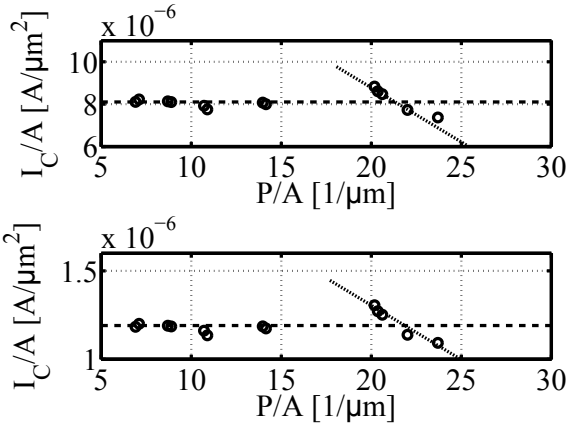


Figure 4.12: Normalized collector current I_C/A vs. aspect ratio P/A without geometry correction ($dw_E = dl_E = 0$) for symmetrical devices

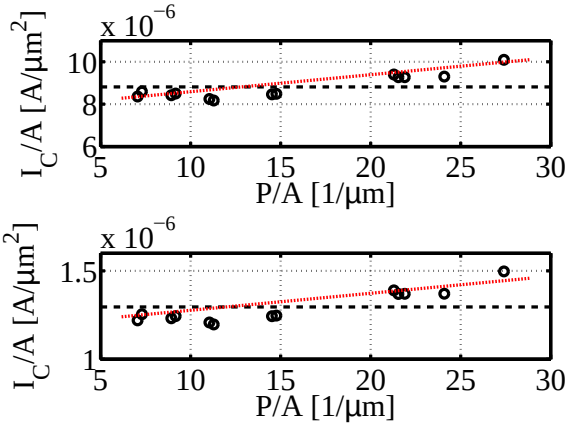


Figure 4.13: Normalized collector current I_C/A vs. aspect ratio P/A with geometry correction ($dw_E = 5nm,\ dl_E = 130nm$) for positive slope

As can be seen in Fig. 4.12, the aspect ratio P/A without geometry correction ($dw_E = dl_E = 0$) already shows a good trend for a variation of w_E in small devices. However given the curve progression (negative slope) for larger devices the geometry scaling needs a correction in lateral l_E direction. A reasonable compromise thus is to be determined experimentally.

In comparison a minor modification of the parameters ($dw_E = 5nm, dl_E = 130nm$) shows the impact on both the aspect ratio as well as the normalized transfer current (cf. Fig. 4.13). However if process scalability is found to be insufficient in later extraction steps, a re-adjustment may be necessary.

4.4 Resistances

Contact resistance is defined as a parasitic resistance. Electrical connections must be made between any semiconductor device in integrated circuits, and the outside terminals. External contacts must be able to withstand temperatures and be compatible with conventional device processing techniques. Usually the device connections are made via ohmic contacts defined as a low resistance junction providing conduction in both current directions between the metal and the semiconductor [95, 96]. The contact resistance thus is a measure for degradation of current flow across a metal-semiconductor interface.

The ohmic contact in particular is defined as a metal-semiconductor contact that has a negligible contact resistance relative to the bulk resistance of the semiconductor material. The respective figure of merit of an ohmic contact is the specific contact resistance (ρ_c, independent of contact area) measured in units of $[\Omega \cdot cm^2]$. However metal-semiconductor interface resistance is desired to be small compared to the resistances of active device and the formation of low resistance ohmic contacts to the device.

The voltage drop over series resistances under high current operation is critical for accurate device modeling. Given the high current through emitter and collector node of the transistor in forward active mode even small deviations in the determination of resistance contributions in the equivalent circuit may cause large errors in circuit simulation. Even though optimized processes keep resistivity values on a low level, the influence on circuit performance through negative feedback may be significant and accurate characterization is crucial.

Theoretically any resistance contribution in the HBT device is considered as a low-field sheet resistance. It may therefore principally be calculated from the device profile using the distribution of process parameters such as doping concentration (N) and carrier mobility (μ) as well as the elementary charge (q) by integration in the direction of the current flow (from x_1 to x_2)

$$r_s = \frac{1}{\int_{x_1}^{x_2} N \cdot \mu \cdot q \, \mathrm{d}x}. \tag{4.3}$$

The inner transistor model (cf. Fig. 2.10) takes into account all intrinsic transistor action including resistances. In general these separate the external nodes from the internal node (*denoted by* ‘) and represent the fact that external terminals are linked to the intrinsic device through series resistances. The parasitic resistances in modern HBTs are thus defined:

- R_{Bx}, the external base resistance between metal interconnect and BE junction
- R_{Cx}, the parasitic collector series resistance
- R_E, the emitter series resistance due to poly-silicon and via contact

In addition to these series resistances (base resistance R_{Bx}, emitter resistance R_E and external

collector resistance R_{Cx}), this section is dedicated to the extraction of the bias-dependent internal base resistance R_{Bi}.

Another aspect that needs to be considered for accurate transistor modeling is related to the generation of noise. Since external resistances are physical ohmic resistances they contribute thermal noise and thus are essential for calculation of the total noise generated in the transistor [97].

4.4.1 Internal Base Node (*B'*)

For proper modeling of distributed effects across the lateral device one has to recall the equivalent circuit of the HICUM model (cf. Fig. 2.10). The (planar) structure under the emitter window (indicated by the dashed box in the EC) is considered the inner one-dimensional transistor. However to take into account three-dimensional effects the device structure is represented by a more complicated network.

As seen in the schematic device cross section specific to the HICUM model (cf. Fig 4.14) the internal transistor (at the B' node) is separated from the base link region (terminal B*) in order to take into account the distributed resistive behavior of the base layer by means of the (bias dependent) internal base resistance $R_{Bi}(V_{BE})$.

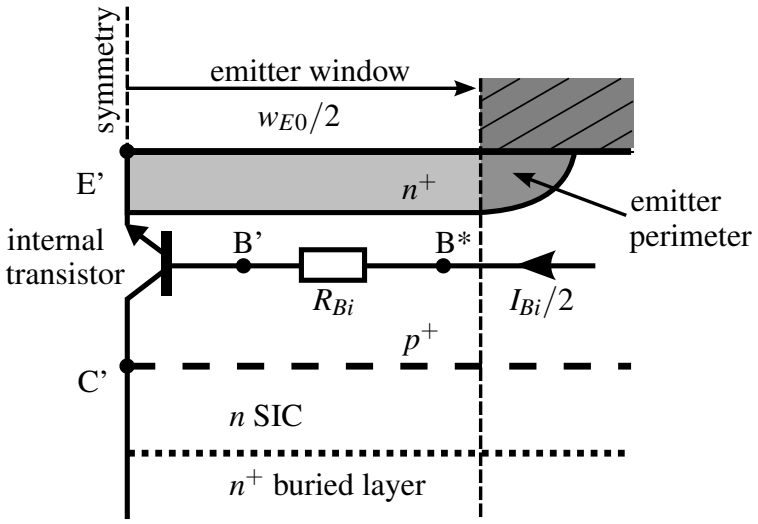


Figure 4.14: Schematic cross-section of the symmetrical BE junction area under the emitter, including internal npn BJT transistor and perimeter region

4.4.2 Base Resistance R_B

In contrast to the emitter and collector series resistances, modeled with constant values the modeling of the base resistance is more complicated due to the combination of the resistance

contributions of the neutral base region and the constant extrinsic base resistance. The internal base sheet resistance R_{SBi} strongly varies with the DC operating point and thus has to be modeled as a function of the transfer current. Several distributed three-dimensional effects in the base have to be taken into account[1]. The base conductivity modulation caused by the transport of emitter-injected charge through the base into the collector and promoted by increased reverse bias across the BC junction, are the dominant factors of influence. 3D effects due to the fact that there is no uniformly constant voltage across the width of the BE junction and emitter crowding are equally important effects for accurate modeling. In the HICUM model R_{Bi} is described as the ratio of zero bias hole charge (Q_{p0}) relative to the hole charge in the base region [31].

The extraction procedure presented here is based on an improvement of the approach presented in [84] using a dual-base ring-emitter device (cf. section 3.5.1). As shown in the equivalent circuit of the measured resistance (cf. Fig. 3.32) a direct measurement between the two independent base contacts B_1 and B_2 using a constant bias ΔV between the terminals is employed. By means of the Klevin technique the current flowing through the terminals is monitored by a first contact, while a second probe gives the accurate voltage reading in the first metal layer (M1).

The resistance equation (4.3) can be simplified by idealization (spatially independent base doping N_B and mobility μ_n over neutral base width w_B) yielding

$$r_{s,B} = \frac{1}{N_B \cdot \mu_n \cdot q \cdot w_B}. \tag{4.4}$$

The inner part of the base resistance is represented by R_{Bi}. For the external base link, a low extrinsic base resistance is achieved by virtue of a silicide layer atop of the heavily-doped polysilicon base contact region. The individual regions are represented by the silicide contribution R_{sil}, the base link region below the spacer denoted R_{lk}. For device modeling purposes, both contributions are combined in one lumped element for the external base resistance, denominated R_{Bx}.

Since the ring-emitter separates the two individual bases the measured resistance represents a parallel configuration of the resistivity chains

$$R_{meas} = R_{Bx}(l_E) + \frac{R_{Bi}(w_E, l_E, V_{BE})}{2}. \tag{4.5}$$

For the measured structures, the extrinsic base link is considered independent of bias and manufactured with constant spatial dimensions for a given emitter window length (l_E). Hence the extrinsic contribution of the base resistance is assumed to be the same for a variation of the device width ($R_{Bx} = f(l_E)$). However the measured intrinsic base resistance is a function of all factors ($R_{Bi} = f(w_E, l_E, V_{BE})$).

[1] resulting in a bias and geometry dependent current crowding function

Some constraints are imposed regarding the measurement range. For the applied base emitter voltage (V_{BE}) two important limits are given: the BE diode shall not be active in forward mode and reverse breakdown has to be avoided. Therefore the V_{BE} range is within the limits of $[-0.5\,..\,0.5V]$.

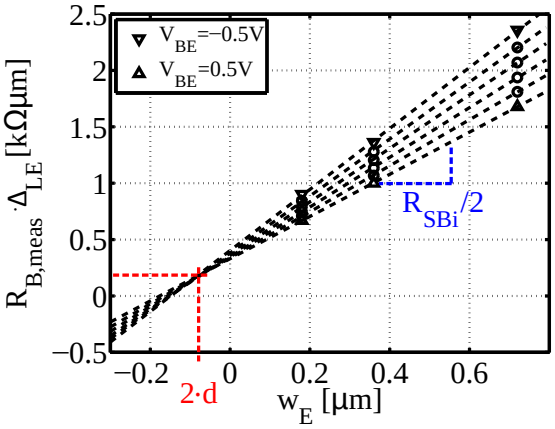


Figure 4.15: Normalized measured base resistance versus emitter width w_E for different base bias values

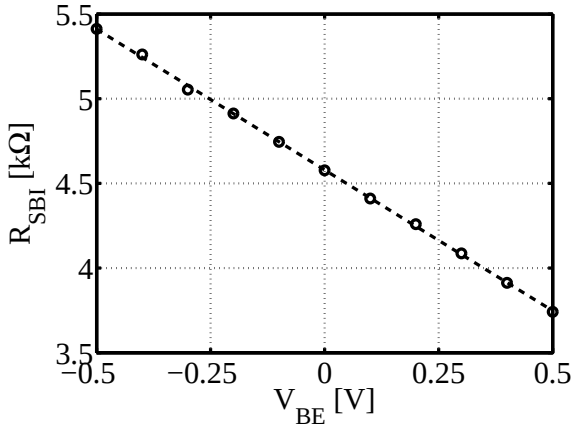


Figure 4.16: Extracted bias dependent intrinsic base resistance R_{sBi} for different V_{BE} junction bias

In order to correct the apparent perimetric effects in the device periphery, two sets of structures using two different lengths ($l_{E,1}$ and $l_{E,2}$) are manufactured. For each set of w_E, the resistance values normalized by the length difference (Δl_E) yield the according intrinsic base resistance progression drawn versus the device width w_E (cf. Fig. 4.15).

From the slope of the $R_{meas}\Delta l_E$ curve versus w_E one then obtains the values of the pinch-base sheet resistance for each individual bias value V_{BE} (cf. Fig. 4.16). The intersect point of all biases represents a geometrical correction factor accounting for the emitter spacer (d). In analogy to the deduction of a two dimensional Δl_E structure, the $R_{meas}\Delta l_E$ product may be normalized by a w_E reference effectively compensating parasitic corner effects through an electrical inner base width ($w_{E,eff} = w_{E,d} - 2 \cdot d$)[1].

$$\Delta_{lw}R_{meas} \cdot \Delta l_E = \frac{R_{SBi}(V_{BE}) \cdot \Delta w_E}{2}. \tag{4.6}$$

The fundamental parameter extraction equation then reads

$$f(R_{SBi}) = R_{SBi} \cdot d + R_{LBx} = \Delta_{lw}R_{meas} \cdot \Delta l_E - \frac{R_{SBi} \cdot w_E}{2}. \tag{4.7}$$

By definition of $f(R_{SBi})$ its curve progression versus the negative R_{SBi} yields the desired values

[1] A negative value for the correction factor d means that after silicon processing the actual emitter window is larger than the initially targeted window width defined in the layout

for d (slope) as well as R_{LBx} (y intercept) and thus presents a more rigorous method than the averaged line intercept in the progression of $R_{meas}\Delta l_E$ versus w_E.

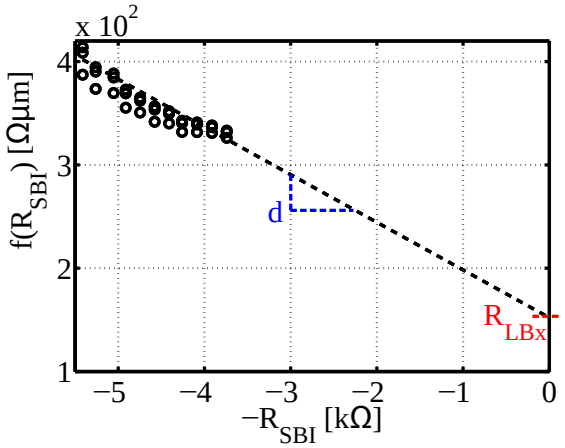


Figure 4.17: Normalized measured base resistance ($f(R_{SBi})$) versus negative bias dependent internal base resistance ($-R_{SBi}$)

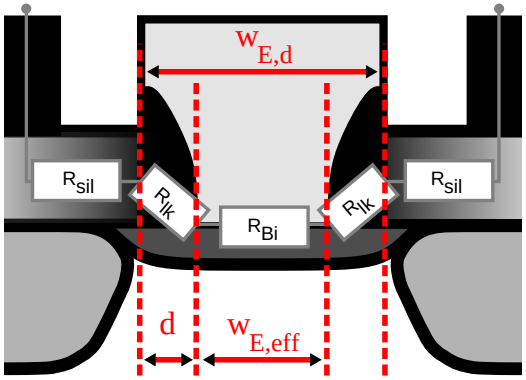


Figure 4.18: Cross section of SiGe SEG HBT with base resistance contributions and electrical base width correction through d

Extrapolation of both parameters, extrinsic base resistance times unit length (R_{LBx}) as well as the geometry correction d is performed by means of a linear regression (cf. Fig. 4.17).

As shown in the schematic cross section (cf. Fig. 4.18) the obtained lineic external resistance may be further partitioned in the contributions of a link region underneath the spacer and the base-silicide interface as well as the silicided region. Therefore additional information regarding the respective spatial dimensions is required (usually taken from the layout definition, cf. Tab. 3.2). Together with the sheet resistance values (R_{Ssil}) specified in the design rules of the employed technology (cf. Tab 3.3) a theoretical value for the link contribution R_{lk} can be calculated.

Circle Impedance Method

In order to verify the results obtained in the DC measurement campaign, a comparison with a normal device structure is desired. A measurement on RF structures may provide this information by means of the circle impedance method [98] with the modified hybrid parameter in order to cancel out the unwanted influence of the emitter resistance

$$h'_{11} = \frac{1}{y_{11}+y_{12}} = \frac{g_x + g_\pi + j\omega C_\pi}{g_x \cdot (g_\pi + j\omega C_\pi)}. \tag{4.8}$$

One can see that in the equivalent circuit the parallel capacitances vanish for infinite frequency ($f \to \infty$, cf. Fig. 4.19). Thus a regression with a fit of the circle equation $r^2 = y^2 + (x - x_0)^2$ gives

the total base resistance $R_B = R_{Bi} + R_{Bx}$ from the separation

$$x_0 = R_B + \frac{r_\pi}{2} = \frac{a}{2} \tag{4.9}$$

$$r = \frac{r_\pi}{2} = \sqrt{b + x_0^2}\,, \text{ with } b = r^2 - x_0^2 \text{ and } a = 2x_0 \text{ gives} \tag{4.10}$$

$$R_B = \frac{a}{2} - \sqrt{b + \frac{a^2}{4}} \tag{4.11}$$

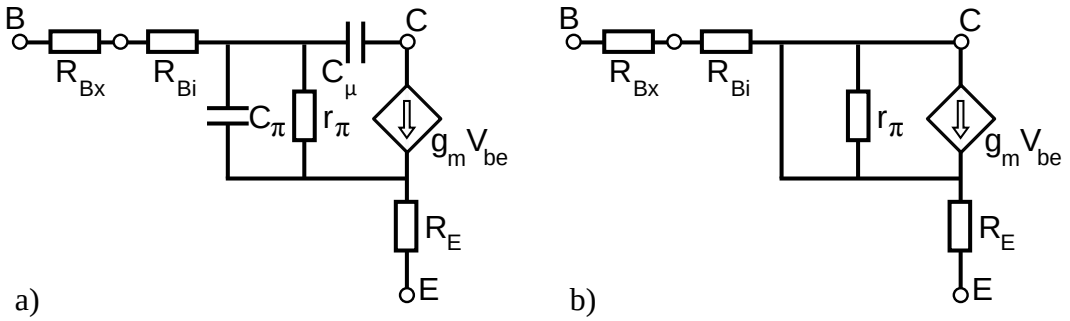


Figure 4.19: Equivalent circuit for R_B extraction from impedance circle method a) at low frequency, b) at high frequency ($f \to \infty$)

Due to the noisiness of the extracted values, the extraction of R_B is shown for different high bias values ($V_{BE} \geq 0.9V$) using S-parameter measurement data in forward active operation with zero BC bias (V_{BC0}).

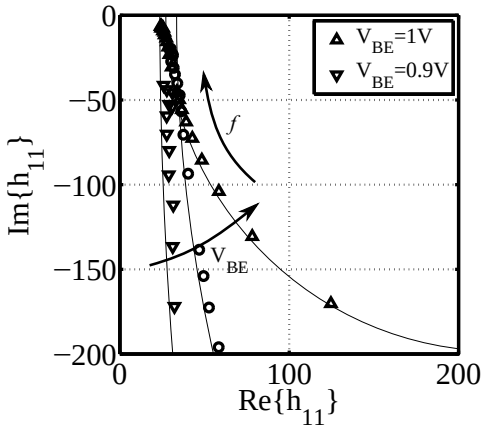


Figure 4.20: Base resistance extracted from forward active S-parameter data by extrapolation of h'_{11} on a circle of constant impedance; *symmetrical device with* $w_{E,min}$, $l_E = 5\mu m$

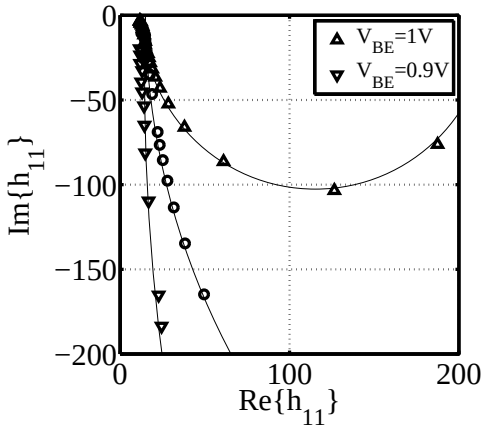


Figure 4.21: Circle impedance base resistance extraction; *symm. device with* $w_{E,min}$, $l_E = 10\mu m$

4.4.3 Collector Resistance R_C

Because of the length of the collector buried layer (BL) region and its comparably low impurity concentration, the external collector (usually) has a relatively large resistance R_{Cx} and thus a strong influence on device characteristics. An accurate measure of the individual external resistance contributions is hence desired.

As for the base resistance there are different methods existing for the collector related resistivity. Besides the extraction by means of equivalent transformation based on RF measurement data, direct extraction methods making use of dedicated test structures are known to give more accurate results. A major benefit using these structures is the separation of individual sheet resistances (R_{SBL} and R_{SSK}) relevant to their position in the device cross section.

The method giving the most accurate results is making use of direct DC measurements on dual collector structures as presented in section 3.5.1. The according cross section of the measured resistance contributions is shown in Fig. 4.22. Based on an existing approach presented in [78] the used structure improves the robustness and reliability of the method. The former approach uses a standard symmetrical bipolar transistor with individual access to the collector terminals (C1 and C2) in fully symmetrical CBEBC configuration whereas the buried layer width to both sides of the emitter window is kept constant. Through a geometry scaling based approach with variation of the emitter width (w_E) the total width of the buried layer was varied, yet the outer dimensions of the structure remained constant ($w_{BL,1} = w_{BL,2} = const.$).

The total buried layer width is given through

$$w_{BL} = w_{BL,1} + w_E + w_{BL,2}. \tag{4.12}$$

As shown in [85], the sophisticated approach makes use of one single emitter width structure with minimal dimensions ($w_E = w_{E,min} = const.$) and a variation of the buried layer itself, providing a wider spread of the total BL width and thus improving accuracy. As seen in the equivalent circuit for DC measurement with forced collector current ($I_{C,1} = -I_{C,2} = const.$) one obtains the total resistance from the measured potential drop ($R_{Cx} = 2 \cdot \Delta V_C / (|I_{C,1}| + |I_{C,2}|)$). Thereafter a multiplication of the measured resistance by the structures length (l_{BL}) yields the normalized product

$$R_{Cx} \cdot l_{BL} = \frac{2 \cdot R_{SK}}{w_{SK}} + R_{SBL} \cdot w_{BL}. \tag{4.13}$$

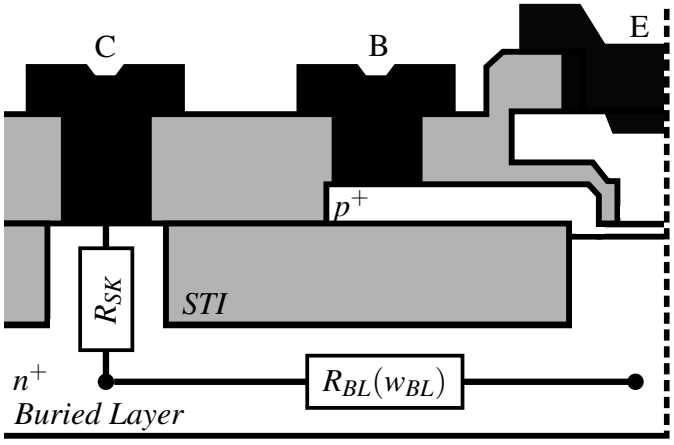


Figure 4.22: Cross section of the BC region in a vertical SiGe HBT device with related collector resistance contributions of buried layer (R_{BL}) and sinker (R_{SK})

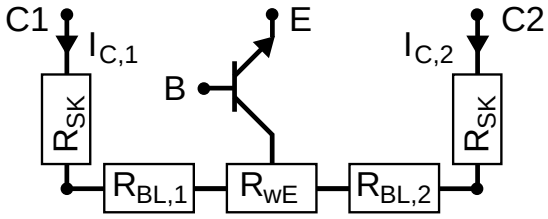


Figure 4.23: Equivalent circuit of collector resistance R_C with measured contributions

To take into account the apparent current spreading in the buried layer during the forced current measurement, an electrical correction factor ($\Delta l_{BL} = dw_{BL}$) can be determined, that minimizes the error of the normalized resistance value. The resulting current is then assumed to flow in a purely parallel current path. Via optimization of the relative error between the four normalized vectors (e.g. least squares approach) an optimum value for the total electrical emitter length with $l_{BL,eff} = l_E + dw_{BL}$ is determined, that is applicable to all geometries by addition to the drawn emitter length (cf. Fig. 4.25). Through a linear regression one can then extract the two technology-specific values of interest (R_{LSK} and R_{SBL}) that allow to re-calculate the respective collector resistance value for any given device configuration or geometry.

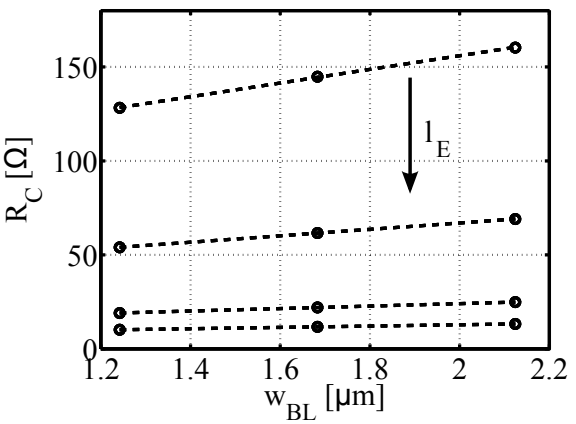


Figure 4.24: Measured collector resistance R_C for different emitter window lengths (l_E) as function of different buried layer widths (w_{BL})

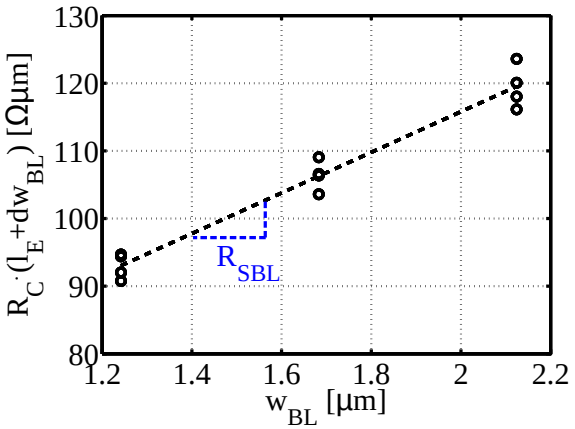


Figure 4.25: Collector resistance (R_C) normalized by the emitter window length (l_E) for different buried layer widths (w_{BL})

Another advantage of the used structure is that the drop in the potential between the C1 and C2 probe gives a possibility to verify the extracted values under forward active operation. Therefore the current in the second contact is forced to zero ($I_{C,2} = 0$), providing the total resistance of buried layer and sinker on one side of the dual collector test structure

$$R_{Cx,fwd} = R_{SK} + R_{BL}(w_{BL}) = \left.\frac{V_{C2} - V_{C1}}{I_{C1}}\right|_{I_{C,2}=0}. \tag{4.14}$$

Like for the base resistance, RF data based methods for the assessment of R_C exist yet the accuracy reached by these approaches is very limited.

4.4.4 Emitter Resistance R_E

The aim of technology development is a low-ohmic electrical contact to the emitter layer (ohmic resistor R_E) achieved by a high impurity concentration in the emitter poly-silicon. Even though R_E presents a relatively small series resistance, the conductivity of the poly-silicon is limited and due to the high emitter current, the inherent potential drop over the series resistance R_E is a critical device parameter for static as well as dynamic performance of the BJT and needs to be determined precisely for proper simulation of the device behavior.

The vertical series resistance contributions and their position in the device structure are shown in the cross section (cf. Fig. 4.26). One can identify an internal resistance contribution $R_{E,int}$ corresponding to the resistance of the emitter poly layer enclosed by the inner spacers. In addition the interface between poly-Si and metal adds to the resistivity as does the first via or contact stripe that provides contact to the first metal layer (M1) summarized in $R_{E,via}$ or $R_{E,con}$.

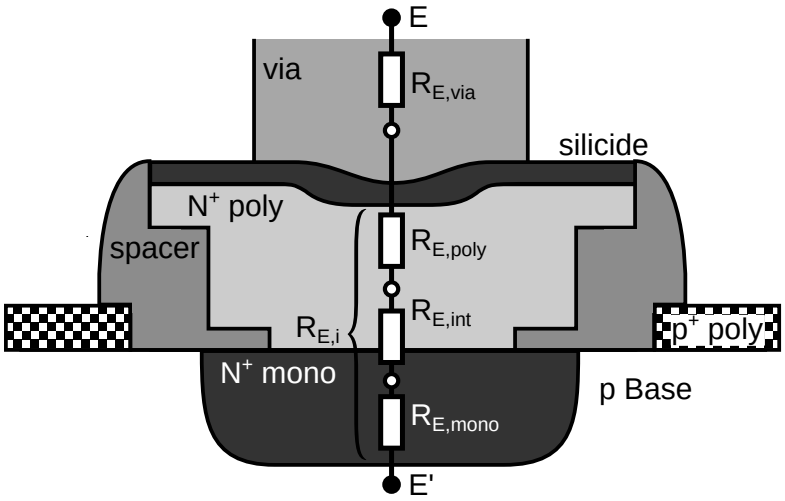


Figure 4.26: Schematic cross section of the emitter structure with contributions from poly-Si and via contact in the upper vertical HBT device

Unlike the base resistance the emitter resistance is modeled as a bias independent ohmic resistor. To date there are many proposals for its extraction. Various experimental techniques for evaluating the emitter series resistance have been presented, either relying on measurement of simple DC *I-V* characteristics or based on RF measurements, yet a specific test structure is not available. Hence a conventional device has to be used for parameter determination.

For verification or a first guess initialization of the emitter series resistance a (simple) method to directly extract the resistance parameters based on actual measured data, without requirement for any special test structures is highly desirable.

DC approach

Being a small resistance contribution R_E generally is difficult to determine from DC measurements under low injection whereas under high injection other effects such as self heating must be taken into account. The popular g_{mx} method is based on measurement of the transfer conductance from the Gummel plot with the collector current I_C as function of the base-emitter voltage V_{BE} at constant base-collector bias ($V_{BC} = \text{const.}$) [99]. Several assumptions are implied and improvements were presented. However the method provides a valuable asset of R_E for a first guess and therefore shall be presented in its simplest form.

From the ideal transistor diode equation follows that I_C varies exponentially with V_{BE}. The small-signal transconductance g_m is then defined as the slope of the transfer characteristics curve (g_{mx}) evaluated at a given DC operating point

$$g_m = \left.\frac{\partial I_C}{\partial V_{BE}}\right|_{V_{CE}} = \frac{I_S}{V_T}\exp\left(\frac{V_{BE}}{V_T}\right) \approx \frac{I_C}{V_T}. \tag{4.15}$$

With the external voltage V_{BE} this is considered as extrinsic conductance. The collector current equation from the model at V_{BC0} is used. Expressing the inner BE voltage $V_{B'E'}$ by the resistances and application of the natural logarithm to both sides of the model equation

$$I_C = \frac{c_{10}\exp\left(\frac{V_{B'E'}}{V_T}\right)}{Q_{p,T}}, \text{ in turn yields} \tag{4.16}$$

$$\ln(I_C) = \ln(c_{10}) - \ln(Q_{p,T}) + \frac{1}{V_T}\left(V_{BE} - I_C \cdot \left(R_E + \frac{R_B}{\beta_0}\right)\right). \tag{4.17}$$

The approximative g_m-method uses the total extrinsic transconductance g_{mx} for a sweep of the BE voltage using idealization (by means of a known internal device temperature or thermal voltage V_T respectively). Assuming constant c_{10} and negligible minority charges ($Q_{jEi}, Q_{jCi} \approx 0$)

the extrinsic conductance g_{mx} obtained from the derivative wrt. the input voltage V_{BE} gives

$$\frac{\partial \ln(I_C)}{\partial V_{BE}} = \frac{1}{V_T} - \frac{R_E}{V_T} \cdot \underbrace{\frac{\partial I_C}{\partial V_{BE}}}_{g_m} = \frac{\partial I_C}{\partial V_{BE}} \cdot \frac{1}{I_C} = \frac{g_m}{I_C} \tag{4.18}$$

$$\frac{g_m}{I_C} = \frac{1}{V_T} - \frac{R_E}{V_T} \cdot g_m \rightarrow \frac{V_T}{I_C} + R_E = \frac{1}{g_m}. \tag{4.19}$$

With the conditions, applicable to HBT devices with high gain ($I_T = I_C, \beta \rightarrow \infty$, and $m = 1$) one can hence extrapolate R_E as shown in Fig. 4.28. Taking into account non-idealities of limited gain and a non-negligible base resistivity one obtains a form where the additional knowledge of parameters can further increase the accuracy

$$\frac{1}{g_m} = \frac{1}{g_{mi}} + r_E \left(1 + \frac{1}{\beta_0}\right) + \frac{r_B}{\beta_0} \tag{4.20}$$

$$g_{mi} = \left.\frac{\partial I_T}{\partial V_{B'E'}}\right|_{V_{CE}} = \frac{I_T}{m_C(I_T) V_T} = \frac{g_m}{1 - R \cdot g_m} \tag{4.21}$$

$$\frac{1}{g_m} = \frac{m_C(I_T) V_T}{I_T} + r_E \left(1 + \frac{1}{\beta_0}\right) + \frac{r_B}{\beta_0} \tag{4.22}$$

However all extrapolation-approaches have some severe limitations. Extracting the series resistance of HBTs using the high-current region, the apparent self-heating of the device adds ambiguity and error sources to the measured data (V_T shift). Neglecting additional effects such as the Early effect, as well as the Kirk effect make the approximations for emitter resistance extraction even more inaccurate. For all the DC methods in presence of self-heating the value of the observed g_m versus current changes since self-heating effects produce a significant shift in the *I-V* curve of HBTs. Finally under very high forward bias, thermal runaway (with R_E dropping towards zero) is a well known problem.

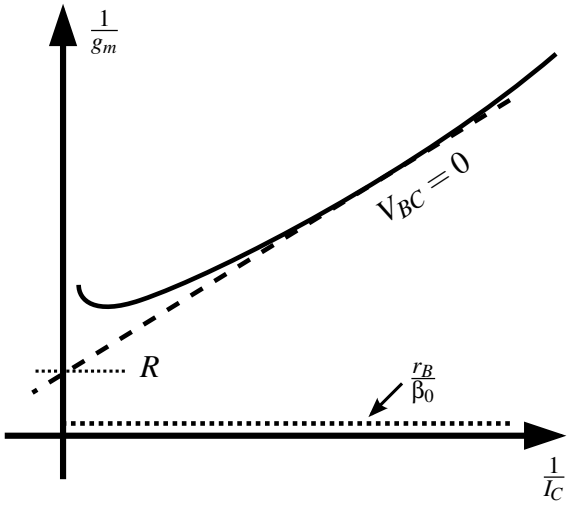


Figure 4.27: Inverse transconductance extraction for R_E using DC measurement data

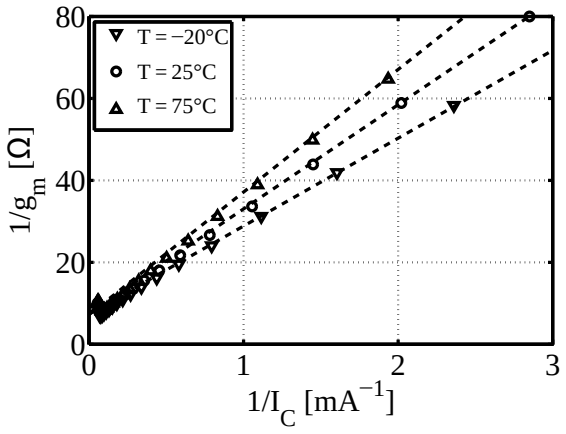


Figure 4.28: Extrapolation approach for emitter series resistance R_E from DC measurement using the g_m method, *single device with 0.2x5µm²*

RF approach

For the estimation from DC measurements, the device is either biased under conditions significantly different from normal operation or influenced by apparent second-order effects. Though DC based methods may yield inaccurate results. In consequence several AC based approaches have been proposed [100–103]. Yet the AC behavior depends highly on the chosen small-signal representation used.

The popular method by Gobert [101] provides information about a number of device parameters, using simple impedance parameter measurements ([**Z**] matrix obtained from S-parameter measurements). For the asset of R_E it is using an extrapolation towards infinite base current (where R_B is assumed to be shunt). However a difficulty of the method on first sight is that the inner collector resistance as well as the base resistance are to be known or assumed negligible otherwise.

$$\Re(Z_{12}) = R_E + \frac{R_{ci}R_{bi}}{R_{ci}+R_{bi}} \tag{4.23}$$

$$\Re(Z_{11} - Z_{12}) = R_{bx} \tag{4.24}$$

$$\Re(Z_{22} - Z_{12}) = R_{cx} \tag{4.25}$$

RF methods commonly make use of devices manufactured in the GSG connection scheme (cf. appendix F) providing access to base and collector terminal whilst the emitter node of the HBT device is tied to the reference (GND). The active transistor is measured under forward Gummel bias conditions, driving the base-emitter diode bias. Eliminating the influence of the BC junction at constant V_{BC} = 0V, effects caused by the base-collector diode as well as avalanche multiplication

or the substrate network are hence considered negligible. The current flowing through the emitter resistance is defined using I_B, at a measured intrinsic base current for moderate to large biases where ideality ($m = 1$) is assumed.

$$i_{re} = i_b + i_c = i_b \cdot (1 + h_{21}) \tag{4.26}$$

$$y_{21} = g_m, \;\; h_{11} = \frac{1}{y_{11}}, \;\; h_{21} = \frac{I_C}{I_B} = \frac{y_{21}}{y_{11}} \tag{4.27}$$

Taking extrinsic capacitances out by assuming they are sufficiently low at low measurement frequencies and summarizing the total base resistance into R_B while taking into account all remaining components in the forward mode, one gets from the equivalent circuit (cf. Fig. 4.29).

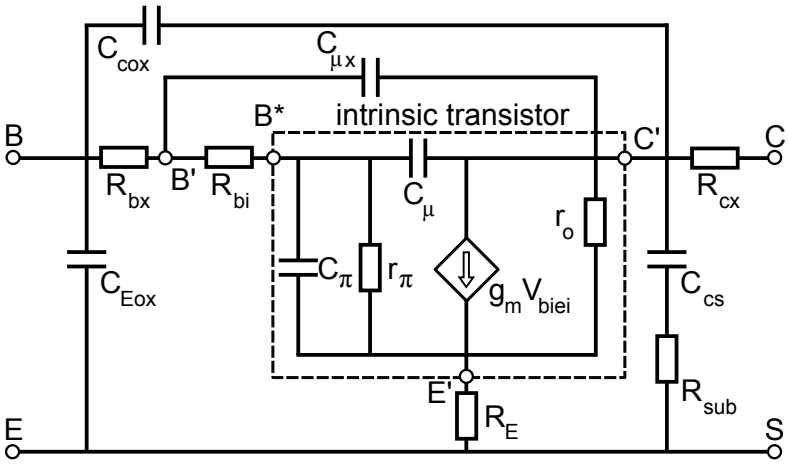


Figure 4.29: Common emitter T equivalent circuit used for emitter series resistance extraction from RF measurement

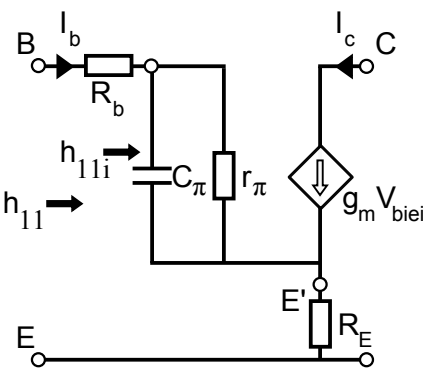


Figure 4.30: Equivalent circuit of HBT for unilateralized network

$$i_b \left(1 + \frac{R_E}{R_B}\right) + i_c \left(\frac{R_E}{R_B} - j\omega C_{bc} R_C\right) = v_{be} \left(\frac{1}{R_B} + j\omega C_{bc}\right) + v_{b^*e'} \left(-\frac{1}{R_B}\right). \tag{4.28}$$

By monitoring the real part of the transfer impedance (open-circuit impedance parameters, $\Re\{Z_{12}\}$) against the inverse collector current ($1/I_C$), the extrapolated intercept at the ordinate gives the value of R_E through linear regression. Yet this representation still uses lots of interdependences between the elements and is not suited for highly accurate estimation of R_E without knowledge of other model parameters.

Even though the method is known to work reasonably well for InP based technologies, the parasitic substrate network connected to the emitter node is of much higher importance for SiGe technologies, causing the assumptions made to be insufficiently accurate. It is hence desirable to isolate the contribution of R_E from all other resistances by means of a more sophisticated method.

The unilateralization-based RF approach

The simplified hybrid-π equivalent circuit of a bipolar transistor in forward bias condition (cf. Fig. 4.29 and appendix section G.3) is the basis for the derivation of the used R_E formulations. All external elements (e.g. capacitors C_{Eox}, C_{Cox} and resistors R_{Cx}, R_{Bx}, R_E, and R_{sub}) are considered as linear lumped elements independent of device bias. The elements inside the dashed line (cf. Fig. 4.29) represent the intrinsic transistor. Using a representation of measured data in the [**h**] matrix one obtains

$$h_{11} = r_b + h_{11,i} + r_e \cdot (1 + h_{21,i}) = r_b + \frac{h_{21,i}}{g_m} + r_e \cdot (1 + h_{21}) = r_b + r_e + h_{21} \cdot \left[r_e + \frac{1}{g_m} \right]. \quad (4.29)$$

The idea behind the unilateralization approach is to remove transfer branches, thus simplifying the admittance matrix and removing interrelations. Unilateralization is a desirable property of active networks, since it implies that input signals are processed only from input port to output port, with no response signal fed back to the input port from the output (input and output are considered completely independent of passive networks and isolated from each other). One consequence is that the lack of internal feedback renders unilateral networks unconditionally stable, yet it is an idealized operating condition.

In general, active networks show vanishingly small y_{12} at low signal frequencies, whereas z_{12} increases towards progressively higher signal frequencies. Considering the special unilateral case of zero internal feedback also means that $h_{12} = 0$. Thus neglecting the internal feedback of the transistor (through reverse isolation with S_{12} = 0) makes it possible to independently solve input and output circuits of the HBT. The elimination of internal feedback of the transistor has two advantageous benefits: one can remove the transfer branches from the equivalent circuit yet the unilateralized current gain can still be considered close to the intrinsic one ($h_{11} \approx h_{11i}$).

Generally one can hence derive the unilateralized admittance parameters ([$\widetilde{\mathbf{y}}$] matrix) from the measured admittance matrix as well as the associated gain (h) as follows:

$$\widetilde{y}_{ii} = y_{ii} + y_{21} \quad \text{and} \quad \widetilde{y}_{21} = y_{21} - y_{12} \quad (4.30)$$

$$\widetilde{h}_{11} = \frac{1}{\widetilde{y}_{11}} \quad \text{and} \quad \widetilde{h}_{21} = \frac{\widetilde{y}_{21}}{\widetilde{y}_{11}} \quad (4.31)$$

The according representations of HBT device elements in the equivalent circuit (cf. Fig 4.30)

are defined as follows:

$$y_{11} = g_\pi + j\omega(C_\pi + C_{jC}) \qquad \Rightarrow \widetilde{y}_{11} = g_\pi + j\omega C_\pi \tag{4.32}$$

$$y_{12} = -j\omega C_{jC} \qquad \Rightarrow \widetilde{y}_{12} = -j\omega C_{jC} \tag{4.33}$$

$$y_{21} = g_m - j\omega C_{jC} \qquad \Rightarrow \widetilde{y}_{21} = g_m \tag{4.34}$$

$$y_{22} = g_0 + j\omega C_{jC} \qquad \Rightarrow \widetilde{y}_{22} = g_0 \tag{4.35}$$

With the standard RF equivalent circuit in its simplified form by means of unilateralized parameters, the emitter resistance R_E can be directly obtained from measured data at low frequency and high current as presented in [104]. A compromise between the low measurement frequency requirement of the presented method and the resolution limitations of the network analyzer is made. From experience a spot frequency of $f_{meas} = 2GHz$ is best suited for parameter extraction.

To further simplify the extraction, the third term of the LHS divided by the (usually very high) gain of RF transistors, allows to neglect the respective contribution

$$r_e + \frac{1}{g_m} - \underbrace{\frac{\omega C_{bc}}{g_m \widetilde{h}_{21}}(r_{bi} + r_{ci})}_{\approx 0} = \frac{\Im(\widetilde{h}_{11} \cdot conj(1 + \widetilde{h}_{21}))}{-\Im(\widetilde{h}_{21})} \text{ [105].} \tag{4.36}$$

In turn, a relation suitable for parameter extraction based on (4.29) through separation of imaginary and real parts of LHS and RHS is obtained

$$\frac{1}{g_m} \approx \frac{1}{\widetilde{h}_{21}} \cdot \frac{1}{\widetilde{y}_{11}} \approx \Re\left(\frac{1}{\widetilde{h}_{21}}\right) \cdot \Re\left(\frac{1}{\widetilde{y}_{11}}\right) = \Re\left(\frac{1}{\widetilde{h}_{21}}\right) \cdot \frac{V_T}{I_b} \tag{4.37}$$

$$r_e + \frac{1}{g_m} = \frac{\Im(\widetilde{h}_{11})}{\Im(\widetilde{h}_{21})} \tag{4.38}$$

$$\Re\left(\frac{1}{\widetilde{h}_{21}}\right) = \Re\left(\frac{1}{\widetilde{h}_{21i}}\right) = \frac{g_\pi}{g_m} = \frac{1}{g_m}\frac{I_b}{V_T} \tag{4.39}$$

$$r_e = \frac{\Im(\widetilde{h}_{11})}{\Im(\widetilde{h}_{21})} - \Re\left(\frac{1}{\widetilde{h}_{21}}\right)\frac{V_T}{I_b} \tag{4.40}$$

By plotting the respective contributions of $\Im(\widetilde{h}_{11})/\Im(\widetilde{h}_{21})$ and $\Re(1/\widetilde{h}_{21}) \cdot V_T/I_B$, a curve suitable for linear regression is obtained, as shown for a single device geometry in Fig. 4.31.

Alternatively R_E can be directly calculated from eqn. (4.40). Yet there are several measured bias points to choose from (cf. Fig 4.32) and neither the practice of taking the absolute minimum nor the exclusive determination of the inflection point of the total R_E as upper limit of the extraction bias is found suitable for automated parameter extraction. A more robust, automated method for the asset of R_E is hence desired for practical application.

A two-step approach is found to provide the best solution. The joint extraction approach for automated parameter extraction hence uses both information: The inflection point (minimum of $\Re(1/\widetilde{h}_{21}) \cdot V_T/I_B$ vs. bias) gives the maximum bias allowed for average calculation. Thereafter automated outlier-detection is performed. The Thompson-Tau algorithm (as proposed by [106, 107]) is applied to the resulting dataset of extracted resistances R_E as a function of BE bias in order to rule out deviating erroneous points (especially at low bias).

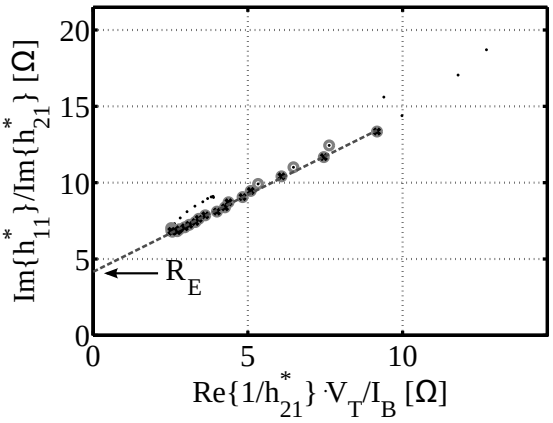


Figure 4.31: Regression analysis of emitter resistance R_E using unilateralized h-parameters; *single device with 0.2x5µm²*

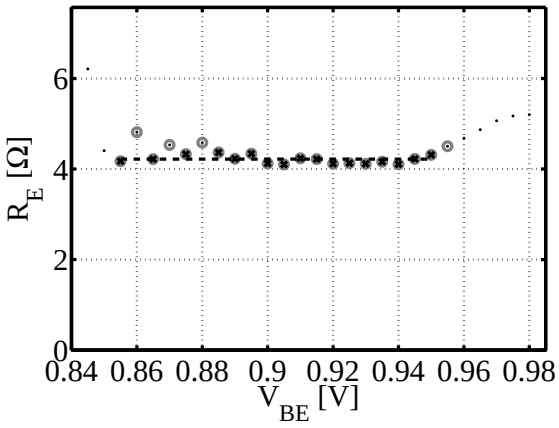


Figure 4.32: Direct extraction of R_E with automated outlier detection as a function of V_{BE} bias

Given the accurately determined individual resistances R_E for a multitude of devices the model parameter for geometry scalable device modeling can be determined as shown in Fig 4.33. Separation of the total resistance of each device into the associated via contact resistance $R_{K,via}$ and the poly-emitter related contribution R_{ke} is then performed by means of the appropriate scaling equations and the specific spatial dimensions of each device (cf. section 3.3.4).

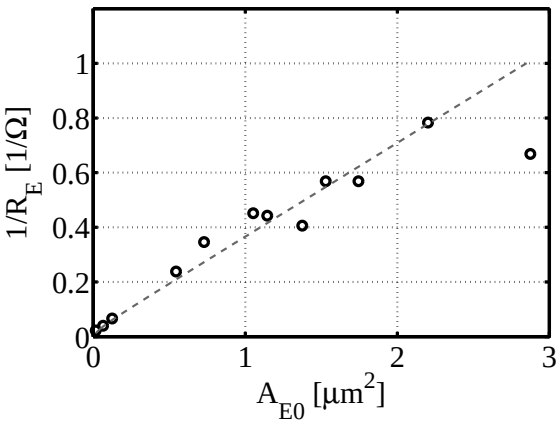


Figure 4.33: Geometry scaling of inverse extracted emitter resistance versus actual emitter area for various geometries

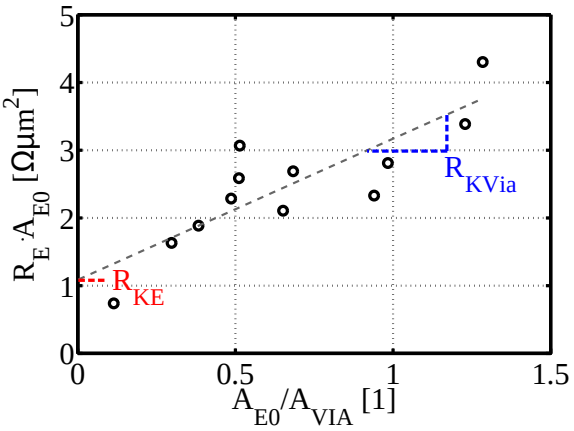


Figure 4.34: Extraction of specific emitter resistance contributions of via contact and polyemitter from regression analysis

4.5 Capacitances

Each of the pn-junctions in a bipolar transistor are charge storing elements. Due to the variation of the depletion width in a semiconductor SCR, the nonlinear charge-storage ($C = dq/dv$) is defined as a small-signal quantity (ratio of the small-signal charge) whereas the total charge in each of the depletion regions consists of the static DC charge plus a small-signal charge contribution ($q_j = Q_{j,DC} + q_{j,AC}$).

Thus for accurate assessment of the device operation under high frequencies, the modeling of charge storage in the transistor requires accurate capacitance characteristics and the extraction of compact model parameters for junction capacitances from measured data is an important and fundamental part in the extraction flow.

For bipolar transistors the most important capacitances are:

- base-emitter capacitance, C_{BE} or C_π
- base-collector junction depletion capacitance, C_{BC} or C_μ
- substrate depletion capacitance, C_{CS}

Given the strong link of AC and DC characteristics in charge based device models the extraction of space charge capacitance parameters is one of the first steps. The required measured data is obtained in the so called ”cold” state of the transistor. Under reverse and low forward bias conditions for all respective junctions. In this off-state of the transistor, the following approximations can be made to simplify the equivalent circuit:

- Negligible transfer current (i_T) which implies infinite values for r_π and r_μ shunting the diffusion capacitances C_{de} and C_{dc} so that they can be neglected

- Series resistance R_B, R_E and R_C can be neglected due to the lack of current gain

Thus the equations to obtain parameter from the measured (deembedded) admittance matrix ([**y**] matrix) are fairly simple:

$$C_{BE} = \frac{\Im\{Y_{11}\} + \Im\{Y_{12}\}}{2\pi f} \tag{4.41}$$

$$C_{BC} = -\frac{\Im\{Y_{12}\}}{2\pi f} \tag{4.42}$$

$$C_{CS} = -\frac{1}{2\pi f \cdot \Im\left\{\frac{1}{Y_{12}+Y_{22}}\right\}}. \tag{4.43}$$

The general equation used for description of the junction capacitances is defined through the derivation of the depletion capacitance relation for a pn-junction with an abrupt doping profile with minor modification[1] [97].

In general the depletion capacitance C_j is thus defined using three fundamental model parameters: The zero-bias junction capacitance is represented through the parameter C_{j0}. The remaining two model parameters are used for the description of bias dependence: the built in potential (Φ_B) is represented by the model parameter (V_D) and the exponent factor (z_j) sets the curvature (fitting coefficient z depends on the junction doping profile). The full model equation thus reads

$$C_j = \frac{C_{j0}}{\left(1 - \frac{V}{V_D}\right)^{z_j}}. \tag{4.44}$$

This equation however is not suited to be directly solved by simple mathematical operations. Therefore after extraction of C_{j0} from the zero-bias intersect a linear optimization approach using the correlation coefficient r is employed for parameter determination using the equation

$$\ln(C_j) = \ln C_{j0} - z_j \cdot \ln\left(1 - \frac{V}{V_D}\right) \text{ , with the normalized form} \tag{4.45}$$

$$\ln\left(\frac{C_j}{C_{j0}}\right) = -z_j \cdot \ln\left(1 - \frac{V}{V_D}\right). \tag{4.46}$$

There are some steps in common for the extraction of junction capacitances. The selection of measured data is done using the off-state bias condition of network NW14 (for C_{BE}) and NW15 (for C_{BC} and C_{CS}) respectively. Measurement data is used from reverse bias up to a medium forward bias, where the diffusion capacitance becomes non-negligible.

In a first step the applicable frequency range is selected, since towards high frequency the

[1] through an additional degree of freedom by means of a user defined exponent factor z

equivalent circuit representation becomes invalid and strong variation of the calculated capacitance is observed. Figure 4.35 exemplary shows the dependence of the measured capacitance of the BE junction at three different bias points for a single device over the whole range of measurement frequency (f_{meas}). As seen in the figure at low frequencies the measured capacitance is found to be relatively stable. Hence as a best practice the measured capacitance data is averaged for each bias point within a range of low measured frequencies. Typically one selects $f_{meas} = [5 \,..\, 20]GHz$.

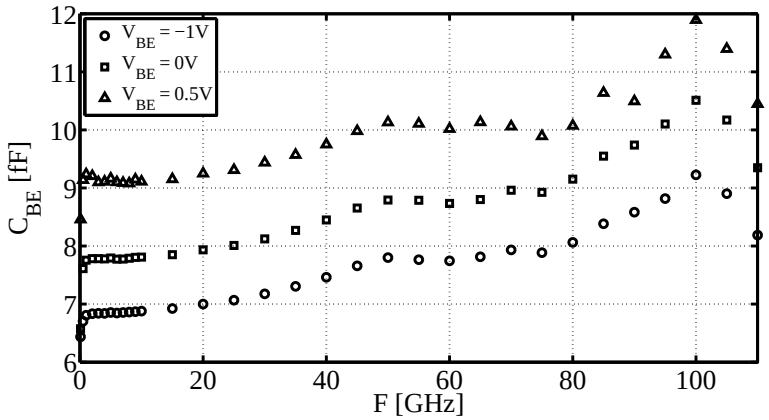


Figure 4.35: C_{BE} capacitance as function of frequency (F); *device with* $w_E = 0.2\mu m, l_E = 5\mu m$

In addition to cold S-parameter measurements, special capacitance test structures with either a larger size (emitter window area of 100x100μm) or parallel matrix connection of multiple devices (n x 0.2x5.0μm^2) with DC pad connection and corresponding deembedding structures can be measured with a CV or LCR meter at lower frequencies ($f_{meas} \leq 1MHz$).

Since most of the extraction procedures for distributed capacitance splitting across the base resistance are applicable to different bipolar compact models (e.g. HICUM, MEXTRAM, VBIC), the fundamental extraction procedure has been detailed in different sources (c.f. [108]).

Yet for accurate modeling one has to consider that junction capacitances and parasitic oxide and intra-metal capacitances are connected in parallel. Thus the total value determined in measurement ($C_{j,meas}$) equals the sum of the bias-independent external part ($C_{j,x}$) plus the bias-dependent internal capacitance (C_j) of the respective SCR (cf. Fig. 3.13). The denominator j here indicates the respective junction.

The contribution of the external capacitance originating from the extrinsic oxide or overlap capacitances ($C_{j,par}$) and the coupling between the terminal contacts ($C_{j,con}$) is thus removed from the measured total value in a very first correction step using the geometry information of the respective transistor area (cf. sect. 3.3.8). The according values of parasitic external contributions ($C_{ox,l}$ and C_{con}) are usually determined by a numerical field solver and stored in the matrix of

device-specific information.

$$C_j = C_{j,meas} - C_{j,con} - C_{j,par} \tag{4.47}$$

$$C_{j,con} = n_{con} \cdot C_{con} \tag{4.48}$$

$$C_{j,par} = P_j \cdot C_{ox,l} \tag{4.49}$$

Using corrected values, the junction capacitance extraction is then performed for each of the individual junctions present in the transistor using the general scaling equation with partitioning approach (partitioning factor X_j) suited for extraction of model parameters (cf. sect. 3.3.1)

$$C_j = C_{jA} + C_{jP} \text{ with} \tag{4.50}$$

$$C_{jA} = X_J \cdot C_j = A \cdot C_{jA} \text{ and} \tag{4.51}$$

$$C_{jP} = (1 - X_J) \cdot C_j = P \cdot C_{jP}. \tag{4.52}$$

4.5.1 Parasitic BC Overlap Capacitance

In order to increase the transit frequency (f_T) as well as the maximum frequency of oscillation (f_{max}), in todays self-aligned SiGe HBT technologies a significant effort is spent to decrease both the base-emitter (BE) as well as the base-collector (BC) capacitance.

The total BC capacitance consists of several contributing elements (cf. Fig. 4.36), with the bias-dependent internal junction capacitance of the BC diode (C_{jC}) being the most significant. However in addition to the junction (bottom area) related depletion capacitance the external capacitance is to be taken into account as well. This capacitance consists of two predominant elements: the bias-dependent peripheral depletion capacitance (C_{jCx}) related to the junction and the parasitic inter-electrode BC capacitance ($C_{BC,par}$), which is considered independent of device bias. Amongst others (metalization etc.) this capacitance takes into account one particular part which is the poly-silicon overlap or oxide capacitance ($C_{BC,ox}$) between the base link and the collector connection (BL).

Even though several different geometries of conventional symmetrical device dedicated to model parameter extraction are manufactured on silicon, $C_{BC,ox}$ cannot simply be separated from other apparent scaling effects. This is due to the fact that the external connection scheme, with the relevant spatial dimensions of the device terminals, is kept constant once the process root is fixed. In consequence the capacitance of interest cannot be directly measured.

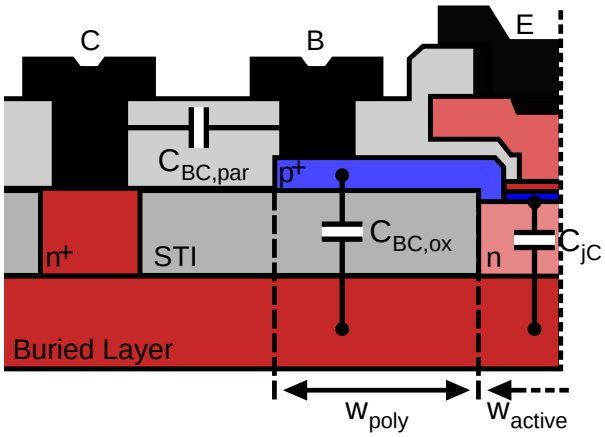


Figure 4.36: Device cross section with position of different C_{BC} contributions; $w_{poly} = f(d_{BE})$

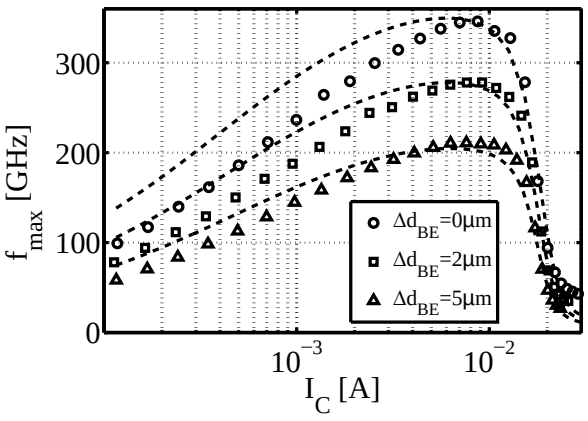


Figure 4.37: Impact of base-emitter separation width between contacts (d_{BE}) on RF FoM f_{max}

In an inverse geometry-scaling approach the basic idea behind the method is to keep inner dimensions of the vertical transistor constant while varying the extrinsic transistor layout. As shown in Tab. 4.3 two reference devices ($l_E = 5$ and $10\mu m$ at fixed $w_E = 0.2\mu m$) were layouted with variation of the extrinsic periphery (through Δd_{BE}) and manufactured on the same test mask as additional modeling structures.

Δd_{BE} \ l_E	5µm	10µm
0µm	$d_{BE} = 0.29\mu m$	$d_{BE} = 0.29\mu m$
2µm	$d_{BE} = 2.29\mu m$	$d_{BE} = 2.29\mu m$
5µm	$d_{BE} = 5.29\mu m$	$d_{BE} = 5.29\mu m$

Table 4.3: Set of symmetrical high-frequency transistors for estimation of parasitic BC overlap capacitance with variation of the separation distance (d_{BE}) between the contact terminals

In order to extract the oxide capacitance ($C_{BC,ox}$) one can assume that C_{jBC} for a given length l_E is constant and a function of bias (V_{BC}). The reference distance between the active emitter and the base ring connection is given by d_{ebcont} (cf. Tab. 3.2) with a nominal value of $290nm$ for the analyzed structures.

From measured data, the subtraction of the capacitance of the reference device ($\Delta d_{BE} = 0$) gives an averaged, bias independent capacitance offset per elongation (Δd_{BE}) as shown in Fig. 4.38, which is suited for a regression analysis of the linear relation

$$C_{BC,meas}(V_{BC}) = C_{jBC}(V_{BC}) + C_{BC,x}(C_{BC,ox}), \tag{4.53}$$

$$\Delta C_{BC}(\Delta d_{BE}) = C_{BC,meas}(V_{BC}, \Delta d_{BE}) - C_{BC,meas}(V_{BC}, \Delta d_{BE} = 0). \tag{4.54}$$

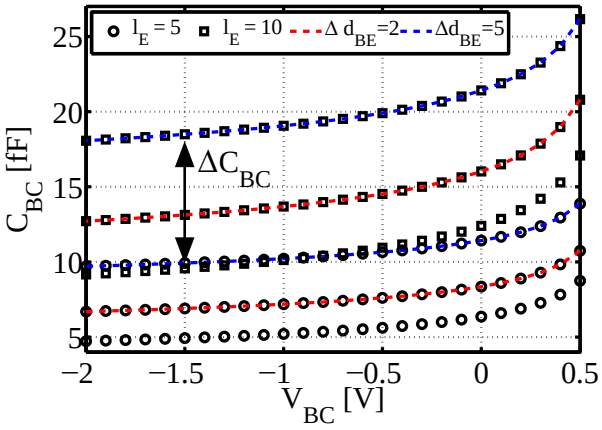


Figure 4.38: Measured total BC capacitance ($C_{BC,meas}$) with variation of overlap for two device lengths; *dimensions in µm*

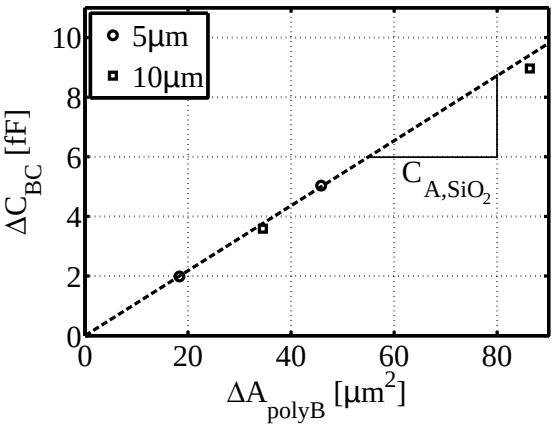


Figure 4.39: Scaling of extracted differential oxide overlap capacitance ($\Delta C_{BC,ox}$) with differential poly-base Area (ΔA_{polyB})

One can consider that the reference structure represents all (constant) non-uniformities of the active region as well as the non-uniform field distribution at the onset of the BE overlap region (cf. Fig. 4.41). Thus the slope of a regression forced through the origin, using the separated capacitance $\Delta C_{BC}(\Delta d_{BE})$ as a function of the respective area enclosed by the BE link region (cf. Fig. 4.40) yields the specific capacitance of the poly-base overlap region.

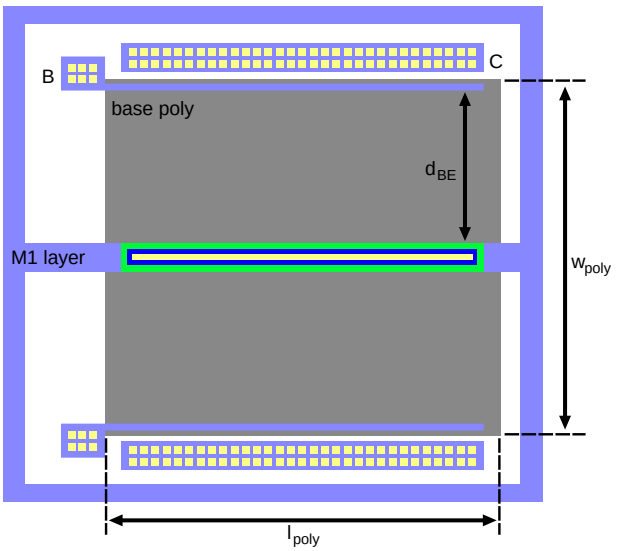


Figure 4.40: Top view of HBT structure for variation of poly base and BC overlap capacitance

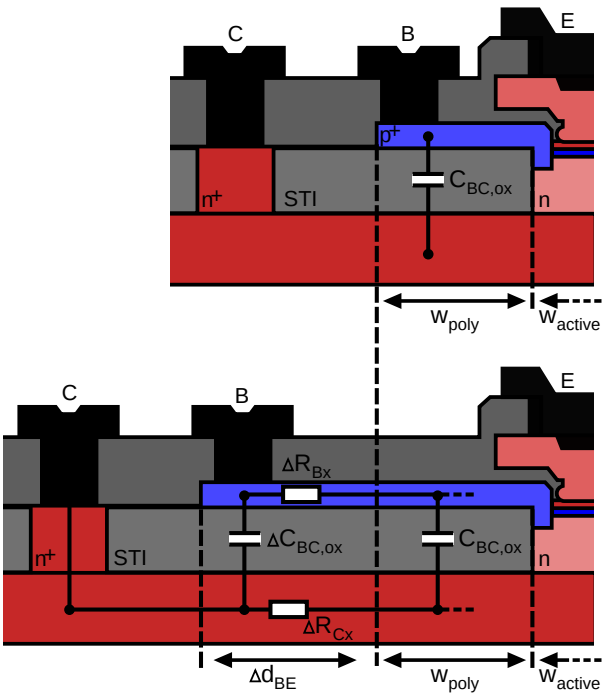


Figure 4.41: Comparison of device cross sections with additional poly-base width Δd_{BE}

The aim of the direct extraction of the oxide capacitance from manufactured silicon is to determine the oxide overlap capacitance from real device measurements. However numerical device simulations provide a useful method to compare theoretical results to the measured data. A simple two-dimensional structure was built based on the cross section as shown in Fig. 4.42. The thickness of the STI, being one of the most critical quantities for correct simulation, was determined from a TEM image analysis as shown in Fig. 4.43.

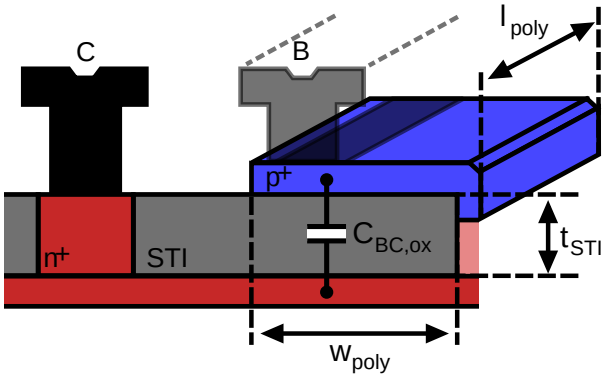


Figure 4.42: 3D cross section view of HBT BC overlap region for variation of w_{poly}

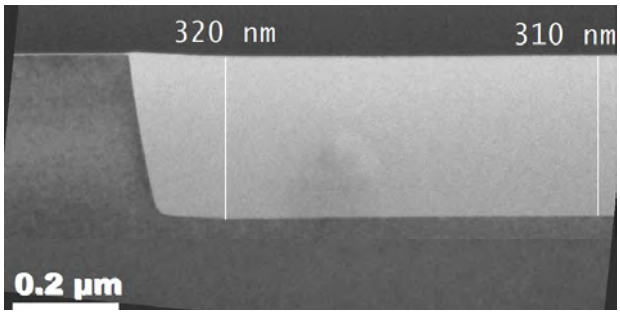


Figure 4.43: TEM imaging analysis of STI oxide layer separating base link and collector

The equivalent structure was built for and simulated in an electromagnetic field solver. The program POICAPS [109] numerically solves the Poisson equation for a given simulation scenario returning the desired field distribution as well as the specific capacitance as shown in Fig. 4.44.

A third estimate can be obtained using the simple plate capacitor-approach while completely ignoring non-idealities. Using the targeted STI thickness defined in the process ($t_{STI} = 340nm$) one can calculate the theoretical normalized surfasic oxide capacitance per area, using the specific material constants of the STI (relative permittivity $\varepsilon_{r,SiO2}$)

$$C_{A,SiO_2} = \frac{\varepsilon_{r,SiO2} \cdot \varepsilon_0}{t_{STI}} = \frac{3.9 \cdot 8.85\ \text{E-}18F}{0.340\mu m^2} \approx 0.102 \frac{fF}{\mu m^2}, \tag{4.55}$$

which comes very close to the experimental result obtained in the linear regression (cf. Fig. 4.39)

$$C_{A,SiO_2,xtr} = 0.109 \frac{fF}{\mu m^2}. \tag{4.56}$$

A 2D electrical field simulation of the BC overlap capacitance for multiple values of Δd_{BE} confirms the scaling of $C_{BC,ox}$ with a variation of the BE separation distance (cf. Fig. 4.45). Given the completely linear slope of the 2D simulation one can conclude, that the influence of the field distribution around the BC contact zone (considered as a constant offset) does not have any significant influence on the linearity of the BC oxide capacitance scaling ($C_{BC,ox}$).

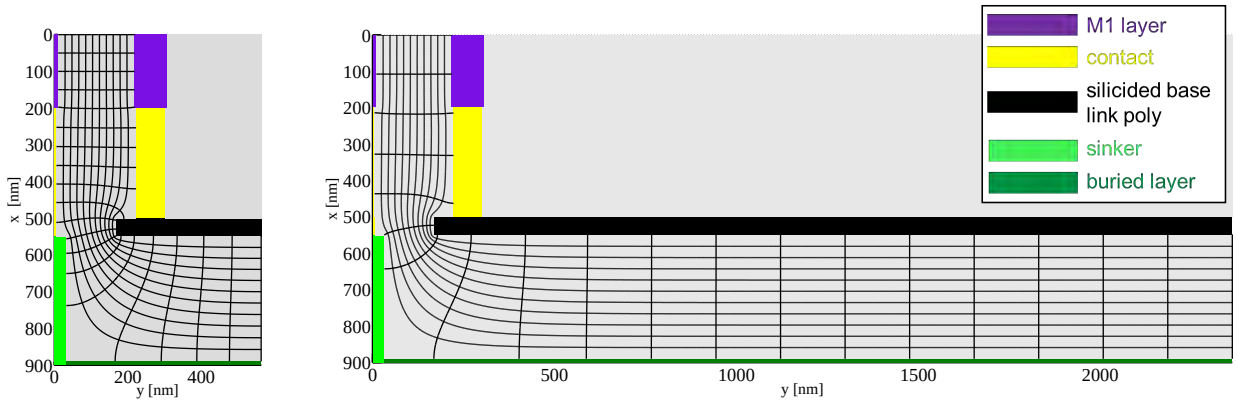


Figure 4.44: Electrical field simulation of BC overlap capacitance of standard device (left) and with additional poly-base width (right, $\Delta d_{BE} = 2\mu m$)

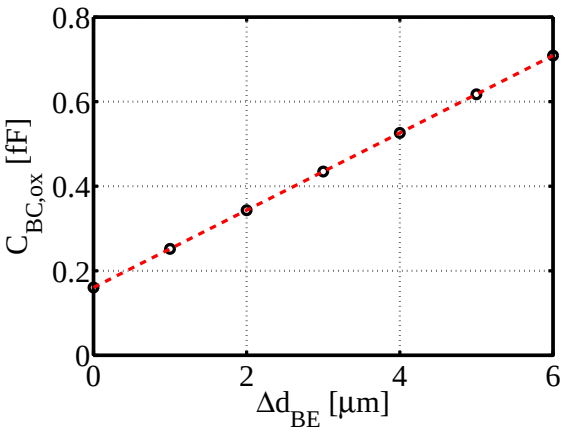


Figure 4.45: Scaling of simulated oxide capacitance with separation distance offset Δd_{BE}

4.5.2 Parasitic BE Spacer Capacitance

Like for the case of the BC overlap capacitance, the RF devices dedicated to model parameter extraction are not suitable for extraction of the parasitic BE capacitance neither. Due to the small separation of base and emitter by means of the L-shaped inside spacer, the most part of the extrinsic BE capacitance originates from the direct periphery of the BE junction rather than the polybase-link or the contact. Accurate on-wafer measurements for the respective zone are desirable to confirm results from theoretical calculus or numerical device simulation.

A trial with dedicated RF structures was performed by means of a requested layout variation with modification of the BE junction. The intention was do design the emitter window with its minimal dimensions allowed in the PDK ($w_E = 0.08\mu m$) in order to have the external elements of the BE junction periphery properly manufactured in silicon. However given the spacer width of a single side of the spacer of about $d_{SK} \approx 40nm$ the intended result was a device with closed emitter window (B and E region isolated by inside spacers) and thus removing the (bias dependent)

capacitance contribution of the BE junction ($C_{j,BE} = 0$) leaving exclusively $C_{BE,meas} = C_{BE,par}$ as shown in Fig. 4.46.

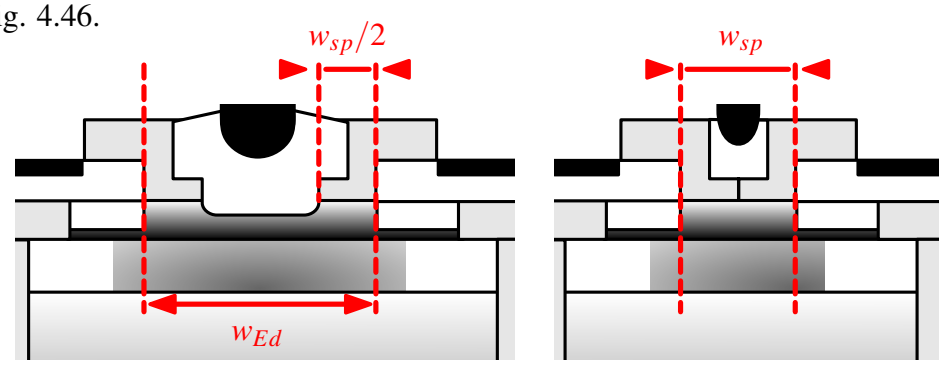


Figure 4.46: Layout variation with modification of the BE junction through reduction of emitter window width; a) $w_E = w_{E,min}$, b) $w_E \approx w_{sp}$

However after silicon processing it was found that the BE junction was still active and the transistor operated normally. In consequence a geometry-scaling approach to estimate the spacer capacitance had to be applied. For each structure the measured zero-bias capacitance ($C_{j,BE0}$) is extracted from several different emitter widths (Fig. 4.47) by assumption of a bias-independent oxide capacitance. Assuming that the capacitance $C_{j,BE}$ scales linearly with the junction width (or bottom area respectively through $A_{j,BE} \propto w_E$) one can deduct a zero-emitter width value ($C_{BE,0}(w_E = 0)$) from the intersect of a regression line with the y-axis (cf. Fig. 4.48). The spacer on the other hand is having a constant dimension. Thus the extrapolated zero-emitter window-value normalized by the emitter periphery length (P_E) yields the lineic spacer capacitance from the slope of a linear regression versus w_E, giving an estimate ($C_{l,SP} = 1.82 fF/\mu m$, cf. Fig. 4.49) with the relation

$$C_{BE,sp} = C_{l,SP} \cdot P_E. \tag{4.57}$$

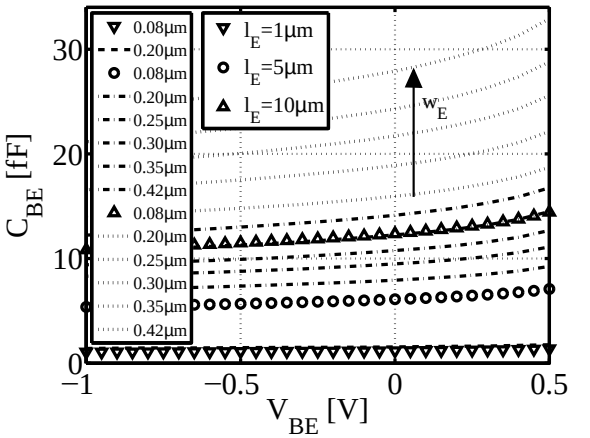


Figure 4.47: BE capacitance scaling versus bias for three different emitter length configurations; *drawn dimensions:* $l_E = 1, 5, 10 \mu m$

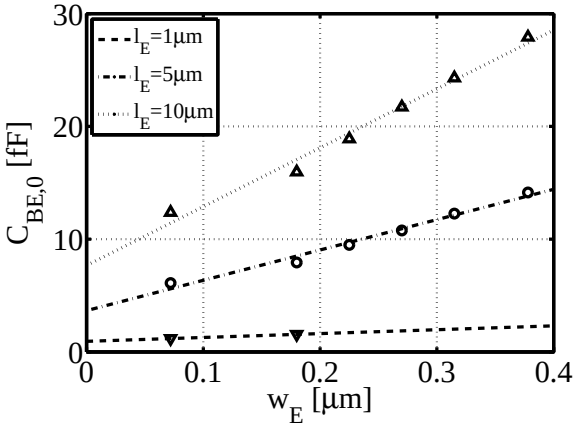


Figure 4.48: Extraction of BE spacer related capacitance at intersect of linear regression from emitter width scaling

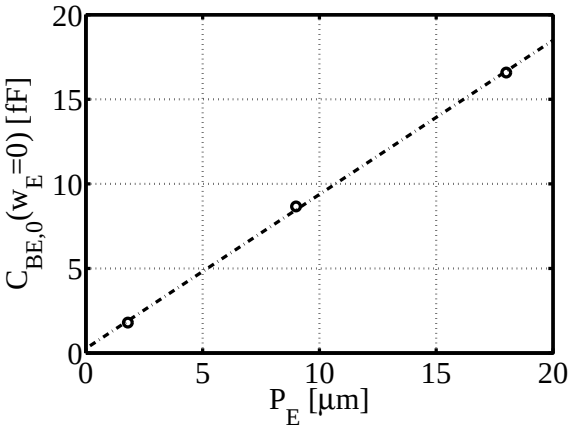


Figure 4.49: Scaling of extracted spacer capacitance ($C_{BE,0}(w_E = 0)$) with emitter perimeter

A comparison with simulated results may serve as a confirmation of the results obtained from the scaling approach. However the actual geometry of the BE spacer region as well as the material composition has to be accurately known. Based on a TEM imaging procedure (cf. Fig. 4.50), the spatial dimensions are determined and a corresponding equivalent structure was built for simulation in a electromagnetic field solver.

The program POICAPS [109] was employed to numerically solve the Poisson equation for the given simulation scenario, returning the desired field distribution as well as the specific capacitance as shown in Fig. 4.51. A sample input file is found in the appendix (cf. App. N). With a total plate-to-plate capacitance of $C_{l,SP,sim} = 1.86 fF/\mu m$ the obtained value from the field simulation corresponds very well with the result obtained by means of the geometry scaling approach.

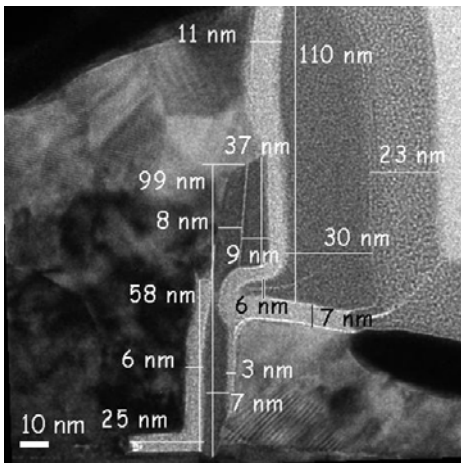


Figure 4.50: TEM imaging analysis to determine dimensions of the BE inside spacer

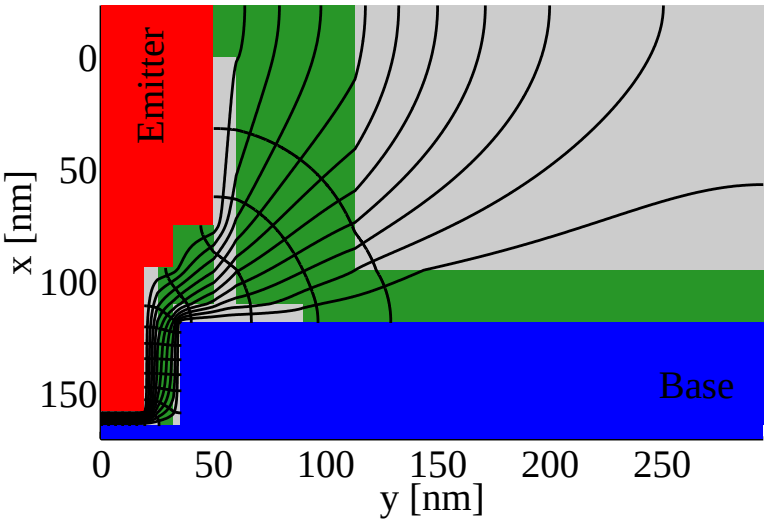


Figure 4.51: Electrical field simulation of inside spacer region between emitter (*blue*) and base (*red*) terminal with differentiation of nitride (Si_3N_4, *green*) and oxide (SiO_2, *grey*)

4.5.3 BE Junction Depletion Capacitance C_{BE}

One of the most critical junctions for proper modeling of RF characteristics is the base-emitter junction. Thus the extraction flow is given in detail for the junction capacitance C_{BE}.

For extraction of geometry scalable junction capacitances a good aspect ratio variation is reached using width scaling (at fixed l_E) rather than length scaling. The parameters describing the bias dependence are assumed to be constant (unitary) for all modeled capacitances leaving only the zero-bias junction capacitance C_{j0} of the respective junction as scalable model parameter to be determined.

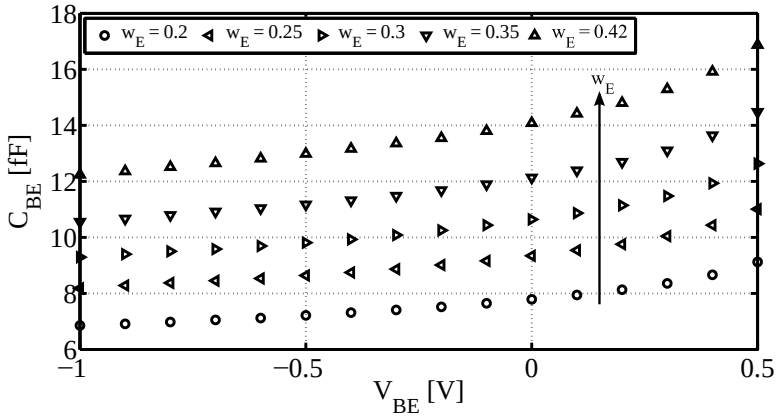


Figure 4.52: Geometry scaling of (averaged) C_{BE} junction capacitance for a selected set of symmetrical devices with $l_E = 5\mu m$ vs. V_{BE} bias for different w_E

Once the zero bias value for each geometry is obtained, a normalization of all curves can be performed by means of the zero-bias value (C_{BE0}) for each respective geometry (cf. Fig. 4.53). A superposition of all normalized curves then gives the curve progression versus junction bias suitable for linear regression analysis.

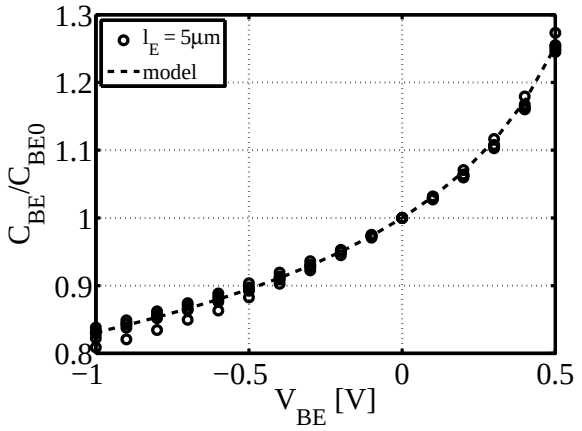


Figure 4.53: Measured BE junction capacitance normalized by zero-bias value C_{BE0} versus bias for various emitter window widths (w_E); *symmetrical devices at fixed* $l_E = 5\mu m$

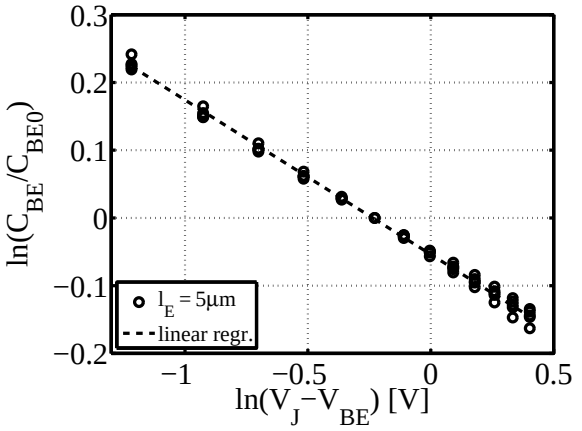


Figure 4.54: Non-linear optimization of bias parameters for BE junction capacitances; *symmetrical devices at fixed* $l_E = 5\mu m$

Afterwards the non-linear equation (4.46) is optimized using the normalized capacitance data. The averaged data is used for analysis of the correlation coefficient r (cf. Fig. 4.54). The maximization of this coefficient gives a best fit for the unitary parameters of the extraction equation

$$\ln(C) = \ln(C_{j0} \cdot {V_D}^z) - z \cdot \log(V_D - V). \tag{4.58}$$

In a best case (after successful optimization of both parameters describing the bias dependence (V_D and z_j)) the regression line and the normalized measured data of all devices (for a defined set of geometries) are superimposed.

Even though the method is not actually a direct extraction approach, reproducibility using this nonlinear regression procedure is found to be reasonably good. Given the fact that the bias dependent parameters are not scaling with device geometry (cf. section 3.3.1) only the partitioning of area-related and perimeter-component remains. The P/A separation approach is used as shown in Fig. 4.55 for splitting the inner and peripheral base-emitter capacitances.

In general capacitances are split as presented in eqn. (4.50) ff.. From slope and intercept of a plot as shown in Fig. 4.55 one can extract the two contributions C_{jBEa} and C_{jBEp} by means of a regression analysis of the zero-bias values for a selected variety of device geometries and configurations.

In HICUM the base-emitter capacitance partitioning between intrinsic and extrinsic part (parasitic BE isolation capacitance, C_{BEpar}) is then taken into account by the partitioning factor F_{BEpar}

$$F_{BEpar} = \frac{C_{BEpar,2}}{C_{BEpar}}. \tag{4.59}$$

In a last step a verification using all available geometries is exercised in order to verify the quality of the model (cf. Fig. 4.56).

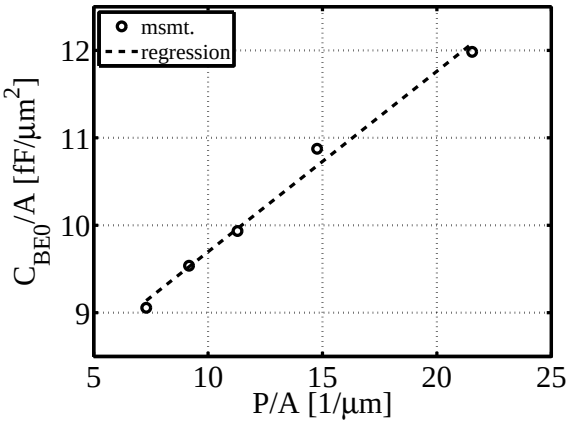


Figure 4.55: Extraction of geometry-scalable BE junction capacitance (*partitioning in C_{BE0a} and C_{BE0p})*

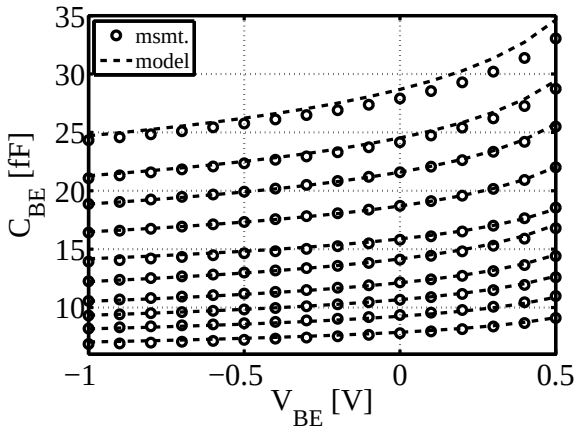


Figure 4.56: Model verification of scalable parameter extraction versus bias for symmetrical devices

Even though the modeling of depletion charges Q_j and capacitances C_j depending on the bias voltage V across the junction is based on classical theory, the model code uses more complex equation. Components for large forward, medium and reverse bias are linked by smoothing functions in the model code. In order to avoid division by zero at high forward bias, a_j is defined as the ratio of the maximum value relative to C_{j0}. This limit on C_j to a maximum value is implemented in order to maintain consistency between measured characteristics and the model. The role of the additional model parameter a_j can be seen in Fig. 4.57.

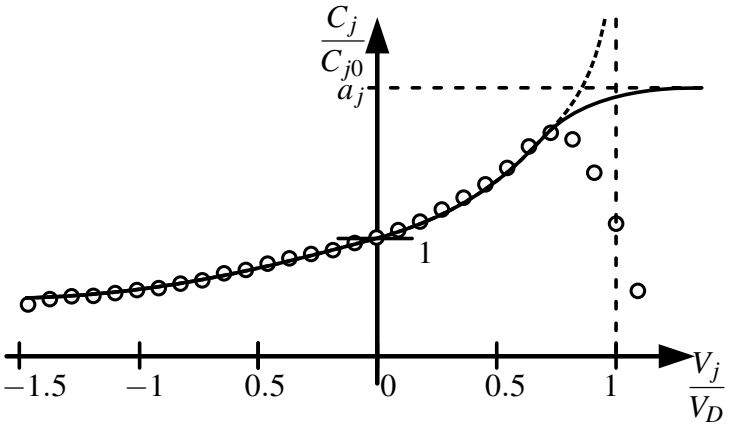


Figure 4.57: Normalized measured depletion capacitance versus bias in comparison with model equations; *Points: measurement, dashed line: classical model equation, solid line: refined model equation with smoothing functions found in HICUM*

4.5.4 BC Junction Depletion Capacitance C_{BC}

The same procedure as outlined for the BE capacitance in sect. 4.5.3 is applied to the measurement data of the BC junction capacitance. As compared to the BE junction the doping ratio of base and collector region is significantly different. In consequence the resistance to junction breakdown is higher for the BC diode and the measurement range under reverse bias is extended to $V_{BC,min} = -2V$ for NW15.

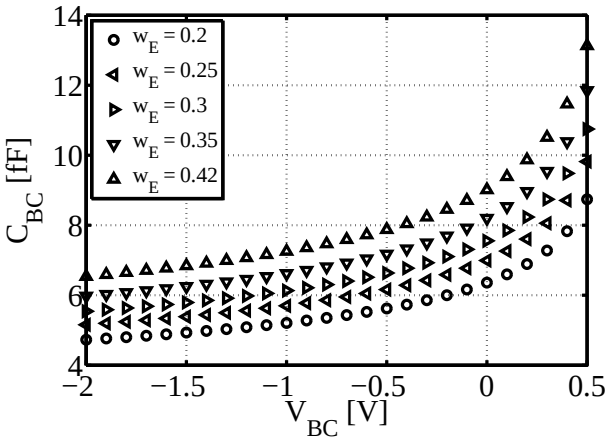


Figure 4.58: Measured BC junction capacitance vs. bias for various emitter window widths w_E [in μm]; *symmetrical devices at fixed* $l_E = 5\mu m$

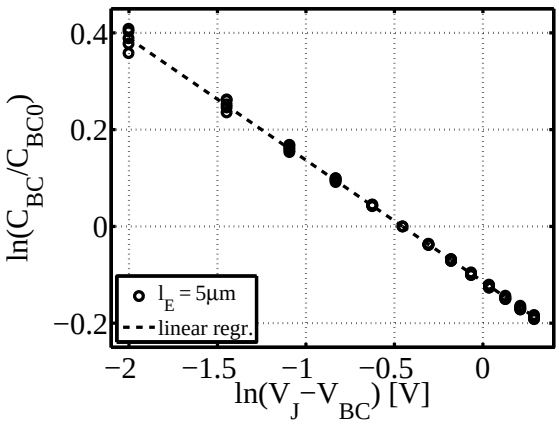


Figure 4.59: Non-linear optimization of bias parameters for BC junction capacitances; *symmetrical devices at fixed* $l_E = 5\mu m$

After correction of back-end parasitics the zero bias value C_{BC0} is extracted from Fig. 4.58. Again the classical model equation in its logarithmic form is used for optimization of the correlation coefficient of the normalized junction capacitances C_N (cf. Fig. 4.59)

$$\begin{aligned} C_N &= \frac{C_{BC}}{C_{BC0}} = \frac{1}{\left(1-\frac{1}{V_D}\right)^z} = \frac{V_D^z}{(V_D-V)^z} \\ \log C_N &= z\cdot\log V_D - z\cdot\log(V_D-V) \end{aligned} \tag{4.60}$$

A P/A separation to obtain the geometry scalable contribution of C_{BC} (cf. Fig. 4.60) is then followed by the verification with a multitude of different measured geometries (cf. Fig. 4.61).

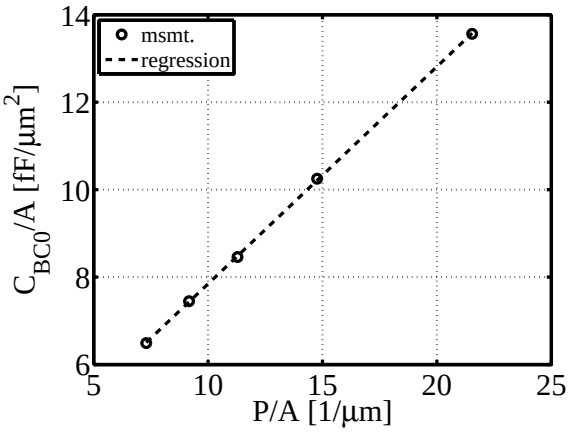


Figure 4.60: Extraction of geometry-scalable BC junction capacitance with partitioning in C_{BC0a} and C_{BC0p} from intersect and slope

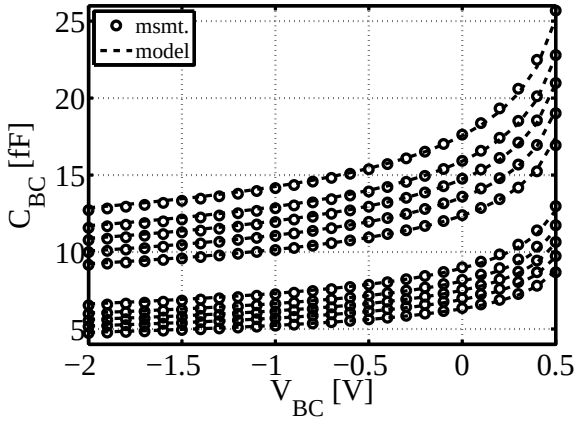


Figure 4.61: Model verification of scalable parameter extraction versus bias for symmetrical devices

4.5.5 CS Junction Depletion Capacitance C_{CS}

In most modern BiCMOS technologies, the basis material for a vertical NPN bipolar transistor used in high-speed applications is a p-doped silicon substrate (cf. section B.1 and Fig. B.1). The substrate is doped very lightly (typically around $N_A \approx 10^{14} cm^{-3}$) giving it a p^- designation. Using an oxide mask the n-type buried layer (BL) is realized by means of an implant process. The highly n^+-doped collector sinker serves as a low-resistance connection to the collector terminal. Given these process steps a pn junction between substrate and collector is formed in vertical direction.

The n-type collector is isolated from the substrate by a (normally) inverse-biased diode. Given the presence of the n^+ buried layer and the p region underneath, C_{CS}, considered as a parasitic capacitance, is modeled as a nonlinear junction capacitance. Since in forward transistor operation the collector-substrate junction is reverse-biased, the capacitance of the collector-substrate junction is usually dominated by the depletion capacitance.

With the area of the collector-substrate junction being considerably larger than the emitter area, the zero-bias value of this parasitic capacitance can be substantial and might dominate the performance of the device. Even though efforts are made to reduce substrate parasitics, the value of C_{CS0} for a given device may reach values of the base-emitter capacitance C_{BE} depending on the used technology. Advanced processes reduce its impact by providing a deep trench isolation (DTI) between adjacent devices or isolating substrates.

The most simple representation of the substrate would be a resistive component linking the substrate node of the vertical BJT to the contact. However for high frequency application the collector-substrate diode as well as coupling needs to be taken into account. Hence the substrate network of BJTs in general includes the substrate-collector depletion capacitance C_{CS}, substrate

resistance R_{SU} (due to the substrate resistivity) and the substrate capacitance C_{SU} (due to the permittivity of the substrate) as seen in the cross section (cf. Fig. 2.10).

In the HICUM model the substrate-collector depletion capacitance C_{CS} is modeled as a junction capacitance and is thus the third bias-dependent capacitance in SiGe HBT devices. With the bias applied to the collector terminal, measured data can be obtained from S-parameter measurement of the NW15 configuration (using eqn. (G.7)) with the simplified π equivalent circuit (cf. Fig. G.1). However due to the very small zero-bias capacitance values measured between substrate and collector node the obtained results are very noisy and likely prone to failure.

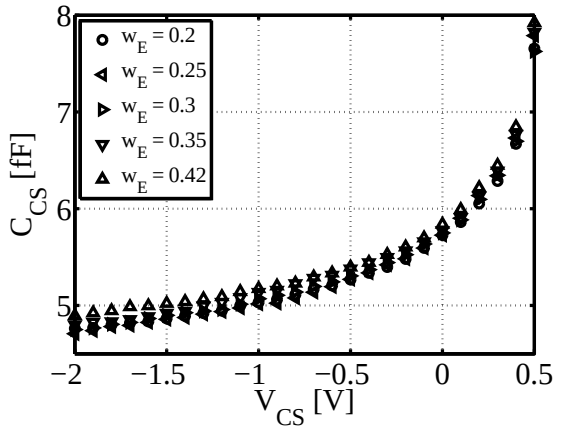


Figure 4.62:]
Measured CS junction capacitance vs. bias for various emitter window widths w_E [in μm]; *symmetrical devices at fixed* $l_E = 5\mu m$

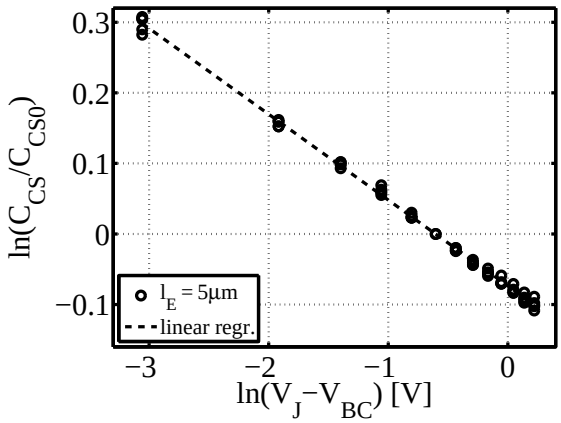


Figure 4.63: Non-linear optimization of bias parameters for CS junction capacitances; *symmetrical devices at fixed* $l_E = 5\mu m$

In contrast to the aforementioned BE and BC contributions, the CS junction capacitance does not directly scale with the emitter area but with the large bottom area enclosed by the DTI (cf. eqn. 3.61). Thus even for a significant variation of the emitter window area the total bottom area of the buried layer remains almost constant. In order to achieve good geometry scaling the approach of taking just one single device length (e.g. $l_E = 5\mu m$) with a variation of the device width is of limited accuracy for the substrate capacitance. Therefore a set of devices with two different lengths has been selected (since $l_{BL} \propto l_{cd} = l_{bd} + 2 \cdot d_{abl}$, cf. eqn. (3.40)).

General scalability seems to be nicely reproduced by a linear regression (cf. Fig. 4.64), however it is to note that the Y-intercept and thus the obtained bottom-area related component of C_{CS} is negative, which in general is not physical.

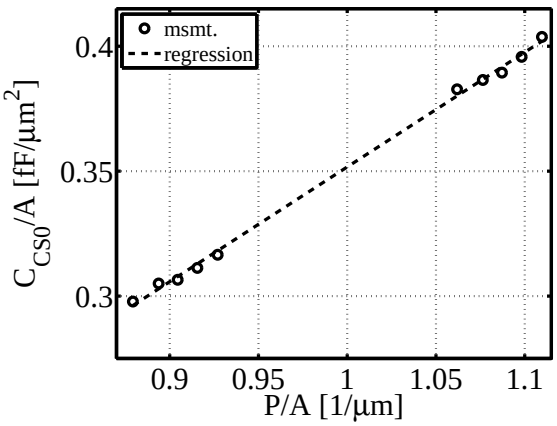


Figure 4.64: Extraction of geometry-scalable CS junction capacitance with partitioning in C_{CS0a} and C_{CS0p} from intersect and slope

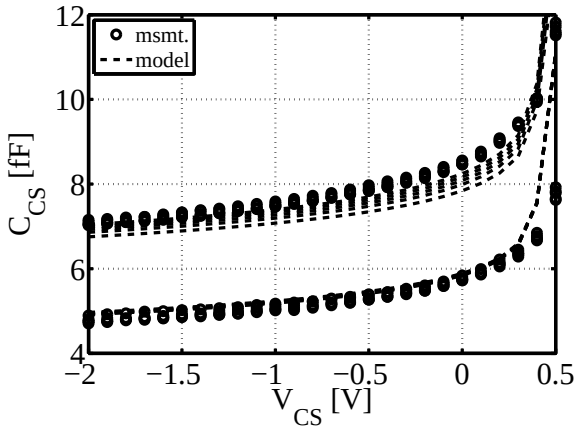


Figure 4.65: Model verification of scalable parameter extraction versus bias for symmetrical devices with different lengths l_E

4.5.6 Collector-Substrate Network

The parasitic influence of the substrate can lead to a significant performance degradation in advanced high-speed and RF circuits. Hence, careful circuit layout is necessary, and shielding measures such as guard rings or trench isolation must usually be employed. However many different technology-related options such as high resistivity substrates, different transistor isolation techniques as well as shielding methods are leveraged in state-of-the-art technologies.

The intra-device substrate coupling is taken into account in the HICUM model by a dedicated network as seen in the equivalent circuit (cf. Fig. 2.10). However there exist several different process variants wrt. to the manufacturing of the substrate connection. The most popular methods are a thick field-oxide, a shallow trench (STI) approach as well as a separation using a deep trench (DTI). For mmW-applications the combination of a deep trench and a buried sub-collector leads to reduced collector capacitance (C_{CS}) and collector resistance (R_{Cx}) parasitics in order to reach highest f_T and f_{MAX} values.

Yet all of the shielding methods have to be covered by a single representation used in the compact model. In order to find the most suitable model, two different approaches are most common: The junction isolation (JI) scheme is used in many process technologies where no deep trench (DTI) is manufactured for substrate isolation. This comes at the expense of a larger transistor layout and less isolation yet does not require additional steps and masks, which significantly lowers the manufacturing cost.

However the deep trench isolation approach is the predominant isolation scheme found in modern bipolar technologies. A good electrical isolation of the individual transistors in integrated circuits allows for significant reduction of proximity and increased circuit density. The deep trench

(filled with an insulator material) is etched deep into the substrate in a first process step limiting the extent of both n-regions (epitaxial and buried layer). The filling of the DTI here is less important yet due to the desired high aspect ratio (few μm in depth) it had to advance with processes as well. Generally it consists of silicon dioxide, a dielectric polyimide or a polysilicon filler. In the analyzed state-of-the-art process a thick oxide liner and undoped poly-silicon is employed to form the DTI with the inherent advantages of:

- reduced device footprint allowing to increase circuit density
- reduced intra-device coupling
- reduced bottom area of the collector substrate capacitance (C_{CS}) lowering switching delay

The goal of the presented extraction procedures is to enable easy yet accurate extraction of the substrate network elements from standard cold S-parameter measurements without the need for special test structures nor increased measurement effort or silicon area.

The Substrate Equivalent Circuit

The popular approach of a network representation with the bulk substrate (resistance R_{SU} and a capacitance C_{SU}) in series with the junction capacitance C_{CS} is introduced in the HICUM model. The according device cross section is shown in Fig. 4.66 along with the equivalent circuit.

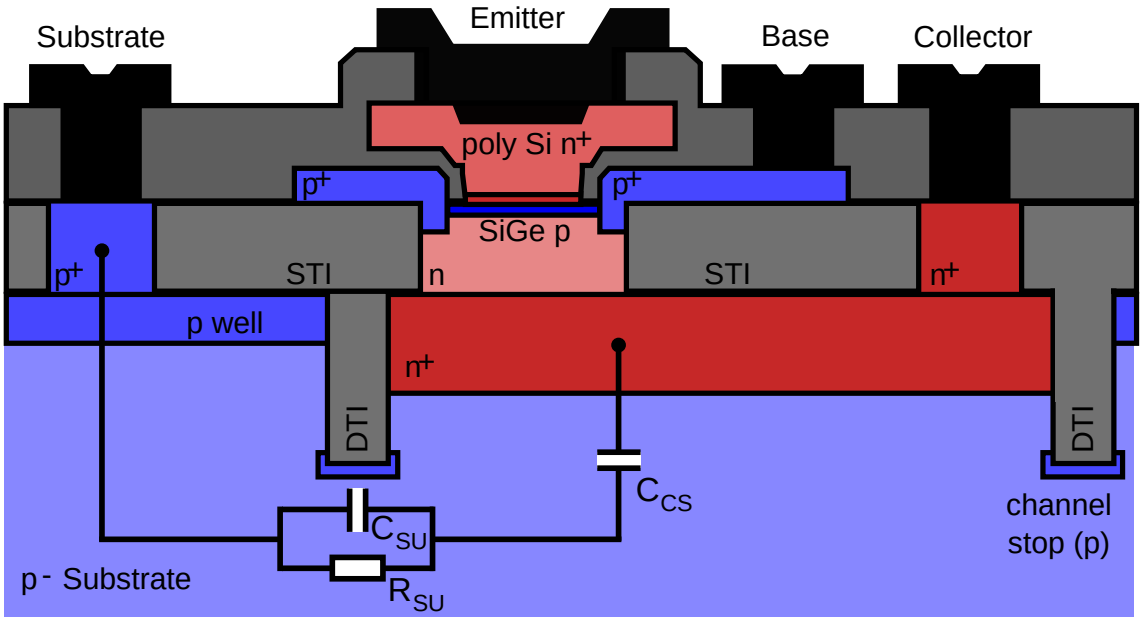


Figure 4.66: Cross section of HBT device in DTI technology with substrate contact and network; textitsingle sided CBE configuration

Substrate Network Parameter Extraction Strategy

For parameter extraction a first step is to find a relation of measured S-parameter to the equivalent circuit elements. The fundamental equation used to extract the substrate complex can be derived from the HICUM equivalent circuit or the simplified representation (cf. Fig. G.5) to the following form suitable for a linear regression:

$$Y_{SUB} = Y_{22} + Y_{12} \text{ and } Z_{SUB} = \frac{1}{Y_{SUB}} \tag{4.61}$$

$$\Re(Z_{SUB}) = \frac{R_{SU}}{1+(\omega \cdot C_{SU} R_{SU})^2} \tag{4.62}$$

$$\frac{1}{\Re(Z_{SUB})} = \frac{1}{R_{SU}} + R_{SU} \cdot (\omega_0 C_{SU})^2 \left(\frac{\omega}{\omega_0}\right) = a_{SU} + b_{SU}\left(\frac{\omega}{\omega_0}\right), \text{ with} \tag{4.63}$$

$$R_{SU} = \frac{1}{a_{SU}} \text{ and } C_{SU} = \frac{\sqrt{a_{SU} b_{SU}}}{\omega_0} \tag{4.64}$$

The required extraction frequency ω_0 is a user-defined value and as best practice a reasonably low value of $\omega_0 = 1GHz$ yields good results. Once the two substrate-parameters are known the junction capacitance may be directly calculated using

$$C_{CS} = -\left[\omega \Im(Z_{SUB}) + \frac{1}{C_{SU}} \frac{(\omega \cdot C_{SU} R_{SU})^2}{1+(\omega \cdot C_{SU} R_{SU})^2}\right]. \tag{4.65}$$

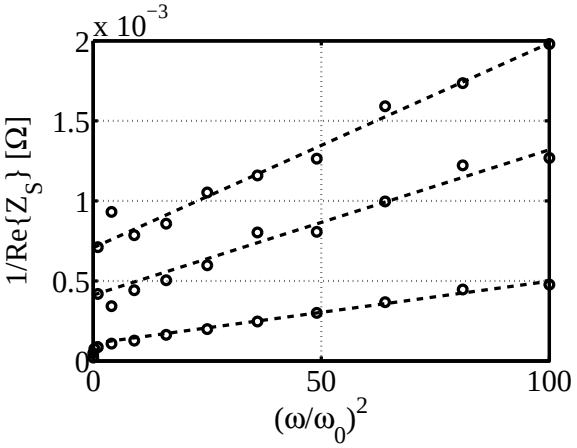


Figure 4.67: Regression analysis of substrate network for different bias values of CS junction (V_{SC}); *single device with 0.2x5µm²*

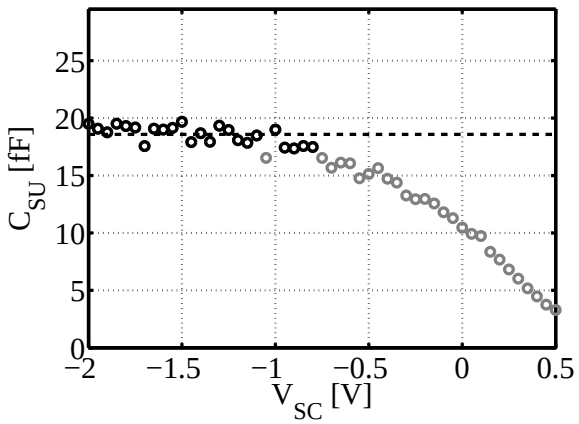


Figure 4.68: Direct extraction of the constant substrate capacitance C_{SU} as function of CS junction bias (V_{SC})

The extraction results from regression analysis as shown in Fig. 4.68 and 4.69 show a strong bias-dependence, thus the extraction of constant contributions (R_{SU} and C_{SU}) are preferably analyzed under strong reverse bias ($V_{SC} < -0.5V$). The parameter of the bias dependent junction

capacitance C_{CS} however can be determined using the entire bias range (cf. Fig. 4.70).

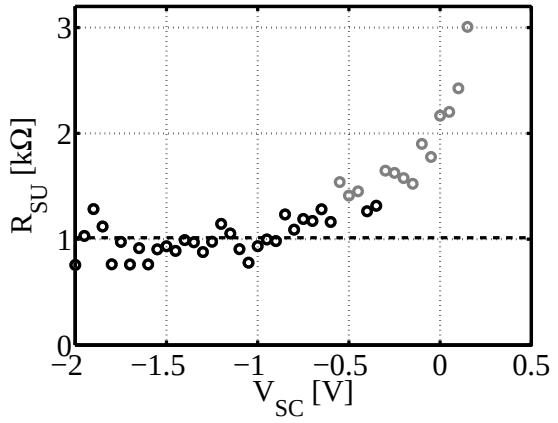


Figure 4.69: Direct extraction of substrate resistivity R_{SU} as function of CS junction bias (V_{SC})

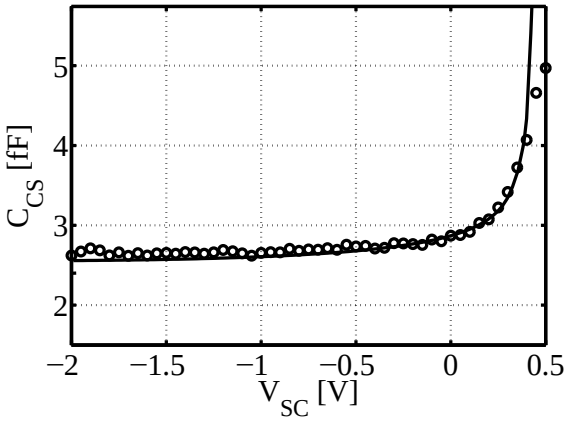


Figure 4.70: Extraction of bias-dependent C_{CS} junction capacitance over bias range (V_{SC})

Another approach directly related to the measured data is the use of a frequency approximation as shown in Fig. 4.71, in order to obtain the desired model parameters by means of extrapolation.

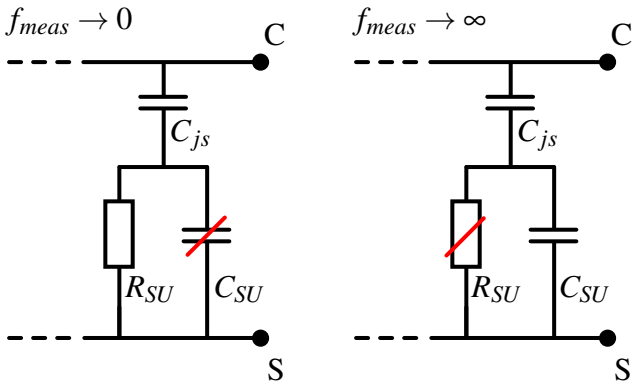


Figure 4.71: Extrapolation of equivalent circuit elements for trench isolation approach as a function of the measurement frequency

For low frequency one can assume the parallel capacitor C_{SU} to be shunt leaving exclusively a series connection of R_{SU} and C_{CS} that can be separated by using the real part of the measured substrate reactance ($\Re(Z_{SUB})$). In contrast, towards high frequency the resistance R_{SU} is assumed to be shunt by the parallel capacitor which gives the series configuration of C_{CS} and C_{SU}.

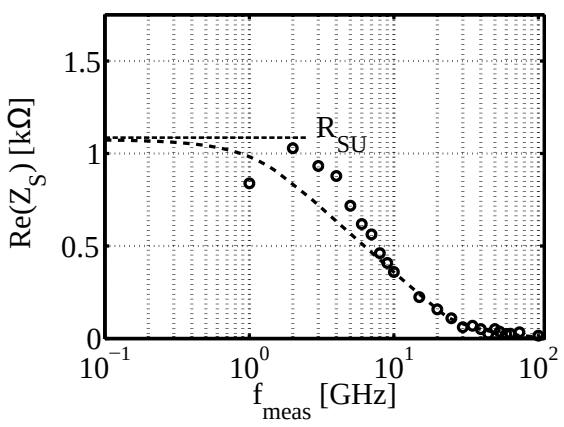


Figure 4.72: Extrapolation of the substrate resistance R_{SU} towards low measurement frequency ($f_{meas} \rightarrow 0$); *single device with 0.2x5µm²*

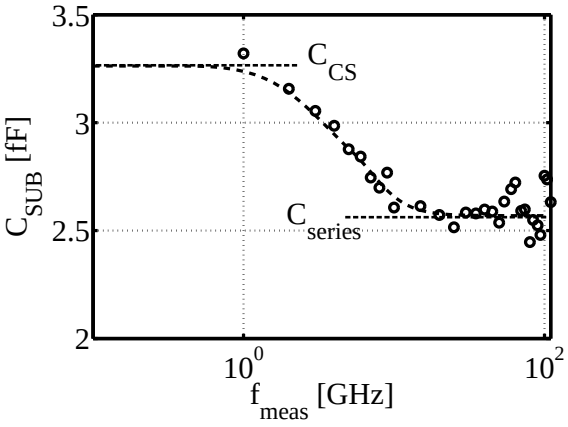


Figure 4.73: Extrapolation of the substrate capacitances C_{SU} and C_{CS} from upper and lower limit of measured frequency range; *single device with 0.2x5µm²*

The properties of the substrate elements C_{SU} and R_{SU} are strongly related to the material composition and linked through the characteristic substrate cut-off frequency. By means of a constant substrate transit time ($\tau_{Si} = \rho_{SU} \cdot \varepsilon_{Si}$), the cut-off frequency of the simple RC network is defined

$$f_c = \frac{1}{2\pi\varepsilon_{Si}\rho_{SUB}} = \frac{1}{2\pi C_{SU} R_{SU}}. \tag{4.66}$$

Once a value for ρ_{SUB} is fixed (cf. Tab 3.3), the scaling within the substrate can be simplified.

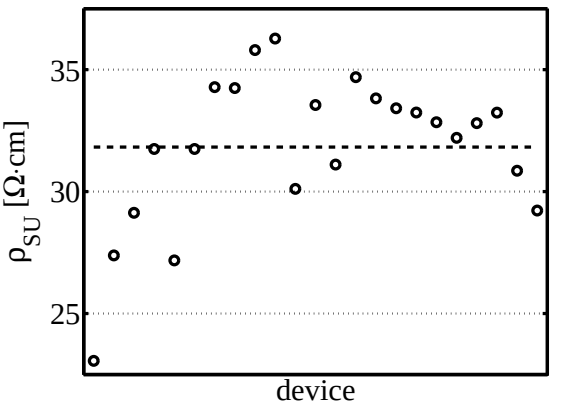


Figure 4.74: Verification of constant product ρ_{SUB} for a number of different structures (device geometries)

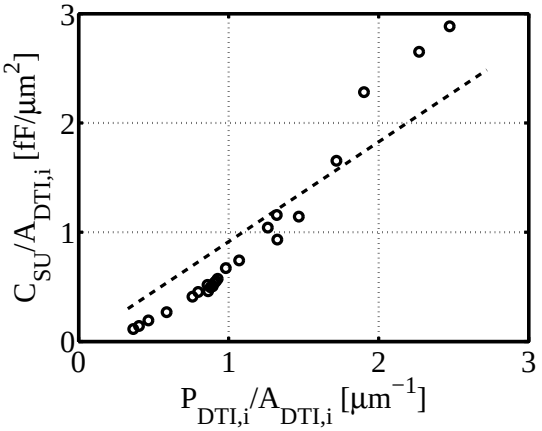


Figure 4.75: Extraction of geometry scaling with a regression forced through the origin ($C_{SU,A} = 0$)

For verification two exemplary geometries are shown with their respective frequency-dependence. As evidenced by the good agreement of the simulated and measured characteristics the native

equivalent circuit implemented in the HICUM model is found to provide sufficient flexibility for modeling of DTI technologies.

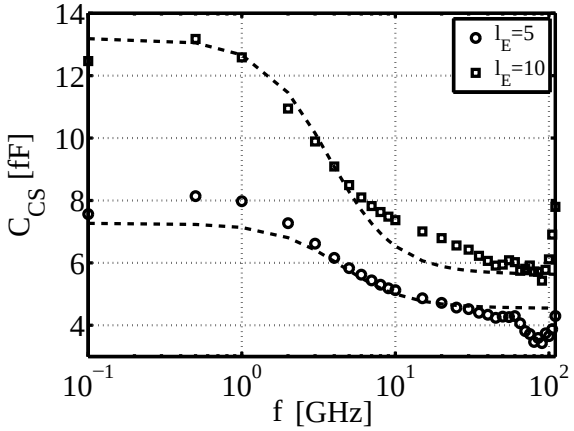


Figure 4.76: Verification of geometry scaling using two geometries for Z_{SUB} network versus frequency; $w_E = 0.2\mu m$, *dashed line: HICUM*

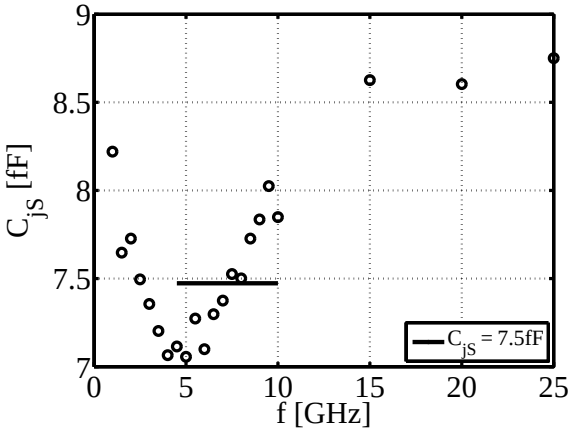


Figure 4.77: Result of direct extraction of C_{jS} as function of measurement frequency; *black line: selected range for averaging*

4.5.7 Variation of the Substrate Network Topology

In addition to a verification of the suitability of the equivalent circuit used in HICUM, the influence of the substrate connection method and layout on the device characteristics has been evaluated. This trial bases on a variation of the employed connection scheme in which the substrate node is realized. A variety of different device configurations was manufactured based on the standard RF device structure with drawn dimensions of $w_E = 0.2\mu m$ and $l_E = 5.0\mu m$.

By default the substrate terminal is connected through a highly p^+ doped ring on the outside of the DTI, with a constant unidirectional (lateral x and y direction) spacing of $d_{DTI,S}$ (cf. Fig. 4.78).

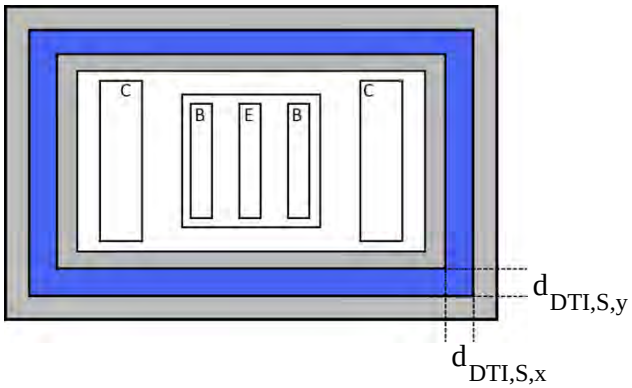


Figure 4.78: Top view of default HBT device with ring-shape substrate connection; *separation distance $d_{DTI,S}$ between substrate (p+) and DTI*

Variation of the Substrate Ring Separation Distance $d_{DTI,S}$

As shown in Tab. 4.4, additional structures were layouted with a variation of the separation distance ($d_{DTI,S}$), while keeping the substrate ring configuration as well as omission of individual sides of the substrate by exclusive connection through two parallel or perpendicular (wrt. emitter direction of l_E) stripes or a single parallel and perpendicular stripe as shown in the respective top views (cf. Fig. 4.79).

$\Delta d_{DTI,S}$ / configuration	$0\mu m$	$2\mu m$	$5\mu m$
ring substrate	$d_{S,S,x}$ = 4.99μm, $d_{S,S,y}$ = 8.13μm	$d_{S,S,x}$ = 8.99μm, $d_{S,S,y}$ = 12.13μm	$d_{S,S,x}$ = 14.99μm, $d_{S,S,y}$ = 18.13μm

Table 4.4: Different substrate trials with variation of separation distance $d_{DTI,S}$ between substrate ring (p+) and DTI and substrate ring configuration; *drawn dimensions in deyice layout between inside edges of substrate plug*

Additional configurations were manufactured with in the following variations (cf. Fig. 4.79):

- two substrate contacts, parallel to emitter window (cf. Fig 4.79 a))
- two substrate contacts, perpendicular to emitter window (cf. Fig 4.79 b))
- one substrate contact only, parallel to emitter window (cf. Fig 4.79 c))
- one substrate contact only, perpendicular to emitter window (cf. Fig 4.79 d))

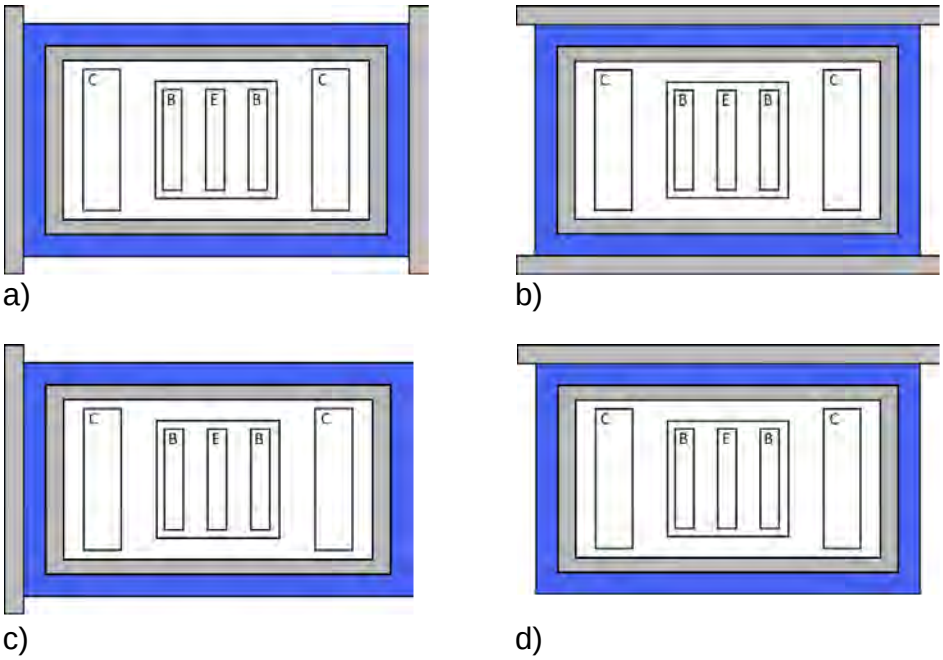


Figure 4.79: Top view of different substrate trials with variation of substrate connection;
a) and c) for double and single parallel, b) and d) for double and single perpendicular connection

In total a number of 14 structures of topology variations with respective SHORT and OPEN deembedding dummy were available for on-wafer RF measurement.

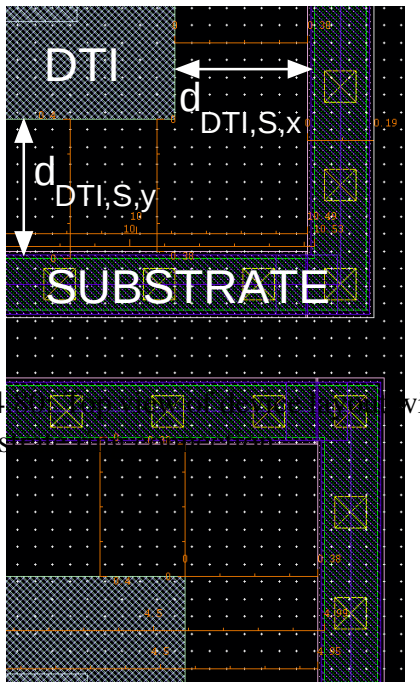


Figure 4.80: [illegible] with DTI and subs[illegible]

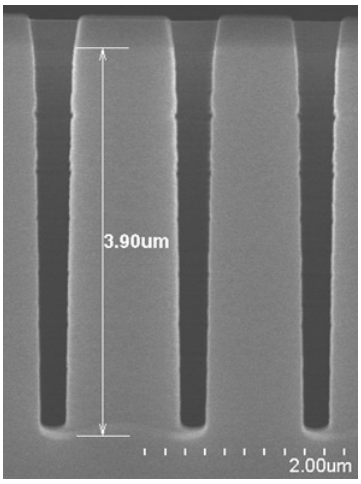


Figure 4.81: Cross sectional view of DTI structure from TEM imaging analysis

The results of parameter extraction for all of the 15 device structures, including the results of the reference layout, are summarized in Tab. 4.5. From a variation of the separation distance one would expect a significant variation of the series resistance R_{SU} due to the elongation of the resistivity path between substrate contact and the device periphery. A similar study has been exercised for HBTs without DTI ring [110]. Yet as evidenced by the summary table comparing the characteristic FoMs of the substrate, neither an increased separation nor the topology change has a significant impact in the used DTI technology. The most sensitive parameter seems to be the substrate capacitance (C_{SU}) whereas the variation of the resistivity R_{SU} and CS junction capacitance C_{CS} vanish within the accuracy of the parameter extraction or measurement resolution respectively.

structure, [$\Delta d_{DTI,S}$ in μm]	$R_{SU}[k\Omega]$, double	$C_{SU}[fF]$, double	$C_{CS}[fF]$, double	$R_{SU}[k\Omega]$, single	$C_{SU}[fF]$, single	$C_{CS}[fF]$, single
ring substrate, [0]	2.1	12.0	11.9	-	-	-
ring substrate, [2]	2.1	11.5	12.0	-	-	-
ring substrate, [5]	1.9	11.1	12.2	-	-	-
parallel stripe, [0]	2.1	12.0	11.9	2.1	10.3	12.3
parallel stripe, [2]	2.0	11.1	11.9	2.0	9.9	12.0
parallel stripe, [5]	2.1	11.0	12.0	2.1	9.1.	12.2
perpend. stripe, [0]	2.1	9.9	12.4	2.0	9.6	12.1
perpend. stripe, [2]	2.1	9.6	12.2	2.1	9.2	12.2
perpend. stripe, [5]	2.1	9.2	12.4	2.2	9.0	12.4

Table 4.5: Extraction results for different substrate trials with variation of the substrate connection separation distance $\Delta d_{DTI,S}$ and topology

The impact on RF characteristics is evaluated by means of the two characteristic cutoff-frequencies $f_{T,peak}$ and $f_{max,peak}$ as shown subsequently. The different configurations are each shown in comparison with the respective variation of the ring substrate as reference (one constant $\Delta d_{DTI,S}$ per figure). The denomination found in the legends is as follows: d for double stripe, s for single stripe, par. for parallel stripes and per. for perpendicular stripes.

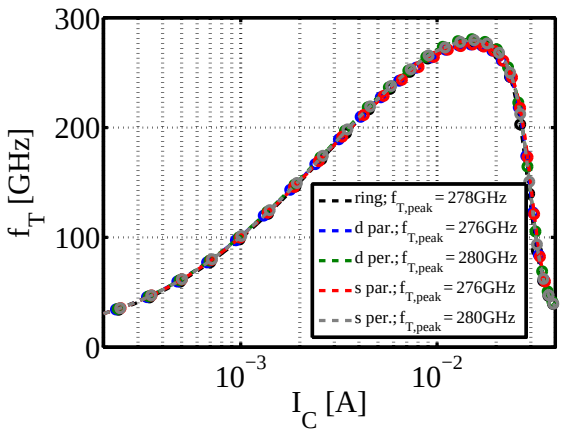


Figure 4.82: Scaling of RF FoM f_T with variation of substrate topology for reference separation ($\Delta d_{DTI,S} = 0\mu m$)

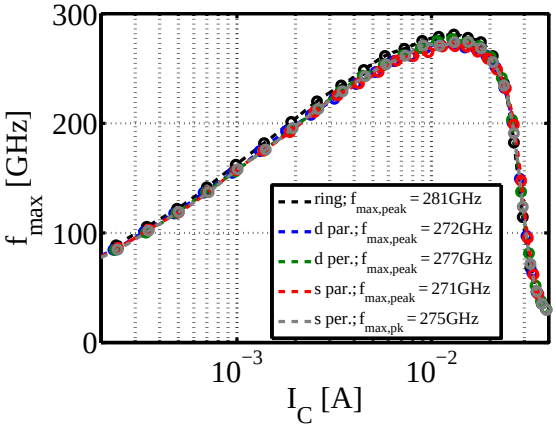


Figure 4.83: Scaling of RF FoM f_{max} with variation of substrate topology for reference separation ($\Delta d_{DTI,S} = 0\mu m$)

Given the precision of the extrapolated peak values of f_T and f_{max} respectively, the variability of RF performance wrt. the reference device is negligible and vanishes within a range that has to be considered as the measurement noise floor (few GHz). Even though the values of the EC elements and especially the substrate capacitance C_{SU} decreased with a less optimal connection topology and separation distance, none of the trials shows a significant variation of the cutoff frequencies. Even for a worst case scenario with a single substrate connector far from the DTI ($5\mu m$) the maximum values of both FoMs f_T and f_{max} decrease by a value of less than $10GHz$ (cf. Fig. 4.84 and 4.85).

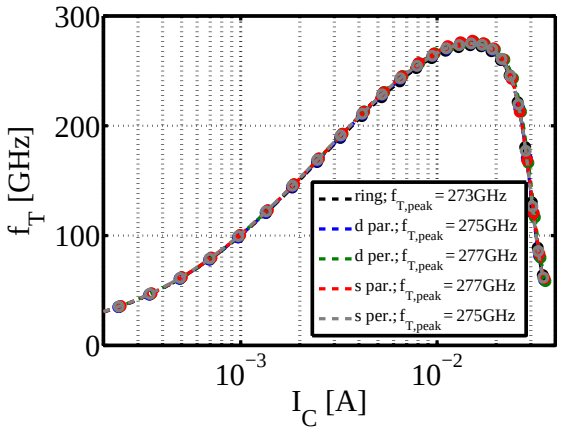


Figure 4.84: Scaling of RF FoM f_T with variation of substrate topology for separation of $\Delta d_{DTI,S} = 5\mu m$

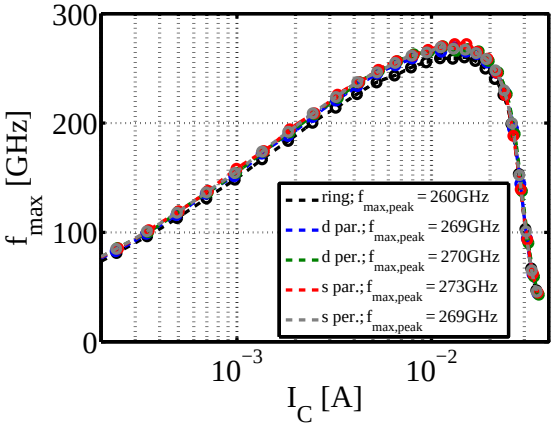


Figure 4.85: Scaling of RF FoM f_{max} with variation of substrate topology for separation $\Delta d_{DTI,S} = 5\mu m$

One can conclude that the DTI structure is very effective in bounding the device and external influences are having almost no impact on the device characteristics.

Deep Trench Isolation Thickness Variation, d_{DTI}

Additionally the impact of a layout change wrt. the DTI itself was evaluated. The default thickness of the DTI defined in the standard process route is fixed at $d_{DTI} = 550nm$ at a default height of the DTI around $h_{DTI} \approx 4\mu m$, giving a large aspect ratio (cf. Fig. 4.81). This study is intended to provide information about changes compared to the default device configuration. A single reference geometry was used ($w_E = 0.2\mu m$ and $l_E = 5\mu m$) and the emitter dimensions were kept constant, whereas the DTI thickness was varied in a range of $400nm$ through $700nm$ (reference plus six additional RF structures). Each device was analyzed with regard to RF characteristics (notably f_T and f_{max}, cf. Fig. 4.86 and 4.87) as well as all substrate-network related EC elements.

In analogy to the variation of the external connection of the substrate terminal, one can clearly see that the RF performance of the measured devices is not influenced by the DTI thickness neither. Figures 4.86 and 4.87 show minimum and maximum DTI width compared to the reference.

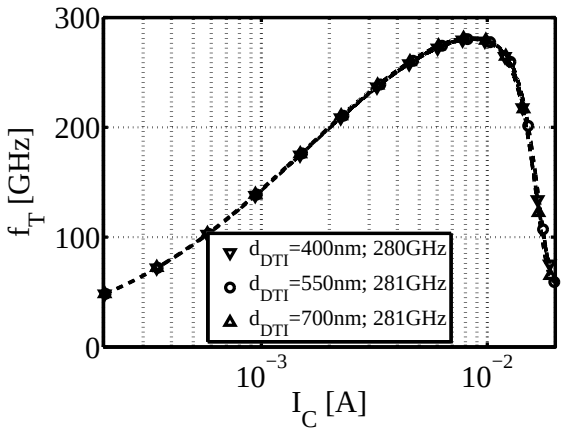


Figure 4.86: Scaling of RF FoM f_T with variation of DTI thickness d_{DTI}

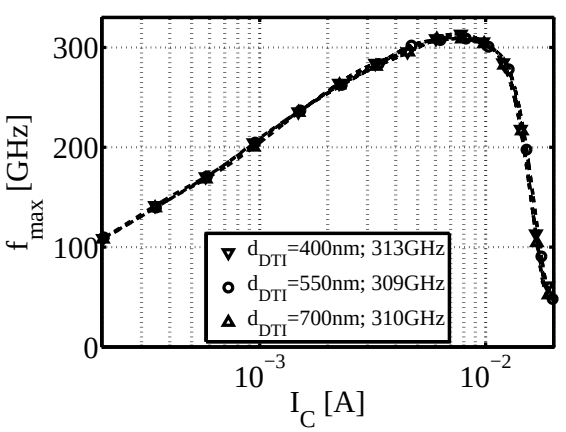


Figure 4.87: Scaling of RF FoM f_{max} with variation of DTI thickness d_{DTI}

As seen in Fig. 4.88 the substrate resistivity increases linearly with the thickness of the DTI. Given the longer current path in the p channel stop pocket (of finite resistivity) at the bottom of the DTI (cf. Fig. 4.66) this behavior is logical and comes as no surprise. Likewise the increasing distance (with increasing DTI thickness) of the separated electrodes (p substrate and n buried layer) to both sides of the DTI causes a lowered capacitance C_{SU}, whereas the CS junction-related bottom area of the transistor remains constant throughout the trials and thus the capacitance C_{CS} does not exhibit a significant variation. However one can clearly see the linear scaling of both substrate-related equivalent network elements as shown in Fig. 4.90 and 4.91

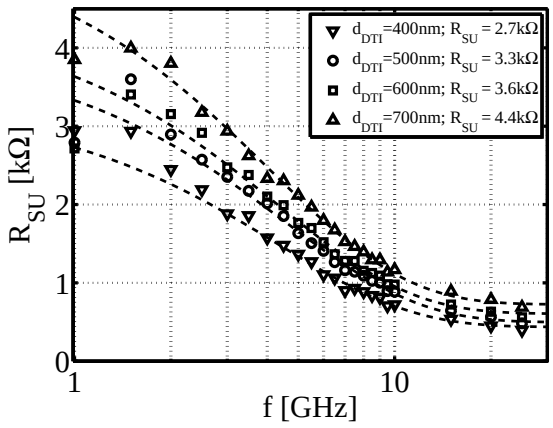


Figure 4.88: Extraction of substrate resistivity R_{SU} from frequency dependence for variation of DTI thickness d_{DTI}

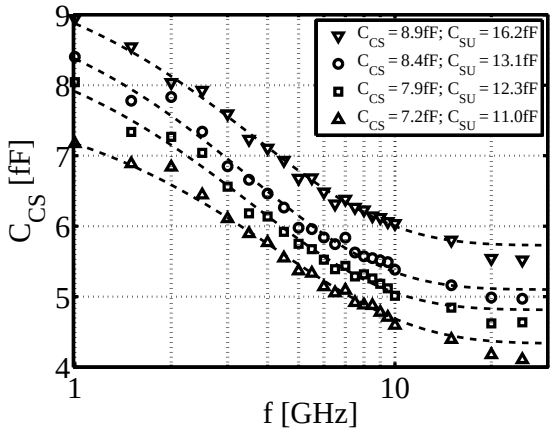


Figure 4.89: Extraction of substrate capacitance contributions C_{SU} and C_{CS} from frequency dependence for variation of DTI thickness d_{DTI}

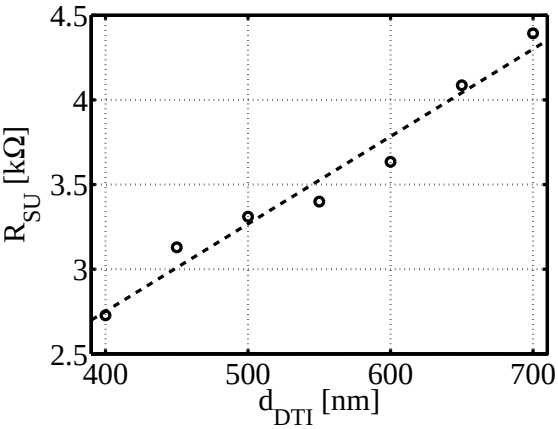


Figure 4.90: Scaling of substrate resistivity R_{SU} with variation of DTI thickness d_{DTI}

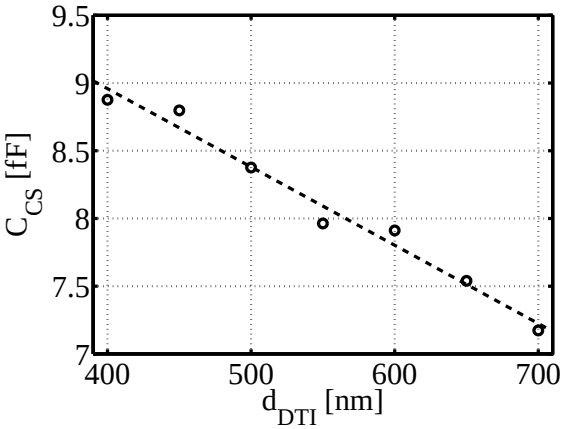


Figure 4.91: Scaling of substrate capacitance C_{SU} with variation of DTI thickness d_{DTI}

Omission of the Trench Isolation

In addition to the thickness variation, a structure without DTI isolation was manufactured (denoted no DTI). As seen in the comparison plots (cf. Fig. 4.92 and 4.93) all characteristic properties of the substrate show a significant variation thus a high dependence on the presence of the substrate isolation through the bounding DTI. The current path to the substrate terminal is shortened, significantly lowering R_{SU}, whereas the increased bottom area gives rise to a significant increase of both the pn-junction related as well as the substrate capacitance (C_{SU} and C_{CS}).

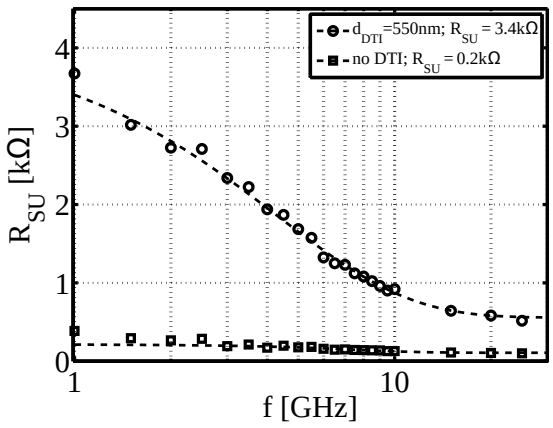


Figure 4.92: Comparison of substrate resistivity R_{SU} for complete omission of the DTI

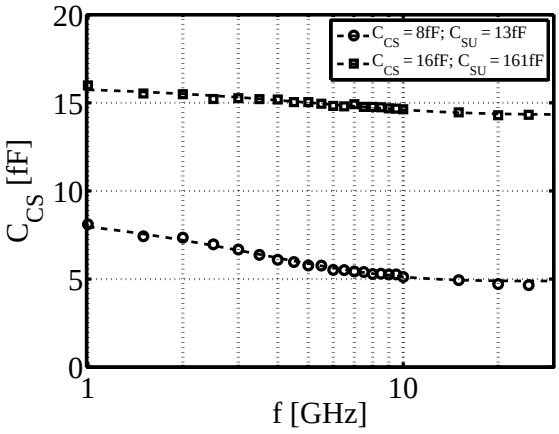


Figure 4.93: Comparison of substrate capacitance for complete omission of the DTI

Despite the large variation of the EC elements in the substrate network, the measured RF characteristics (cf. Fig. 4.94 and 4.95) show that the impact on any of the RF characteristics is vanishing in the measurement precision and thus found to be completely negligible.

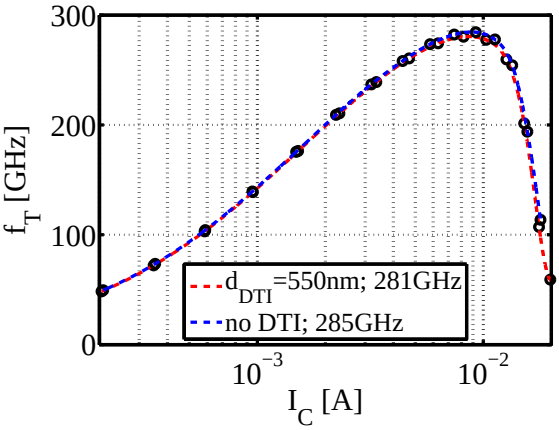


Figure 4.94: Comparison of RF FoM f_T for complete omission of the DTI

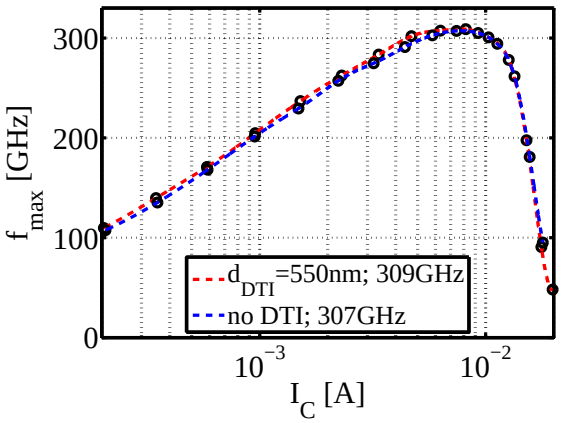


Figure 4.95: Comparison of RF FoM f_T for complete omission of the DTI

Given the presented results, one can conclude that the extrinsic device topology of the substrate shows little to no influence on any of the analyzed device characteristics. However in order to completely conclude the effectiveness of the DTI thickness, a comprehensive analysis of analog RF circuits or logic arrays would be required since the main purpose of the DTI is the intra-device isolation, that cannot be quantified with individual RF structures[1].

[1] with the actual transistor area being largely inferior compared to the total area of the RF measurement periphery

4.6 Low-Current Parameters

After assessment of all external elements and junction capacitances, the modeling of the inner transistor can be performed. In a first step the parameters governing the low-current operation of the device are extracted. The commonly used Gummel plot (measurement network NW1) shows the two important device currents (I_B, I_C) versus V_{BE} bias, under the condition that V_{BC} is held constant. The corresponding plot in logarithmic scale ($\ln(I_C)$ and $\ln(I_B)$) is shown in Fig. 4.96.

For several extraction steps the fundamental diode equation (cf. eqn. (2.27)) is taken as a reference. The normalization of the forward diode current by the saturation current (with exponential dependence on junction bias) provides a helpful method for finding either intermediate variables or precise model parameters values. Therefore rearranging the original equation by division of I_S simply gives the form which is suited to be used with the natural logarithm of both sides removing the exponential to

$$\frac{V_D}{mV_T} = \ln\left(\frac{I}{I_S}+1\right). \tag{4.67}$$

Using an iterative solution approach, a starting value for V_D is guessed and used to fit the rhs. of the equation. Minimizing the correlation error between measurement and model, a value for the saturation current (I_S) is obtained, that is taken to repeat the extraction using a new initial value for V_D. This new value now substituted on the lhs. and so forth. The iteration ideally converges to give the desired values of V_D and I_S.

4.6.1 Base Current I_B

The HICUM formulation of the base current (I_B) is independent of the transfer current (I_T) using individual emission coefficients and saturation currents. The base current is sub-divided in its two components (ideal and non-ideal) namely recombination current (dominant at very low injection) as well as medium to high-current. The complete description in the model uses the equation below

$$I_{BEi} = IBEIS\cdot\left[\exp\left(\frac{V_{BE}}{m_{BEi}\cdot V_T}\right)-1\right] + IREIS\cdot\left[\exp\left(\frac{V_{BE}}{m_{REi}\cdot V_T}\right)-1\right] \tag{4.68}$$

The set of parameters determined in this extraction step are the non-ideality factors M_{BEI} and M_{REI} as well as the saturation currents I_{BEIS} and I_{REIS} giving the respective current densities required for the scalable equation.

The characteristic to be analyzed is the forward Gummel DC measurement data at zero BC bias ($V_{BC} = 0V$). To determine the applicable bias range one can directly look at the I_B characteristic as

well as the first order derivative of I_B to identify non-ideality range (cf. Fig. 4.96 and 4.97). For scalable parameter extraction a reliable bias range is to be defined. Due to maturity related issues in the process[1] the application to data from the BiCMOS55 technology is limited.

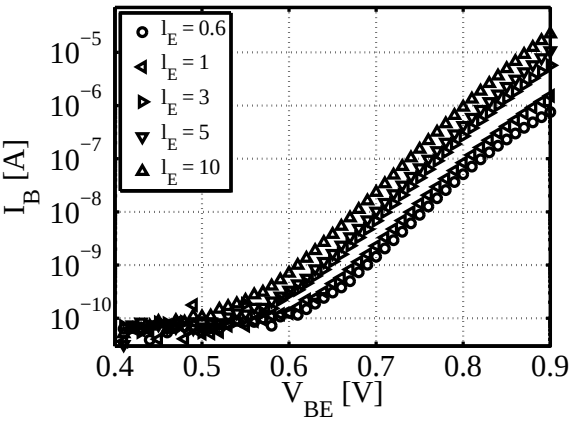


Figure 4.96: Base current I_B in forward Gummel characteristic (V_{BC0}) at fixed emitter width ($w_E = 0.2\mu m$), *drawn emitter length (l_E) in µm*

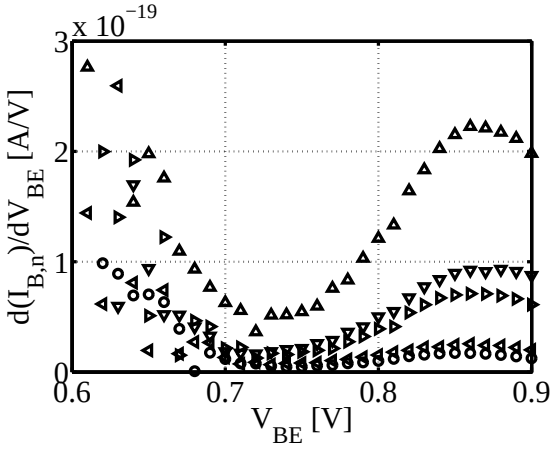


Figure 4.97: First order derivative of I_B to identify non-ideality range at fixed emitter width ($w_E = 0.2\mu m$)

As can be seen in the characteristics, bias values below $V_{BE} = 0.6V$ cannot be taken due to the resolution limit of the measurement instrument, whereas bias values above $0.75V$ are already too far in the high injection regime. Additionally a restriction on the upper bias limit is imposed due to the fact that one assumes the measured current to be identical to the inner device currents ($I_{BE} = I_{BEI}$) and thus the influence of link resistances needs to be reasonably small (R_{Bx} negligible).

$$I_{BEi} = I_B \tag{4.69}$$

$$\frac{I_B}{\exp\left(\frac{V_{BE}}{M_{BEI} \cdot V_T}\right)} = I_{BEIS} + I_{REIS} \cdot \left[\exp\left(\frac{V_{BE}}{V_T}\left(\frac{1}{M_{REI}} - \frac{1}{M_{BEI}}\right)\right)\right] \tag{4.70}$$

In an optimization of the correlation coefficient (r) the two non-ideality parameters (M_{BEI} and M_{REI}) are then determined, so that the curve of the normalized base current ($I_{B,n}$) versus the exponent factor $\exp(V_{BE}/V_T(1/M_{REI} - 1/M_{BEI}))$ becomes a straight line (cf. Fig 4.98). The same procedure is then applied to the extrinsic contribution in order to determine the corresponding non-ideality parameters (M_{BEP} and M_{REP}, cf. Fig. 4.99)

In order to yield reliable separation of intrinsic and peripheral contribution of the base current the extraction is done using a variable emitter width w_E with at fixed emitter length l_E. Typical values for the extracted non-ideality factors are $M \approx 2$ for recombination factors M_R and $M_B \approx 1$ in the ideal range.

[1] base recombination effects are stronger than expected and will be corrected by changes in the process route

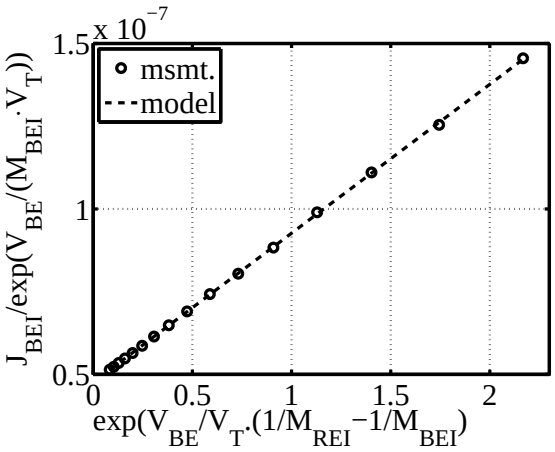


Figure 4.98: Correlation optimization of normalized intrinsic base current density from forward Gummel characteristic (V_{BC0})

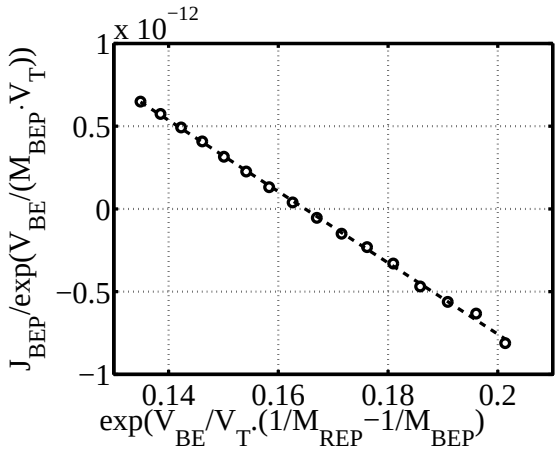


Figure 4.99: Correlation optimization of normalized extrinsic base current density from forward Gummel characteristic (V_{BC0})

In addition the BE tunneling effect is taken into account by a third set of base-current parameters describing I_{BET} [111], yet due to the encountered base-current anomalies this set of parameters was not taken into account for modeling. After manually re-adjusting the set of parameter the verification of scalability and the general fit of the model is performed (cf. Fig. 4.101).

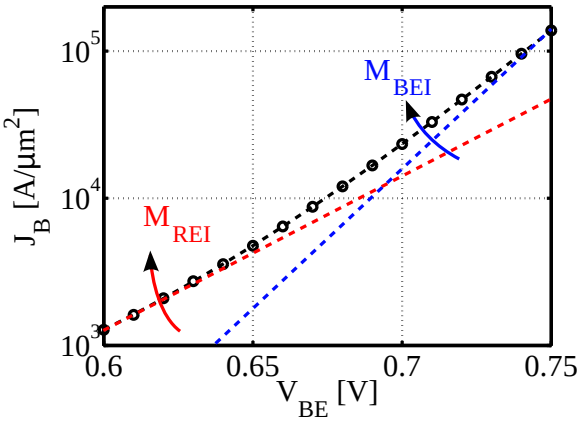


Figure 4.100: Verification of extracted non-ideality factors on current density plot;
[*parameter optimization wrt. to slope for recombination (M_{REI}) and low injection (M_{BEI})*]

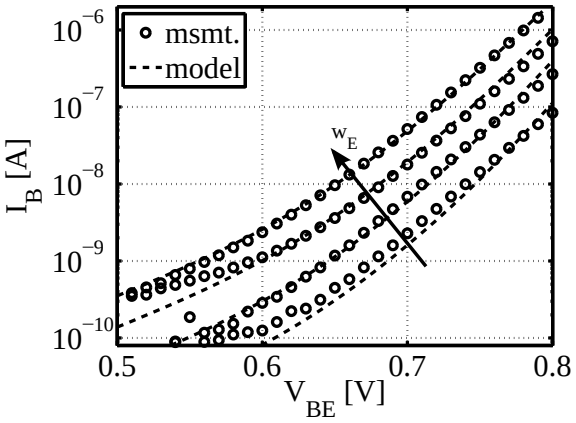


Figure 4.101: Verification using different geometries in forward Gummel characteristic (V_{BC0});
[*model (dashed line) vs. simulation (points)*]

4.6.2 Avalanche Breakdown

For conscious circuit design the breakdown voltage needs to be accurately predictable. The avalanche breakdown in the BC junction is taken into account by the compact model through the increase in current for voltages higher than BV_{BC}. It is generally reflected by an additional factor in the current expression. The simplified BC breakdown model in HICUM takes into account weak avalanche under sufficiently low current density operation [111].

The empirical relationship used to describe avalanche breakdown employs the so called multiplication factor M, which is defined as the ratio of the current density leaving the zone of impact ionization to the current density entering that region. An empirical relation between M and the desired avalanche current is given by $I_C = J_n A_E = M \cdot I_T$ with $I_C = I_{AVL} + I_T$. Thus the avalanche current is defined $I_{AVL} = (M-1) \cdot I_T$

The mechanism of avalanche breakdown in BJTs depends on the circuit configuration (common-emitter or common-base). Compared to common base configuration, the breakdown voltage is generally lower for common emitter mode due to the amplification effect within the transistor. The effect of avalanche breakdown can therefore be observed in DC measurements of RF devices in both collector and base current with the latter being the much more sensitive quantity. For parameter extraction the avalanche current I_{AVL} is obtained from the base current I_B as a function of V_{BC} (msmt. data of forward Early characteristic, NW5). A reference value is taken from the base current at low CB voltage (e.g. $V_{CB} = 0$, cf. Fig. 4.102)

$$I_{AVL} = I_B(V_{BC} = 0) - I_B(V_{BC}). \tag{4.71}$$

From measured data the multiplication factor M is obtained from:

$$M = \frac{I_C}{I_C - \Delta I_B} = 1 + \frac{I_{AVL}}{I_C - I_{AVL}}. \tag{4.72}$$

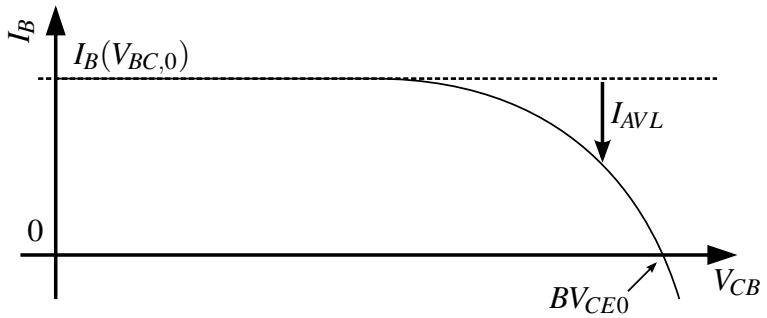


Figure 4.102: Base current variation under of impact ionization in the BC SCR caused by high negative BC bias

Towards high CE bias, strong electric fields cause impact ionization in the BC SCR. In the respective curve (cf. Fig. 4.102) the voltage where I_B crosses abscissa gives the characteristic CE breakdown voltage BV_{CEO} from the addition of the respective junction biases

$$BV_{CEO} = V_{BE} + V_{CB}(I_B = 0). \tag{4.73}$$

The collector-base weak avalanche current, is represented by a current in parallel with the internal base-collector junction as shown in the equivalent circuit (cf. Fig. 4.103). It is modeled using the forward transfer current I_{TF} as well as the electric field E, which is represented through the BC depletion capacitance C_{jCi} resulting in the model equation

$$I_{AVL} = F_{AVL} \cdot I_{TF} \cdot (V_{DCi} - V_{BiCi}) \cdot \exp\left[-\frac{Q_{AVL}}{C_{jCi}(V_{DCi} - V_{BiCi})}\right]. \tag{4.74}$$

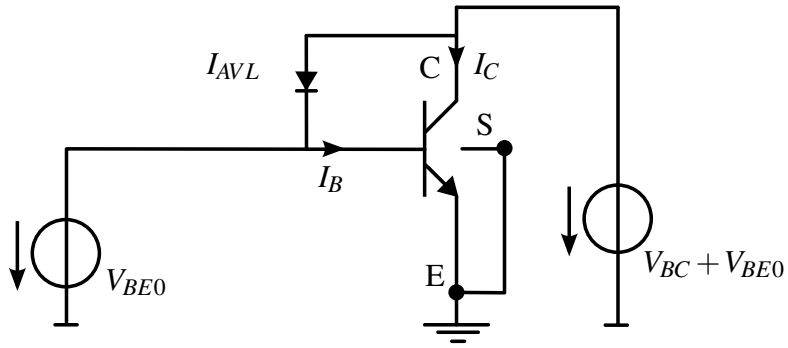


Figure 4.103: npn BJT equivalent circuit of the weak avalanche effect

Unlike the transfer current the main portion of the avalanche current is generated in the internal transistor. The ratio of perimeter to area for the avalanche current is hence different from the scaling of the transfer current so that $A_{AVL} \neq A_{E,eff}$ calling for a modified ratio of perimeter to area avalanche current γ_{AVL}[1].

Thus in a first step the respective contributions of perimeter and bottom area are separated using a multitude of available geometries (cf. Fig. 4.104). Since at high current densities the electric field changes, the area contribution decreases resulting in a bias dependent ratio γ_{AVL}. Selecting an average value over a given range is thus mandatory (cf. Fig. 4.105).

[1] depending on internal and external collector doping, the current ratios are roughly the same as for the transfer current

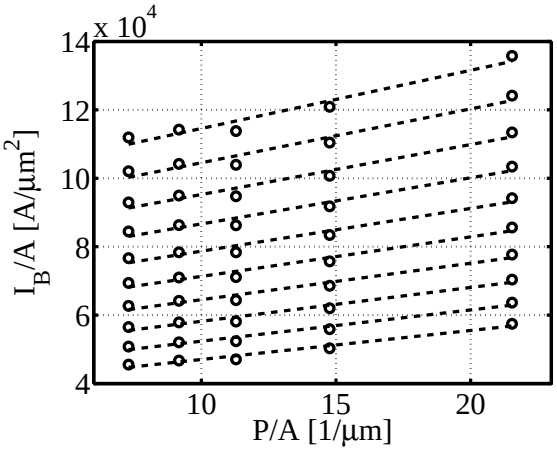


Figure 4.104: Separation of base current contributions for available device geometries (P/A separation) at different BC bias conditions

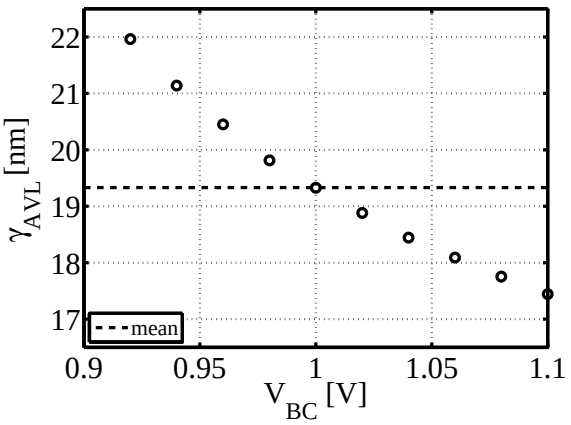


Figure 4.105: Avalanche current related ratio of perimeter and area (γ_{AVL}) as a function of device bias

As seen in the model equation, both HICUM parameters F_{AVL} (avalanche current factor) and Q_{AVL} (exponent factor) are very sensitive to the accuracy of extraction results for the internal BC depletion capacitance parameters. Using the multiplication factor M as deviation from ideality and replacing the transfer current by the measured collector current $I_{TF} = I_{Co}$ yields

$$I_C = M \cdot I_{Co} = I_{Co} + I_{AVL} \text{ with } M = 1 + \frac{I_{AVL}}{I_{Co}}. \tag{4.75}$$

Identical transformation of model equation and measured data gives the form

$$1 - M = \frac{I_{AVL}}{I_{Co}} = F_{AVL} \cdot (V_{DCi} - V_{BiCi}) \cdot \exp\left[-\frac{Q_{AVL}}{C_{jCi}(V_{DCi} - V_{BiCi})}\right], \tag{4.76}$$

which is suitable for a direct parameter extraction from experimental data in the form of

$$\ln\left[\frac{M-1}{V_j}\right] = \ln(F_{AVL}) - \frac{Q_{AVL}}{C_{jCi0} \cdot V_{DCi}^{Z_{Ci}}} \cdot V_j^{Z_{Ci}-1}. \tag{4.77}$$

From a plot of $\ln\left[(M-1)/V_j\right]$ versus $V_j^{Z_{Ci}-1}$ one obtains hence Q_{AVL} from the slope of a linear regression whereas F_{AVL} is deducted from the ordinate intercept (cf. Fig. 4.106).

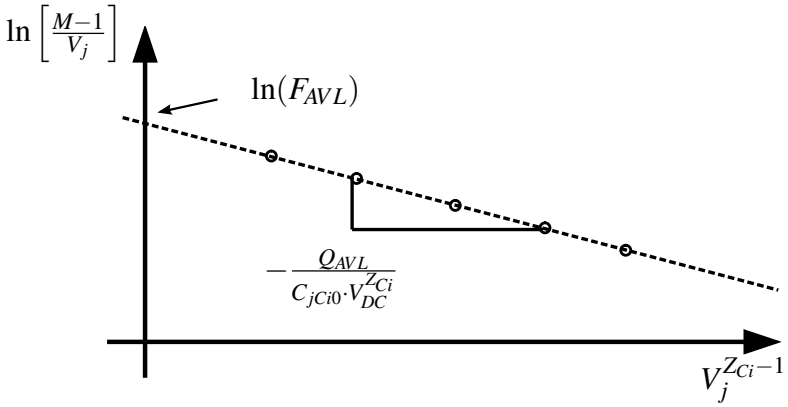


Figure 4.106: Direct avalanche parameter extraction from linear regression

Since avalanche parameters are usually determined before assessment of high injection parameters and self-heating effects a sufficiently low bias is recommended for parameter extraction. For the extraction of the geometry-scalable avalanche parameter set (F_{AVLU} and Q_{AVLU}) the forward Early measurement (NW5) is used. The bias range is limited to operation at medium BE junction bias (e.g. V_{BE} = 0.7V, V_{CB} = [0V .. 1.1V]) in order to eliminate self-heating effects. Furthermore the bias range is capped by definition of a minimum bias $V_{CB,min}$, defined as point where strong avalanche effects become visible (cf. Fig. 4.109).

By means of normalized measured data from different geometries one can then isolate the multiplication factor M (cf. Fig. 4.107). Using the BC junction related model parameters for the built-in potential V_{DC} and exponent factor Z_C (cf. section 4.5.4) the ionization rate parameters are obtained from slope and intercept of the linear regression analysis (cf. Fig. 4.108).

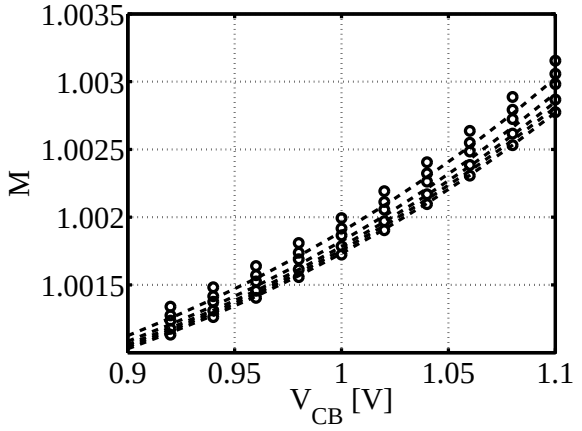


Figure 4.107: Extraction of the avalanche multiplication factor M for a selected BC bias range; *different geometries at* $w_E = 0.2\mu m$

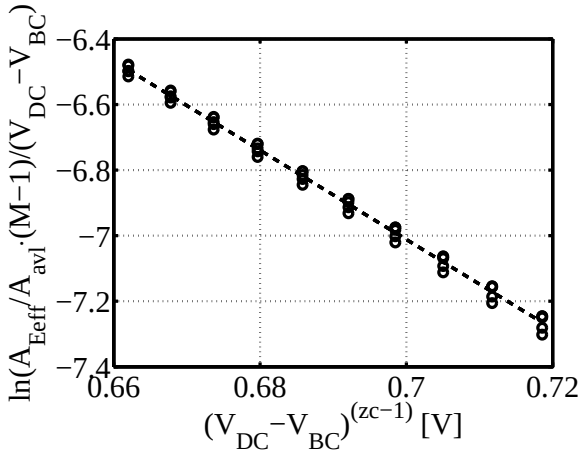


Figure 4.108: Direct avalanche model parameter extraction from linear regression using normalized currents of multiple geometries

After adjustment of γ_{AVL} the normalized base current $I_{B,N}$ remains largely uninfluenced by the

variation of device geometry. Thus for re-adjustment and optimization of the two parameters, a selected number of reference geometries is found sufficient. A circuit simulation with non-normalized base current confirms the scalability of the model (cf. Fig 4.110).

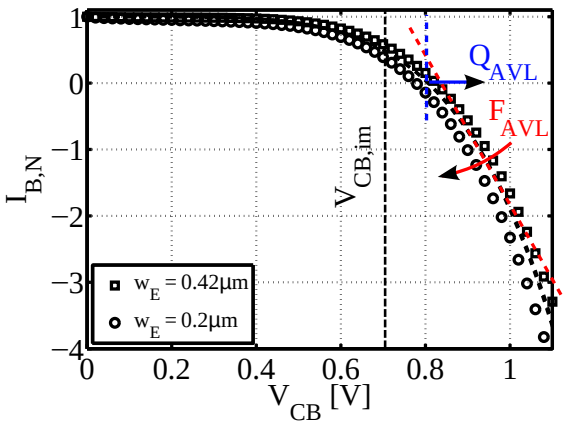


Figure 4.109: Deviation of the normalized base current (I_B) from its reference value ($I_B(V_{BC0})$) towards high electrical fields

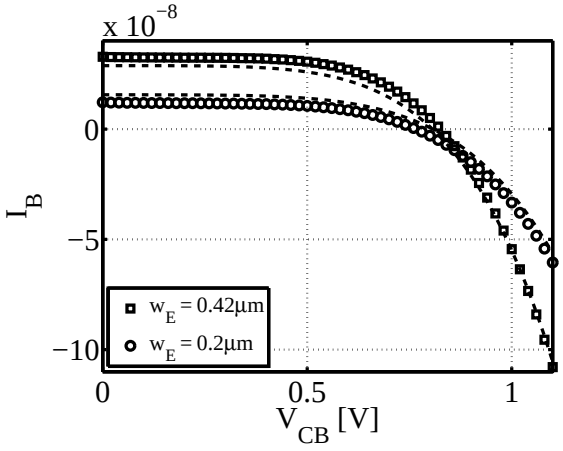


Figure 4.110: Verification of parameter and model scalability for two extreme geometries ($w_{E,min}$ and $w_{E,max}$)

4.6.3 Collector Current I_C

The main current in the bipolar transistor is described by the integral charge control relation (ICCR). A weighted integral of the majority carrier concentration in the base is the foundation of the charge control principle, describing the relation between charges and currents in a bipolar transistor. Its dependence on the main current makes it an implicit equation (rather than a explicit expression) yet allows the HICUM model to cover modern hetero-junction transistors realized in SiGe technologies. The issue of non-homogeneous spatial structures and material compositions (B, E, and C can differ) is taken into account by a representation of each region by separate model components.

Ratio of Periphery to Area Specific Collector Current (γ_C)

In order to have good process scalability the ratio of periphery and area specific collector current contribution is determined in a first step of the DC current extraction flow, related to the collector or transfer current respectively [68]. The fundamental concept is presented in Fig. 3.7 of section 3.1.2. From the forward characteristic one separates the perimeter current per unit circumference (I_{CP}), which is then related to the internal current density per unit area ($I_{CA} = J_C$). As shown in Fig. 4.111, the ratio of $I_{T,p}$ to the total current I can be easily calculated from device measurement of the collector current in forward operation. The bias range is bounded to $V_{BE} = [0.5..0.72V]$ in order to get robust results without the influence of high injection effects. Given the favourable aspect ratio, a width scaling at constant l_E is preferred. The value of γ_C is then obtained through averaging in the selected bias range and can be verified using all available geometries (cf. Fig. 4.112).

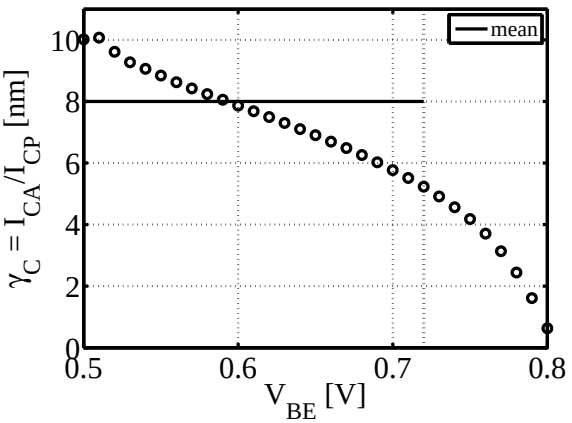


Figure 4.111: Averaged extracted ratio of periphery- to area-specific collector current (I_C) from fwd. Gummel msmt. at V_{BC0} as function of V_{BE} bias for devices with $l_E = 5\mu m$

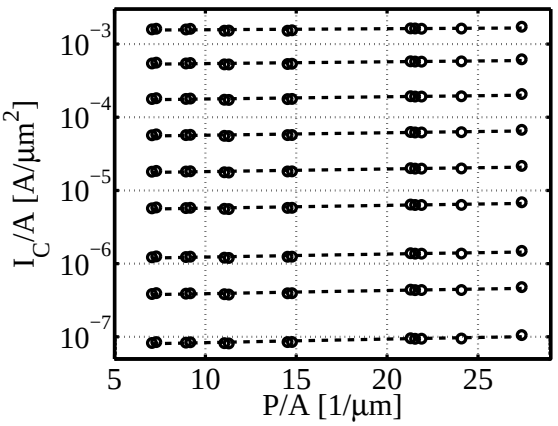


Figure 4.112: Verification plot of normalized collector current (I_C) for all available geometries with CBEBC configuration for selected V_{BE} bias points

Extraction Strategy for the h_{jEi} Complex

Apparent differences in material properties of neighboring regions (e.g. intrinsic carrier concentration n_i and carrier mobility μ) are represented by compact model parameters (weighting factors) associated with the different junctions. As presented in section 2.5 a significant effort was made to adapt the model formulation for recent high-speed SiGe technologies with aggressively scaled vertical SiGe base profiles. However given the increased flexibility, the model formulation for the BE junction related model parameter h_{jEi} became more difficult (as compared to HICUM L2.2x) and the number of model parameters was increased, causing the approach of direct solution through parametric nonlinear transformation to become obsolete.

There are different extraction strategies for the emitter related space charge weighting factor parameters of the new model: a direct parameter extraction method was presented [112] using a transformation by means of the Lambert W function [113], yet iterative approaches exist [10, 91, 114].

All methods have a few basic assumptions in common. The extraction is performed at low forward bias range of the BE junction (low to medium current densities), so that neither high injection effects nor series resistances dominate the electrical behavior of the transistor. This low current precondition allows for the following simplifications:

- the internal bias at the BE junction is equivalent to the terminal bias [$V_{BEi} \approx V_{BE}$]
- the transfer current i_T equals the measured collector current I_C [$i_T \approx I_C$]

At zero BC bias ($V_{BC} = 0$) the influence of Q_{jCi} is eliminated. Now Q_{jEi} is the largest charge contribution at lower current densities and only the bias dependence of h_{jEi} is relevant [115]. The total GICCR approach can then be simplified as shown below

$$J_T = \frac{I_T}{A_E} = qV_T \frac{\mu_{n0} n_{i0}^2}{\int_{x1}^{x2} h(x)p(x)dx} \frac{1}{n_i^2} \exp\left(\frac{V_{BE}}{V_T}\right) \approx \frac{c_{10}}{Q_{pT}} \exp\left(\frac{V_{BE}}{V_T}\right). \tag{4.78}$$

The classical approach using a regression analysis for determination of the (SGP) parameters I_S, V_{ER} and I_{KF}, uses the transfer current expression with the normalized collector current $I_{C,n}$

$$I_{C,n} = \frac{I_C}{I_S} \cdot \exp\left(\frac{-V_{BE}}{V_T}\right) \tag{4.79}$$

Under low injection condition the GICCR constant c_{10} as well as the hole charge Q_{p0} were estimated using simple DC transfer current measurement at $V_{BC,0}$, so the equation could be simplified (m_{BEi} = 1) to the form of

$$I_C = \frac{c_{10}}{Q_{p0} + Q_{jEi}} \exp\left[\frac{V_{BE}}{V_T}\right], \tag{4.80}$$

where under idealized assumptions the total hole charge Q_{pt} is reduced to the contributions of the charge of the inner BE SCR (Q_{jEi}) as well as the model parameter for the zero-bias hole charge (Q_{p0}). With Q_{jEi} fixed through the previously extracted junction capacitance only two essential model parameters are left to be determined. The quotient c_{10}/Q_{pT} can be expressed in a form of normalized collector current (I_{Cn}, which additionally takes the saturation current I_S into account), that is convenient for graphical interpretation of the measured DC data

$$\frac{c_{10}}{Q_{pT}} = \frac{c_{10}}{Q_{p0}+Q_{jEi}} \text{ and} \tag{4.81}$$

$$\left.\frac{c_{10}}{Q_{pT}}\right|_{meas} = \frac{I_C}{\exp\left[\frac{V_{BE}}{V_T}\right]} = I_S \cdot I_{C,n} \tag{4.82}$$

The normalized collector current (I_{Cn}) is an important measure wrt. the BE junction. For real devices the measured collector current is different from the ideal transistor characteristic described by the exponential dependence of the transfer current in the forward Gummel plot as shown in Fig. 4.114. This effect is due to the BE charge (Q_{jEi}) and an indicator for the reverse Early effect. The collector current normalized by the ideal diode equation thus is always less than 1 (cf. Fig. 4.113) gives a deviation from the ideal BE bias dependence (cf. Fig. 4.114).

For recent SiGe based HBT technologies this deviation however was found to be significant, and the negative slope of the normalized collector current was not satisfactory reproduced by a constant weighting factor h_{jEi} anymore.

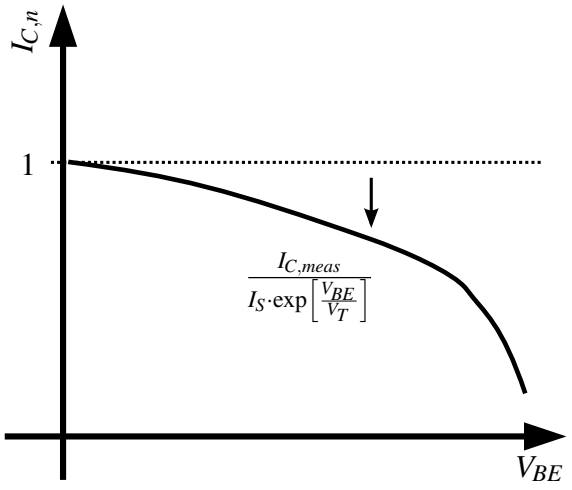


Figure 4.113: Normalized collector current (I_{Cn}) with deviation from ideal (exponential) diode relation (I_S determined from regression)

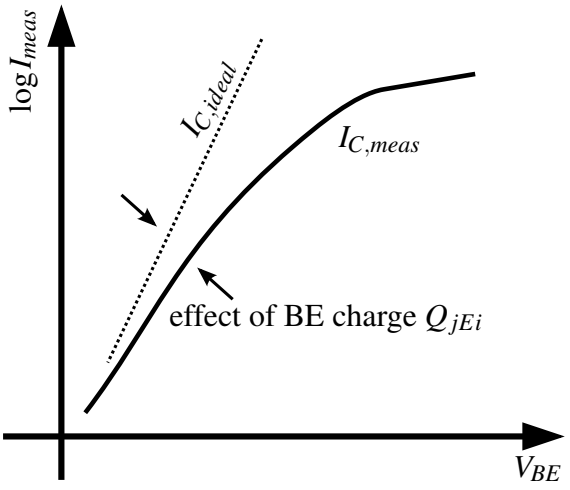


Figure 4.114: Deviation of measured collector current (I_C) from ideal exponential characteristic due to additional BE charges

The BE charge in the model is defined through the depletion capacitance C_{jEi} (whose parameter set is determined in the first extraction steps of the flow) as well as the position-dependent

weighting factor m

$$Q_E = \int_0^{V_{BEi}} m(x_e(u)) \cdot C_{jEi}(u) du, \text{ with} \tag{4.83}$$

$$m(x,T) \propto \frac{\exp\left[\frac{T_0 V_G(x)}{T\ V_{T0}}\right]}{\mu(x,T_0)} \tag{4.84}$$

In eqn. (4.83) $x_e(u)$ represents the bias dependent BE SCR boundary within the base region. By factoring out the charge Q_{jEi}, the emitter related weighting factor is obtained

$$h_{jEi} = m(x_e(V_{BEi})) = \frac{\int_0^{V_{EBi}} m(x_e(u)) \cdot Q_{jEi}(u) du}{Q_{jEi}(V_{BEi})}. \tag{4.85}$$

Even though the new model implementation brought a higher degree of freedom for the modeling of the reverse Early effect it added ambiguity to the model extraction flow. The first methods to determine the weighting factor parameter set (cf. [10, 91]) were indirect extractions based on parameter optimization-loops with a high sensitivity to selected bias ranges.

The complex of parameters introduced in HICUM L2.3x with the newly introduced formulation for the weighting factor reads

$$h_{jEi} = h_{jEi0}(T_0) \frac{\exp\left(a_{hjEi}\left(1-\left(1-\frac{V_{BEi}}{V_{DEi}}\right)^{z_{Ei}}\right)\right)-1}{a_{hjEi}\left(1-\left(1-\frac{V_{BEi}}{V_{DEi}}\right)^{z_{Ei}}\right)}. \tag{4.86}$$

As formulated in the release manual [116] the factor a_{hjEi} of the model equation (4.86) related to the BE weighting factor is substituted by u, which represents a modified internal base-emitter voltage (smoothing function respectively), that is dependent of a new model parameter r_{hjei} [50]. This voltage is different from the junction voltage v_j

$$u = vj_{zz} = a_{hjEi}\left(1-\left(1-\frac{v_j}{V_{DEi}}\right)^{z_{Ei}}\right) \text{ , with } v_j = f(r_{hjei}) \approx V_{BEi}. \tag{4.87}$$

Substituting the h_{jEi} complex yields

$$h_{jEi} = h_{jEi0}(T) \cdot w(u) \text{ with } w(u) = \frac{\exp(u)-1}{u}. \tag{4.88}$$

However a function in the form of $x \cdot \exp(x) = a$ is injective, thus has no direct inverse making it difficult to find a direct parameter extraction strategy. For the special case used in the model (cf. eqn. (4.86)) the Lambert W function [113], a twofold set of branches of the inverse relation of the function $z = f(W) = W\exp(W)$, is suited to solve the respective expression. Using a numerical

analysis of the Lambert W function $x = W(a)$ [117], one can obtain the inverse of the lower of the two real function branches (W_{-1}). Yet therefore some identical transformations are required.

By definition the bias dependence of the weighting factor h_{jEi} uses a redefined junction voltage v_j as well as an additional temperature dependent model parameter $a_{hjei}(T)$

$$u = a_{hjei}(T) \cdot f(v_j). \tag{4.89}$$

In order to solve the model formulation, the multivalued inverse $x = W(w)$ is defined with

$$w = x \cdot \exp(x) \quad \text{and} \quad \exp(u) = w \cdot \left(u + \frac{1}{w}\right). \tag{4.90}$$

After further identical manipulations, one obtains the suitable form

$$-\frac{1}{w}\exp\left(-\frac{1}{w}\right) = -\left(u + \frac{1}{w}\right) \cdot \exp\left[-\left(u + \frac{1}{w}\right)\right]. \tag{4.91}$$

Given this representation, the inverse W_{-1} is giving access to the desired parameter u as function of bias

$$u = -\frac{1}{w} - W_{-1}\left[-\frac{1}{w}\exp\left(-\frac{1}{w}\right)\right]. \tag{4.92}$$

At each temperature T the model parameter $a_{hjei}(T)$ can now be calculated with identical transformation of the model equation (cf. Fig. 4.115 and 4.116)

$$a_{hjei}(T) = \frac{u}{1 - \left[1 - \frac{v_j}{v_{dei}(T)}\right]^{z_{ei}}}. \tag{4.93}$$

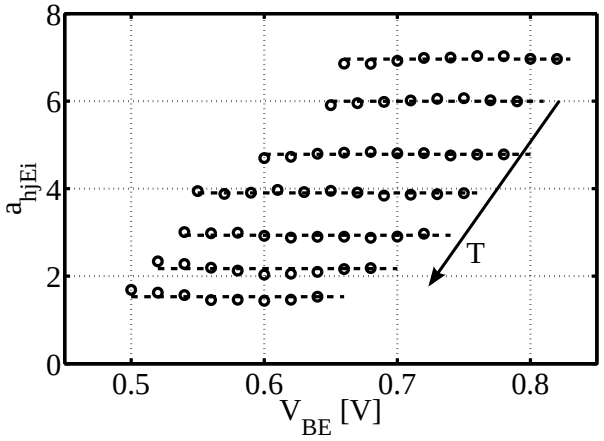


Figure 4.115: Extracted weighting factor $a_{hjei}(T)$ as a function of V_{BE} bias at different ambient temperatures T

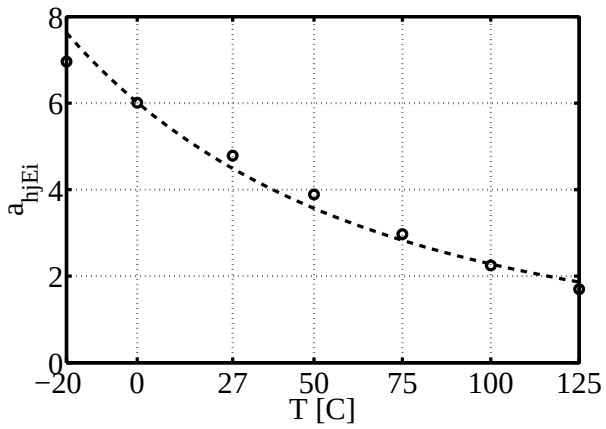


Figure 4.116: Averaged value as function of temperature for extraction of temperature coefficient; *dashed line: model equation*

The direct extraction flow is based on a simplified model representation with SGP originated parameters for saturation (I_S) and knee current (I_{QF}) respectively. The inverse scaled collector current (I_{T*}) can then be expressed from in a very general form using Early parameters P for the respective junction together with the related charges q_j

$$\frac{1}{I_{T*}} \approx \frac{1}{I_S} + \frac{P_E \cdot q_{jE}(V_{BE}) + P_C \cdot q_{jC}(V_{BC})}{I_S} + \frac{1}{I_S} \cdot \frac{I_T}{I_{QF}} \tag{4.94}$$

With the measurement data selected for zero BC bias (V_{BC0}) the terms related to the BC junction can be neglected. Furthermore inserting the HICUM representations as

$$q_{jE} = \frac{\exp(u)-1}{u}\, Q_{jEi}(V_{BE}) \;\text{ and }\; P_E = h_{jEi0}(T), \tag{4.95}$$

yields the formulation suitable for extraction by a multi-variable linear regression based on measured collector current data

$$\frac{1}{I_{T*}} \approx \frac{\Theta_{qp0}}{I_S} + \frac{h_{jein}(T)\,\Phi_{ET}}{I_S} \cdot \frac{\exp(u)-1}{u} + \frac{I_T}{I_S I_{QF}}\Theta_{t0}, \text{ with} \tag{4.96}$$

$$h_{jEin} = h_{jEi0}(T) \cdot C_{jEi0}(T_0) \cdot \frac{V_{DEi}}{Q_{p0}}. \tag{4.97}$$

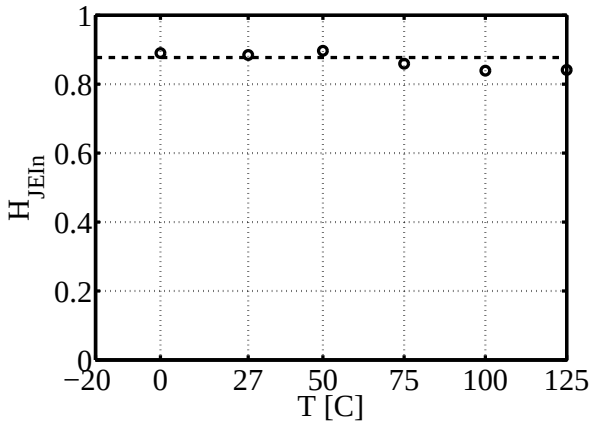


Figure 4.117: Averaging of normalized weighting factor h_{jEin} over temperature ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

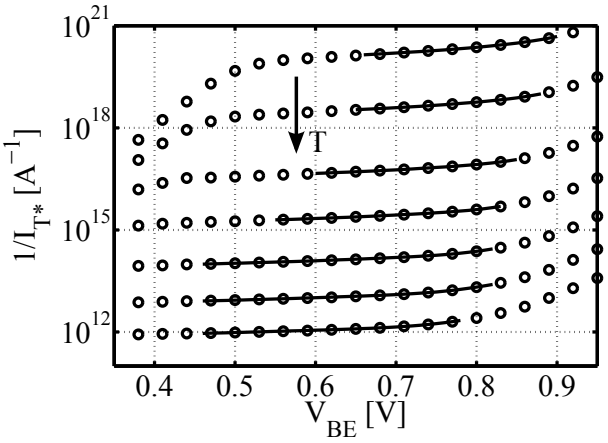


Figure 4.118: Verification with scaled inverse transfer current I_{T*} at various temperatures, *solid line: model within selected bias range*

Given the averaged weighting factor one can then recalculate the value of the multivalued inverse of the Lambert function from

$$w = \left(\frac{I_S}{I_{T*}} - \Theta_{qp0} - \frac{I_T \Theta_{t0}}{I_{QF}} \right) \frac{1}{h_{jEin} \cdot \Phi_{ET}}. \tag{4.98}$$

In order to take into account the apparent temperature dependence for extraction at temperatures different from the nominal temperature T_0 the factor Θ represents the temperature scaling function of the respective parameter in subscript as implemented in the model.

The iteratively optimized auxiliary variables (of the SGP model) can be used with their respective representation through charges and weighting factors in the HICUM model

$$I_{QF} = \frac{Q_{p0}(T_0)}{t_0(T_0)\, h_{f0}(T)} \text{ and } I_S = \frac{c_{10}(T)}{Q_{p0}(T_0)}. \tag{4.99}$$

Using these auxiliary variables, obtained at different temperatures during extraction, one can further extract the temperature parameters of the respective model equation. The saturation current (I_S) is known to be directly linked to the GICCR weighting factor c_{10}, whereas the knee current (I_{QF}) is assumed to obey the same temperature scaling rule as the newly introduced temperature dependence of the weighting factor h_{f0}.

$$I_S(T) \propto c_{10}(T) = c_{10}(T_0) \left(\frac{T}{T_0} \right)^{\zeta_{CT}} \exp \left[\frac{\Delta V_{GBE}(T=0)}{V_{T0}} \left(1 - \frac{T_0}{T} \right) \right] \tag{4.100}$$

$$I_{QF}(T) \propto h_{f0}(T), \text{ with } I_{QF}(T) \exp \left[\frac{-\Delta V_{GBE}}{V_{T0}} \left(1 - \frac{T_0}{T} \right) \right] \tag{4.101}$$

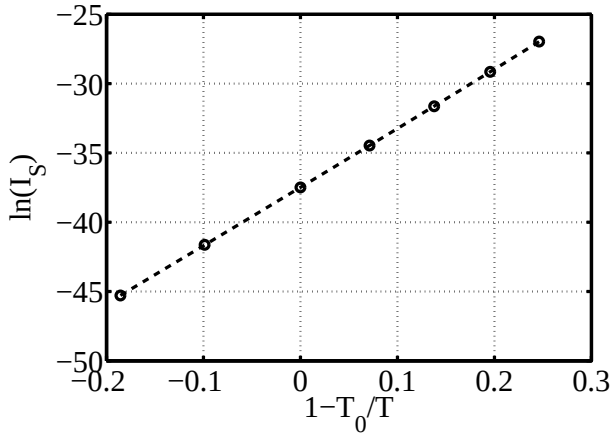


Figure 4.119: Extraction of temperature scaling parameters using logarithmic extracted saturation current (I_S) for single geometry

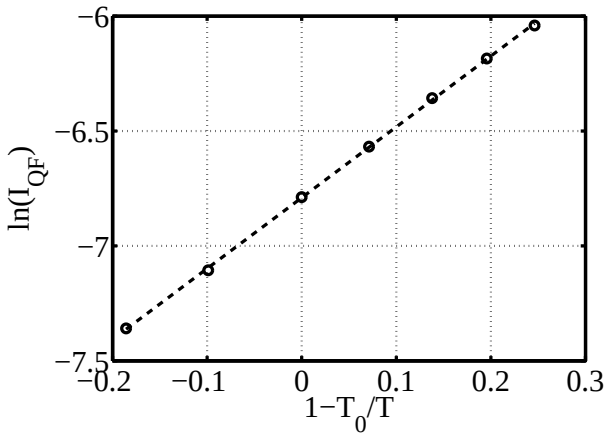


Figure 4.120: Extraction of temperature scaling parameters using logarithmic extracted knee current (I_{QF}) for single geometry

The immediate verification of the extraction results can be made with a plot of the normalized collector current versus bias and temperature as shown in Fig. 4.121 and confirms the suitability of the new model formulation as seen in Fig. 4.122.

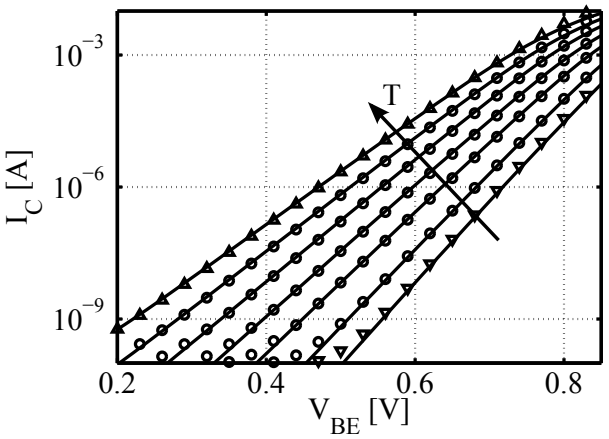


Figure 4.121: Model verification of temperature scaling using the collector current (I_C) in fwd. Gummel characteristic;
single geometry, $w_E = 0.2\mu m$ and $l_E = 5.0\mu m$

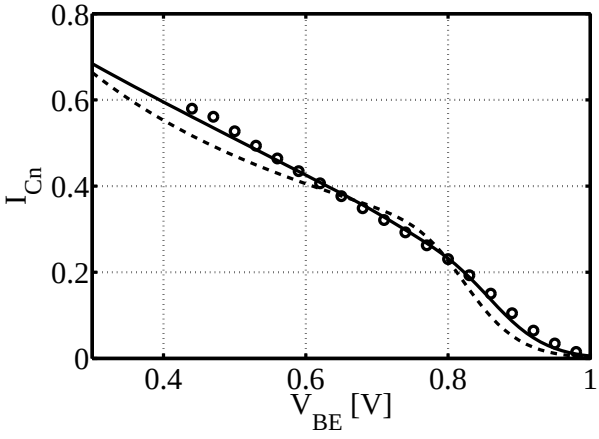


Figure 4.122: Model verification versus bias of the BE junction using the normalized collector current ($I_{C,n}$) at T_0; *dashed: HICUM L2.2x, solid: HICUM L2.3x*

The Indirect Extraction Approach for the h_{jEi} Complex

The direct extraction approach is well suited for highly accurate and physics-related parameter determination due to the use of a transformation of the model equation. However the method is sensitive to proper selection of bias ranges and thus less practical for comprehensive geometry scalable parameter extraction.

The former transfer current extraction as proposed in [89], is based on the same simplifications of the original ICCR. The transfer current equation is first simplified for the zero BC bias case, where the depletion charge is canceled out ($V_{BC} = 0 \rightarrow Q_{jCi} = 0$). These circumstances make the exponential term for the BE junction bias dominant compared to the BC term allowing to completely neglected its influence

$$I_C \cong i_T = \frac{C_{10}\left[\exp\left(\frac{V_{BEi}}{V_T}\right) - 1\right]}{Q_{P0} + h_{jEi}Q_{jEi}} \approx \frac{C_{10}\left[\exp\left(\frac{V_{BEi}}{V_T}\right)\right]}{Q_{P0} + h_{jEi}Q_{jEi}} \tag{4.102}$$

$$Q_{jEi} = \frac{C_{10}}{h_{jEi}}\frac{\exp\left(\frac{V_{BEi}}{V_T}\right)}{I_C} - \frac{Q_{P0}}{h_{jEi}} \tag{4.103}$$

With the BE junction capacitance C_{jE} determined in a previous extraction step the charge Q_{jEi} at a given bias is known and hence the parameter extraction might be performed. The parameters, that are determined in this extraction, are scaling with geometry since only the ratio of GICCR factor c_{10} and the zero bias hole charge Q_{P0} is geometry-independent. However only the ratio C_{10}/h_{jEi} as well as Q_{P0}/h_{jEi} might be determined whereas the individual model parameters C_{10} and Q_{P0} remain unknown until the extraction of AC parameters (cf. sect. 4.9).

To overcome the limitations of the extraction of the ratios C_{10}* and Q_{P0}* a new method was proposed recently. The method is again based on the same preconditions as outlined above, yet the partitioning of the current contributions is done differently.

$$I_C \cong i_T = \frac{C_{10}\left(\exp\left(\frac{V_{BE}}{V_T}\right)\right)}{Q_{P0} + h_{jEi}Q_{jEi}} \quad \text{, with normalization through } \; Q_{p0} \tag{4.104}$$

$$I_C = \frac{C_{10}\cdot\exp\left(\frac{V_{BE}}{V_T}\right)}{Q_{P0} + \frac{h_{jEi}}{Q_{P0}}Q_{P0}Q_{jEi} + \frac{h_{jCi}}{Q_{P0}}Q_{P0}Q_{jCi}} = \frac{I_S\cdot\exp\left(\frac{V_{BE}}{V_T}\right)}{1 + h_{jEi0}Q_{jEi} + h_{jCi0}Q_{jCi}} \tag{4.105}$$

Here h_{jEi0} and h_{jCi0} are two temporary weighting parameters.

$$h_{jEi0} = \frac{h_{jEi}}{Q_{P0}} \;\text{ and }\; h_{jCi0} = \frac{h_{jCi}}{Q_{P0}} \tag{4.106}$$

At medium forward bias of the BE junction the model equation is again simplified due to the vanishing BC charge and becomes independent of the zero bias hole charge Q_{P0} as follows:

$$I_C\big|_{V_{BC}=0} = \frac{I_S\cdot\exp\left(\frac{V_{BE}}{V_T}\right)}{1 + h_{jEi0}Q_{jEi}}. \tag{4.107}$$

After identical transformation, the formula suitable for parameter extraction reads

$$\frac{\exp\left(\frac{V_{BE}}{V_T}\right)}{I_C} = \frac{1}{I_S} + \frac{h_{jEi0}}{I_S} Q_{jEi}. \tag{4.108}$$

Using the effective emitter area calculated after determining the periphery to area ratio (γ_C) the two contributions (J_{CA} and I_{CP}) are separated in a first step (cf. Fig. 4.125 and 4.124) for a selected sub-set of devices (e.g. at constant l_E).

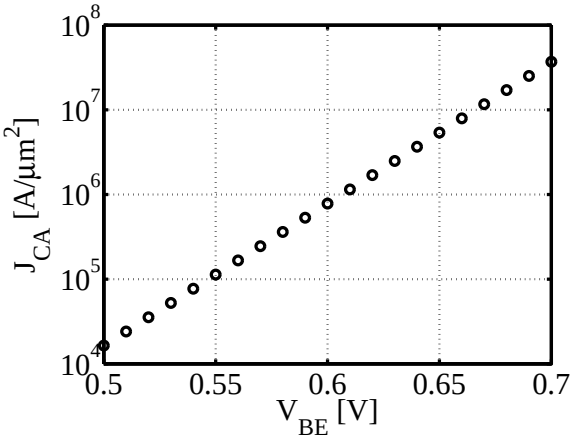


Figure 4.123: Area-effective collector current density; (*symmetrical devices with* $l_E = 5\mu m$)

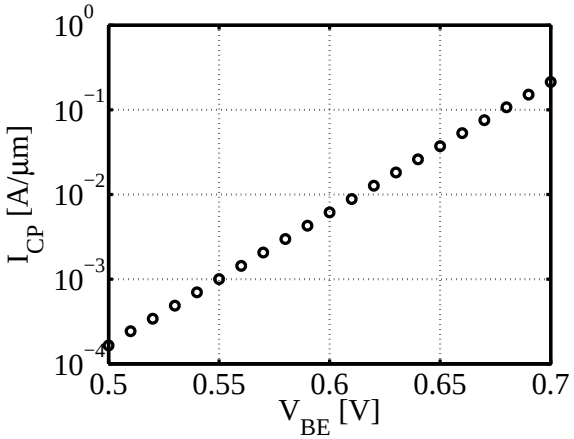


Figure 4.124: Perimeter related collector current contribution; (*devices with* $l_E = 5\mu m$)

Using the ideal junction relation and extrapolation of respective saturation currents (J_{SA} and I_{SP}) one obtains the according normalized currents (cf. Fig. 4.125 and 4.126).

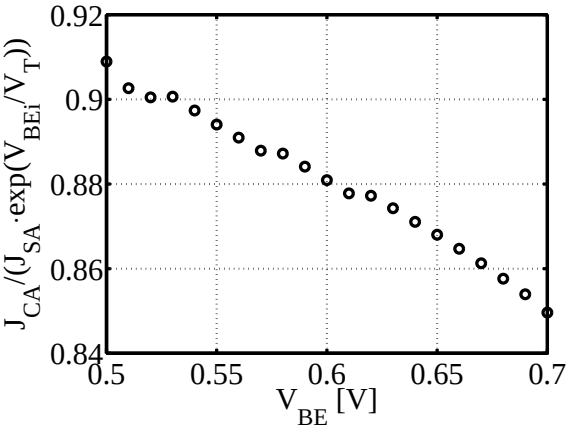


Figure 4.125: Area-effective collector current density normalized by means of the according saturation current density J_{SA}

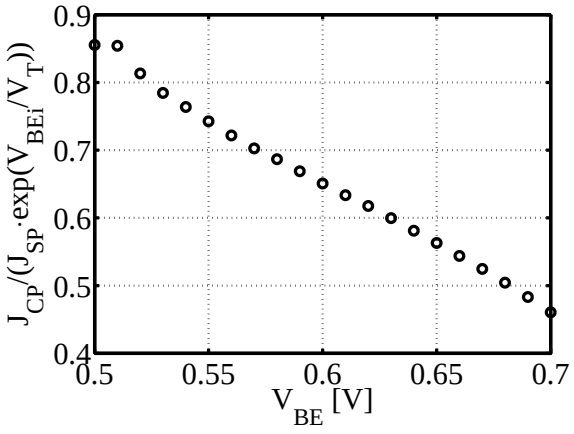


Figure 4.126: Perimeter related collector current contribution normalized with the lineic saturation current I_{SP}

The actual parameter extraction procedure is then performed using the area-related collector current density ($I_{C,n}/A$ or J_{CA} respectively) separated from measured data in the previous step. In the initial model evaluation phase it was concluded that the model parameter r_{hjEi}, responsible for smoothing and upper bias limitation, has no significant impact and can be kept at its pre-defined (default) value. Given the BE junction parameters V_D, Z_E already determined in previous steps the maximization of the correlation coefficient of q_{jEa^*} versus the inverse normalized collector current density ($\exp(V_{BE}/V_T)/J_{CA}$) is performed in an iterative cycle in order to find the best-fitting value of a_{hjei} at fixed r_{hjei} within the specified V_{BE} bias range.

Therefore the normalization of Q_{jEa} is performed by multiplication with the exponential bias-factor as a function of the parameter a_{hjei}, which can then be determined in optimization loops

$$q_{jEa^*} = Q_{jEa}\frac{\exp(u)-1}{u}. \tag{4.109}$$

As a result of this iterative optimization one may then calculate a normalized area-effective hole charge for extraction of GICCR constant as shown in Fig. 4.127. In a next step intermediate parameters J_{Sa} and H_{jE0U} are extracted from slope and intersect of a linear regression analysis (cf. Fig. 4.128). At this point the desired model parameter Q_{P0U} cannot be ultimately determined since there is a multitude of combinations of the unitary parameter ratio of C_{10U} and Q_{P0U} that yield identical curve progression. Q_{P0U} has to be re-adjusted under high collector current, yet can be fixed at an initial value (e.g. $50fC/\mu m^2$) in order to verify the accurate determination of A_{HJEI} and C_{10U}.

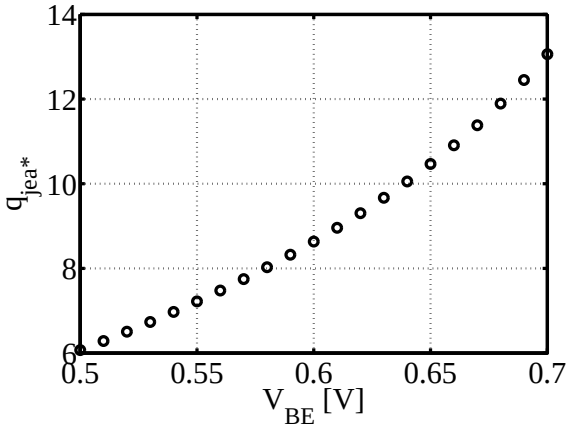


Figure 4.127: Normalized area-effective hole charge for extraction of GICCR constant

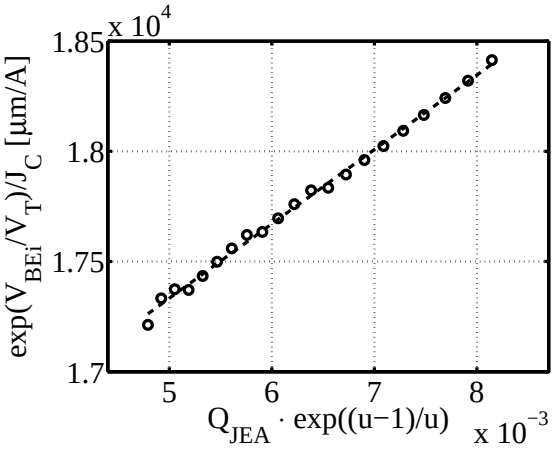


Figure 4.128: Result of iterative correlation coefficient optimization

In a final step the parameter verification can be done using the forward Gummel characteristic at low bias for a selected number of devices. The comparison of scaling for several geometries using

the normalized current (cf. Fig. 4.129) confirms the scaling of the determined set of parameters. Here the accuracy of the collector current modeling under low and medium bias can be verified.

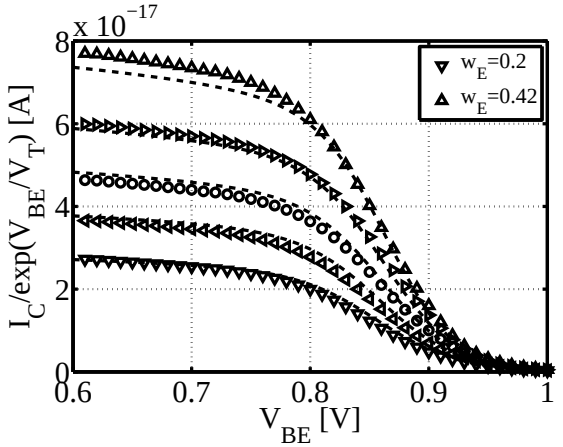


Figure 4.129: Normalized collector current (I_{Cn}) from Gummel characteristic as a function of BE bias, ($l_E = 5\mu m$)

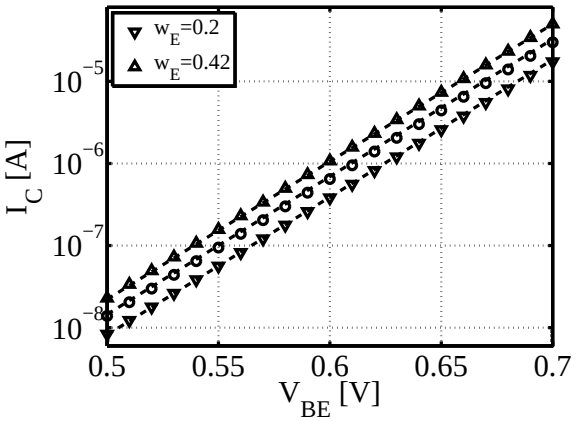


Figure 4.130: Verification of absolute collector current (I_C) in Gummel plot for different device geometries

A verification of both currents in the base and the collector can be made using their relation through the forward current gain ratio (β) of the transistor. Due to its sensitivity to both currents the plot of β obtained from the Gummel characteristic as a function of bias can serve as a first indicator if all domains are correctly modeled (cf. Fig. 4.131 and 4.132).

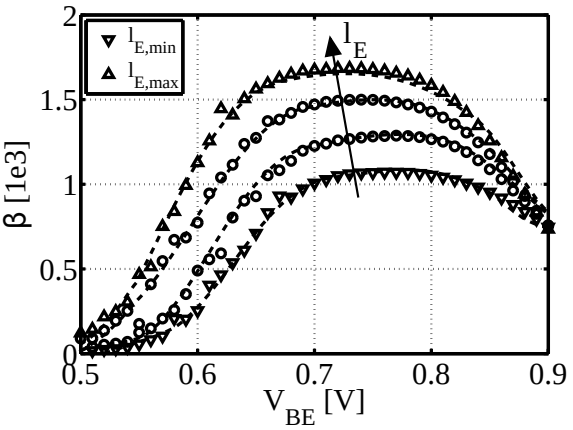


Figure 4.131: Current gain β versus BE bias; *length scaling for devices with* $w_E = 0.2\mu m$

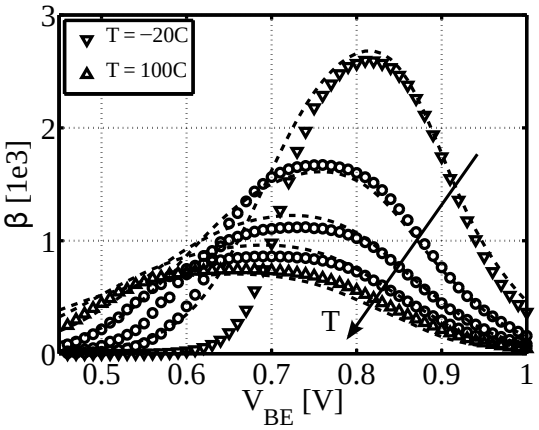


Figure 4.132: Current gain β versus BE bias at different ambient temperatures; *verification for single device with* $w_E = 0.2\mu m$ *and* $l_E = 5.0\mu m$

4.6.4 Additional Charge Weighting Factors

In order to correctly model HBTs there are additional weighting factors to model the band gap difference between base and emitter (BE junction) as well as base and collector (BC junction). As described in Chapter 2.1 the intrinsic carrier concentration n_i may be changing through band gap narrowing due to high doping or Ge incorporation with significant impact on the charge weighting. Though especially for high-speed double-HBT transistors the GICCR weighting factors are significantly different from their default values ($h_{fE} = 1$ and $h_{fC} = 1$) and should be initialized carefully. Weighting factors H_{FE} and H_{FC} can be used to fine-tune the model accuracy under high current densities of I_C.

$$h_{fE} = \frac{\overline{\mu_{nB} n_{iB}^2}}{\mu_{nE} n_{iE}^2} \quad \text{and} \quad h_{fC} = \frac{\overline{\mu_{nB} n_{iB}^2}}{\mu_{nC} n_{iC}^2} \tag{4.110}$$

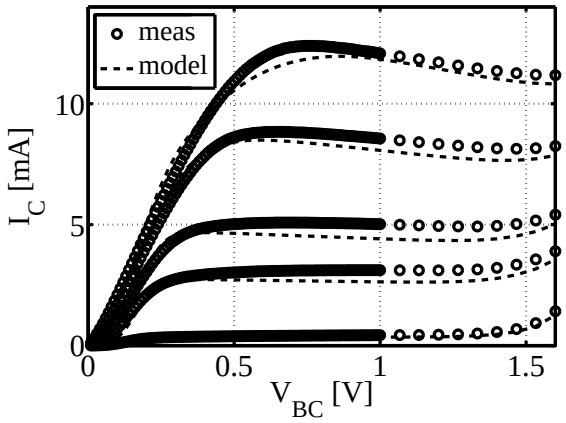


Figure 4.133: Collector current (I_C) in output characteristic for different $I_B = const.$, for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

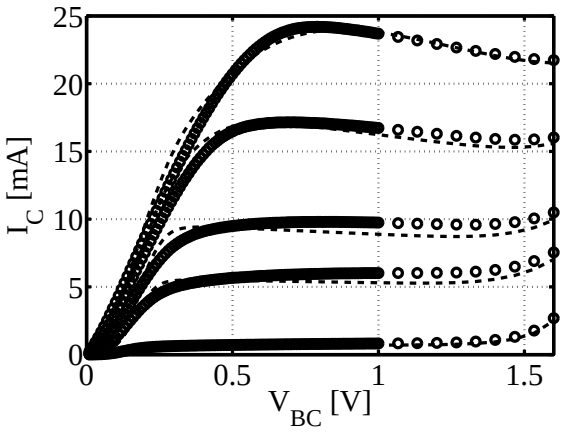


Figure 4.134: Collector current (I_C) in output characteristic for different $I_B = const.$, for single device ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

4.7 Temperature Dependence

This section is dedicated to model parameters linked to the changes of model parameters and device characteristics resulting from environmental temperature variation and self-heating. To accurately model temperature dependence, the scaling with (junction) temperature is a substantial part of the semi-physical compact model HICUM. A rather simplified set of expressions representing underlying physical mechanisms is implemented, yet found suitable for compact modeling. The physical effects that are responsible for the temperature dependence of semiconductor devices are mostly related to the impact on material parameters. Thus in general the parameters describing the temperature dependence are unitary parameters.

Since time for measurement and extraction is limited , only a selected (reduced) subset of configurations as well as characteristics is required. The data used and presented subsequently is obtained from the same RF device structures as the RF data at room temperature. The temperature range typically spreads from -20C through 100C in the six steps referenced in table 4.6. Extensions to both sides are possible (e.g. [-40 ... 150C]). Due to the high effort of RF measurements and limitations of the instrument, the range of (normalized) temperature is slightly reduced compared to the full range shown in Fig. 3.42, yet still sufficiently large for the intended purpose.

$T[C]$	-20	0	25	50	75	100
$T[K]$	253	273	298	323	348	373
T/T_{ref}	0.85	0.92	1.00	1.08	1.17	1.25

Table 4.6: Set of measured temperatures on RF devices for model extraction and verification with a subset of configurations

In the latest model release (ver. L2.3x) only minor adaptations to the existing temperature scaling rules were made. The most significant change is the effect on the two weighting factors H_{JEI0} and H_{F0} in order to model SiGe devices with strongly graded Ge profile.

4.7.1 Resistances

Temperature dependence of the sheet resistances can be derived from the according temperature dependence of the mobility. In order to keep the model equations as simple as possible, the mobility itself is modeled using an empirical temperature equation (cf. eqn. 4.111). Accordingly the temperature scaling of all resistances present in the HICUM equivalent circuit is modeled using the temperature exponent factor ζ (Z_{ETA} respectively), applied to the normalized temperature (T_n).

The model parameters for temperature dependence can in principle be obtained repeating the extraction steps presented in the respective section at different ambient temperatures.

$$\mu(T) = \mu(T_0)\left(\frac{T}{T_0}\right)^{\zeta_\mu} \qquad \text{empirical mobility scaling} \qquad (4.111)$$

Since DC measurements of resistances are accomplished in less time once the system is set up, the measurement range for both R_B as well as R_{Cx} was extended to a range of -40C through 150C. The temperature scaling coefficients are obtained by simple least squares curve optimization of the respective scaling equation applied to the extracted sheet resistances. Due to the split in internal and external contribution, the base resistance has two individual temperature parameters (Z_{ETARBI} and Z_{ETARBX}). Typically these are significantly different due to their relation to the used materials[1].

$$R_{BI}(T) = R_{BI0}(T_0)\left(\frac{T}{T_0}\right)^{Z_{ETARBI}} \qquad \text{temperature scaling for } R_{Bi} \qquad (4.112)$$

$$R_{BX}(T) = R_{BX}(T_0)\left(\frac{T}{T_0}\right)^{Z_{ETARBX}} \qquad \text{temperature scaling for } R_{Bx} \qquad (4.113)$$

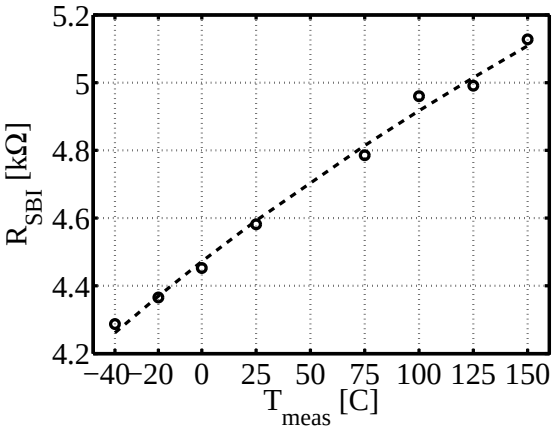


Figure 4.135: Temperature scaling of the inner base sheet resistance (R_{SBI}), obtained from extraction using measured tetrode data at different ambient temperatures T_{meas}

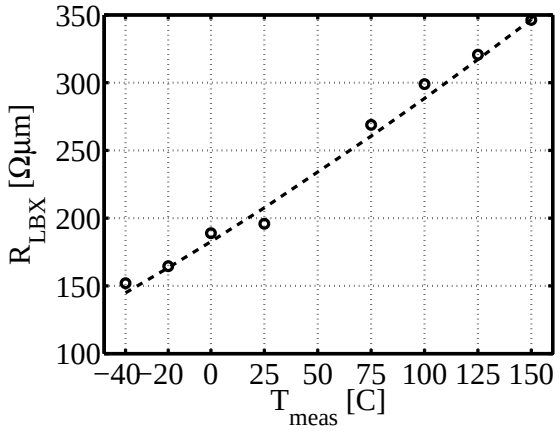


Figure 4.136: Temperature scaling of the extrinsic base link resistance (R_{LBX}) obtained from direct extraction using tetrode measurements

Since the emitter resistance is determined using RF measurement data in forward active mode, the parameter Z_{ETARE} is obtained in the same range as shown for the RF characteristics. The via resistance R_{KVIA} is considered to be less affected by changes of ambient temperature thus the temperature scaling is attributed to the poly-silicon link contribution R_{KE}. With the electrical parameters obtained at each of the measured temperatures, the extraction of the temperature coefficient related to the emitter poly-silicon resistance (Z_{ETARE}) can be performed (cf. Fig. 4.137).

[1] in the analyzed process $Z_{ETARBI} \approx 0.3$ and $Z_{ETARBX} \approx 1.4$ were determined

The collector resistance is modeled in the same way using one scaling parameter Z_{ETARCX}. The two contributing resistivities are analyzed individually in a first step yet the obtained exponent factors yield almost the same value. Thus a single, averaged value is included in the model library.

As can be concluded from the curve progression in Fig. 4.137 a negative temperature exponent was found for the total emitter series resistance R_{KE} ($Z_{ETARE} < 0$) whereas the external collector resistance rises with increasing temperature[1].

$$R_E(T) = R_E(T_0)\left(\frac{T}{T_0}\right)^{Z_{ETARE}} \qquad \text{temperature scaling for } R_E \qquad (4.114)$$

$$R_{CX}(T) = R_{CX}(T_0)\left(\frac{T}{T_0}\right)^{Z_{ETARCX}} \qquad \text{temperature scaling for } R_{Cx} \qquad (4.115)$$

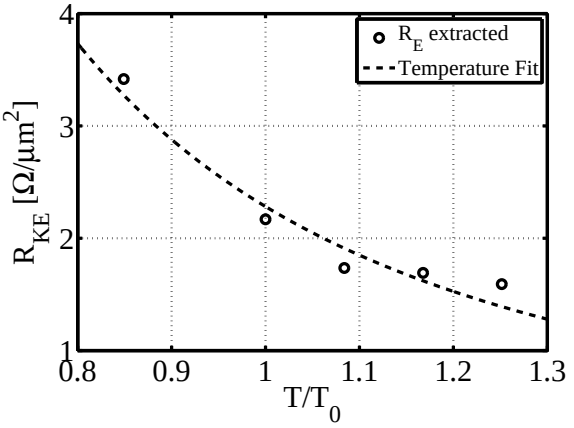


Figure 4.137: Temperature scaling of the poly-emitter resistance (R_E) obtained from direct extraction from RF measurements on multiple geometries

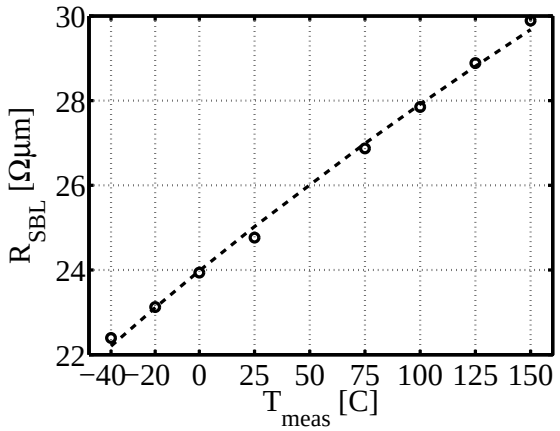


Figure 4.138: Temperature scaling of the buried layer contribution of the extrinsic collector resistance (R_{SBL}) obtained from direct extraction

With the collector resistance extraction performed at each temperature step, slightly different values for the geometry correction dw_{BL} are obtained. Since the model library uses one single value, an averaging over temperature is performed in order to obtain a unique value.

[1] values of $Z_{ETARE} \approx -0.7$ and $Z_{ETARBX} \approx 0.45$ were determined

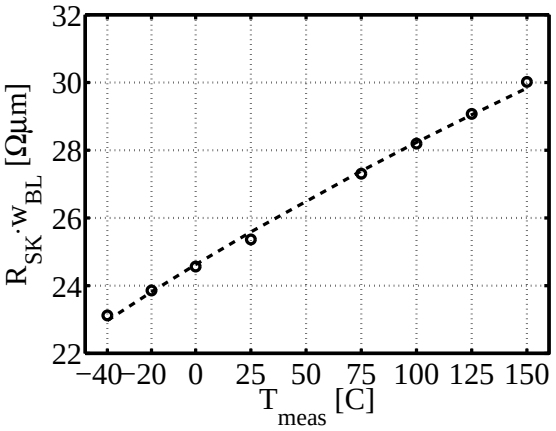


Figure 4.139: Temperature scaling of the sinker contribution of the extrinsic collector resistance (R_{LSK}) obtained from direct extraction

4.7.2 Capacitances

At the basis of the corresponding model equations (cf. eqn. 3.236 in sect. 3.4) the verification of the temperature scaling of the capacitances is a good indicator for the fitting of the two bandgap-voltages V_{GB} and V_{GE}.

A verification of the temperature scaling of the base-emitter junction capacitance C_{BE} is preformed using some selected individual geometries (cf. Fig. 4.140 and 4.141). Even though a trend of the extracted zero-bias junction capacitance is visible, the change of charge storage is known to be much less sensitive to variation of the ambient temperature than the resistances.

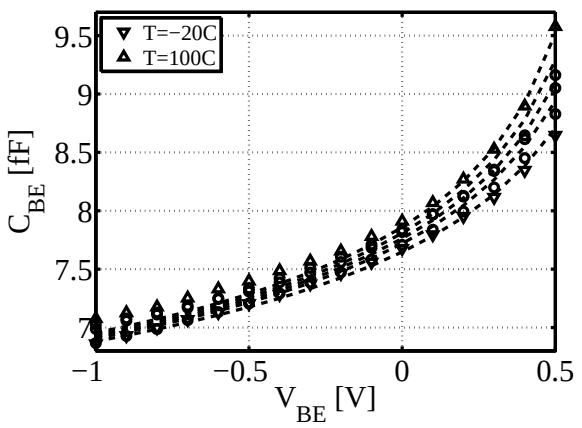


Figure 4.140: Verification of temperature scaling of the base-emitter junction capacitance C_{BE} of a single device $w_E = 0.2\mu m, l_E = 5\mu m$

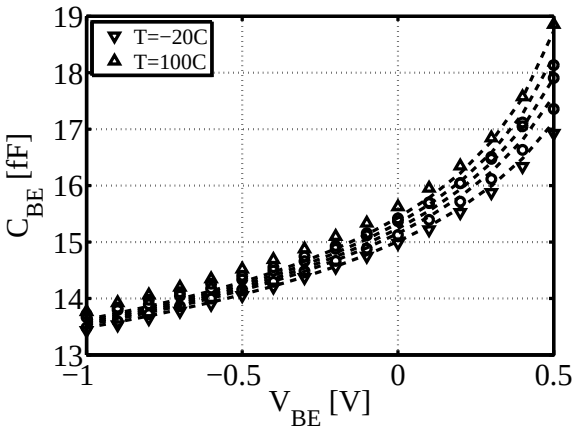


Figure 4.141: Verification of temperature scaling of the base-emitter junction capacitance C_{BE} of a single device $w_E = 0.2\mu m, l_E = 10\mu m$

The exact same verification can be done for the BC junction, where the temperature sensitivity is again very low.

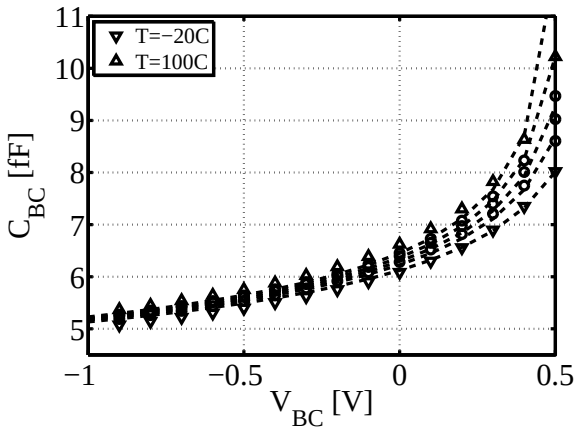


Figure 4.142: Verification of temperature scaling of the base-collector junction capacitance C_{BC} of a single device $w_E = 0.2\mu m, l_E = 5\mu m$

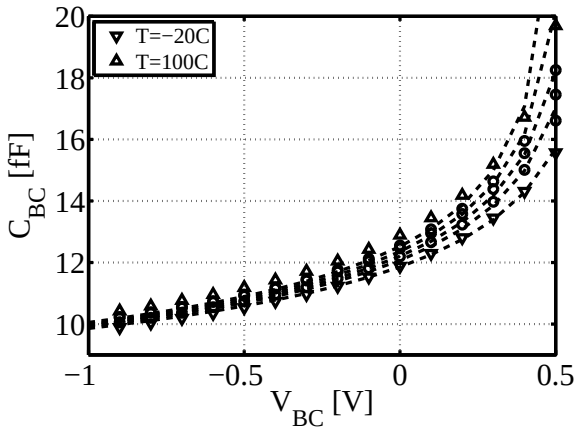


Figure 4.143: Verification of temperature scaling of the base-collector junction capacitance C_{BC} of a single device $w_E = 0.2\mu m, l_E = 10\mu m$

4.7.3 Base Current

The temperature scaling equation describing the base current is governed by two parameters: The exponent factor (Z_{ETABET}) and the band-gap voltage of the BE junction (V_{GE}). Since the band-gap energy E_G is related to the respective voltage through $V_G = E_G/q$, for most current components, the average value of the band-gap between adjacent regions is taken as reference for temperature scaling. As the band-gap voltage is a strong function of the thermal voltage (V_T, cf. eqn. H.3), these two parameters are giving sufficient flexibility for modeling.

In order to obtain the effective band-gap parameters used for the temperature scaling in the BE junction scaling equation (cf. eqn. 2.29) the following calculations for effective band-gap voltage are implemented in HICUM:

$$V_{gBEeff} = \frac{V_{gBeff} - V_{gEeff}}{M_{JR}}. \tag{4.116}$$

The parameter in the temperature equations for the BJT model are extracted using the measured base current in forward operation, preferably at zero BC bias.

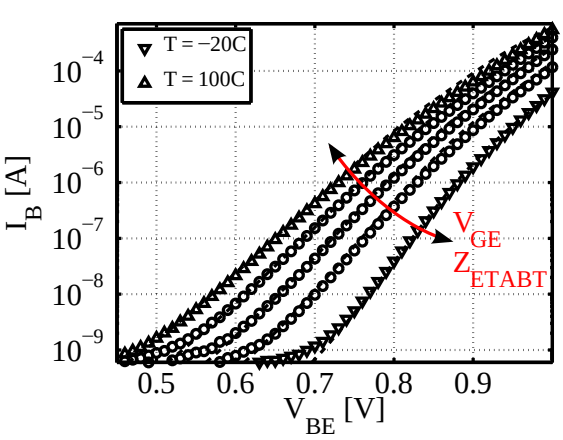


Figure 4.144: Temperature scaling of the base current (I_B) in forward Gummel characteristic for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

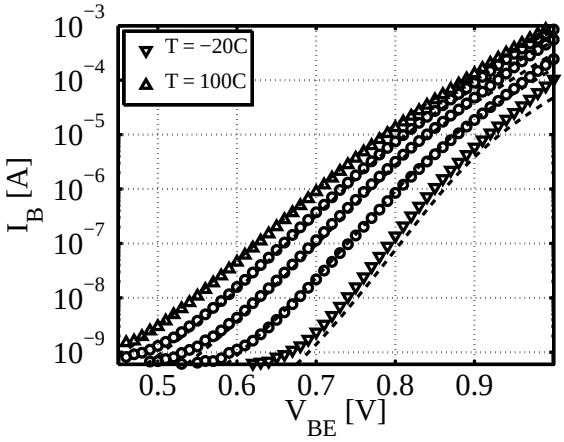


Figure 4.145: Temperature scaling of the base current (I_B) in forward Gummel characteristic for single device ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

4.7.4 Avalanche Current

As shown in equations (3.248) and (3.249), the avalanche breakdown current temperature dependence is scaling linearly with the (unitary) exponential factors A_{LFAV} and A_{LQAV}. Thus experimental data from one single device (e.g. reference device) is analyzed at different temperatures as shown in Fig. 4.146.

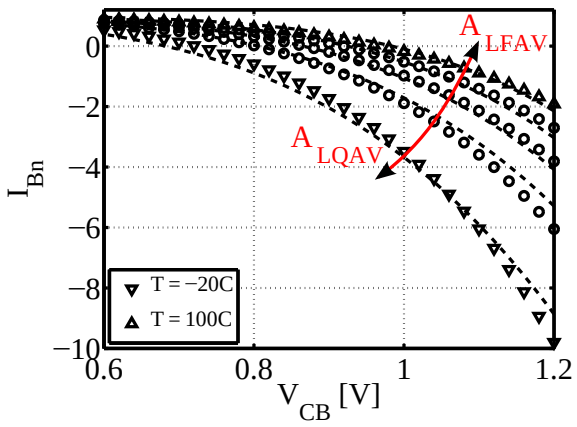


Figure 4.146: Temperature scaling of the avalanche current (I_{BAVAL}) in Early characteristic for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

4.7.5 Collector Current

The dependence of the transfer current on temperature in the low current range is modeled using unitary parameters independent of the device geometry. For the correct temperature dependence of the forward collector current the Gummel characteristic at V_{BC0} is analyzed using a single device geometry (and a second one for verification). The parameters used to model the influence of ambient temperature are the exponent factor Z_{ETACT} and the base band-gap voltage V_{GB}.

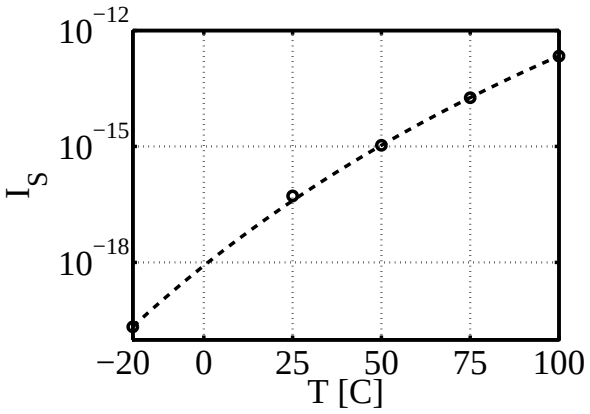


Figure 4.147: Temperature scaling of the saturation current (I_S) for single geometry ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

$$C_{10}(T) = C_{10}(T_0)\left(\frac{T}{T_0}\right)^{Z} \cdot \exp\left[\frac{V_{GB}}{V_T}\left(\frac{T}{T_0} - 1\right)\right] \tag{4.117}$$

$$V_{GB} = \frac{\ln\left(\frac{I_S(T_1)}{I_S(T_0)}\right) - Z\ln\left(\frac{T_1}{T_0}\right)}{\left(\frac{T_1}{T_0} - 1\right)} \tag{4.118}$$

with $Z = Z_{ETACT} = 3$ for initialization

These parameters however are best obtained with the respective weighting function parameters describing the bias dependence (cf. sect. 4.6.3). Since the value of V_{GB} can be determined from the temperature dependence of the saturation current I_S or the GICCR transfer-current constant C_{10} (cf. eqn. 4.118), only the exponent factor Z_{ETACT} remains to be adjusted.

After conscious optimization of the parameter set the model can be validated using the measured collector current in the forward Gummel plot.

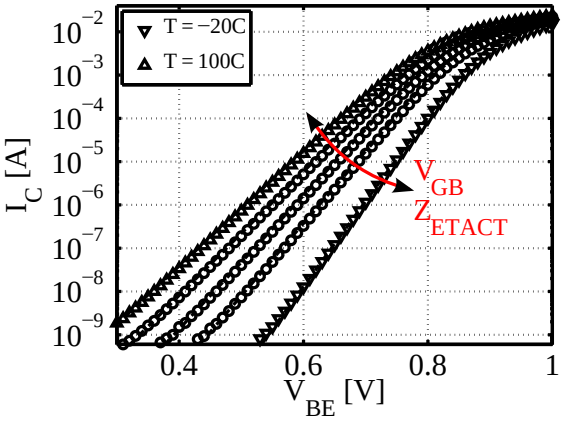


Figure 4.148: Temperature scaling of the collector current (I_C) in forward Gummel characteristic for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

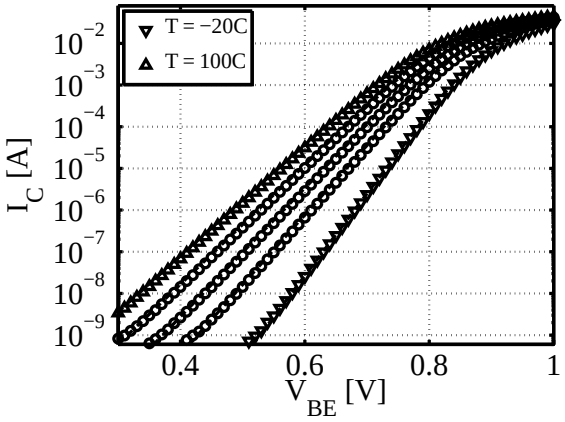


Figure 4.149: Temperature scaling of the collector current (I_C) in forward Gummel characteristic for single device ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

The spread of the I_C curves in reverse mode is adjusted using the band-gap voltage parameter V_{GS} respective to the substrate diode.

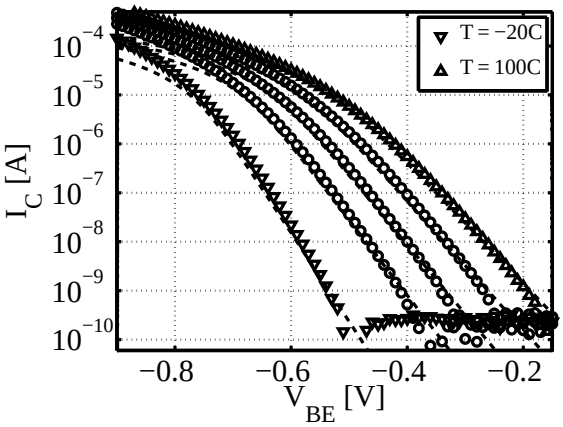


Figure 4.150: Temperature scaling of the collector current (I_C) in reverse operation for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

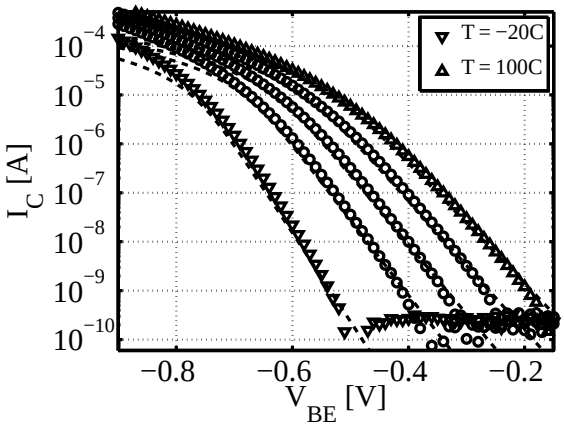


Figure 4.151: Temperature scaling of the collector current (I_C) in reverse operation for single device ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

4.7.6 Temperature Dependence of the Transit Time

Due to various effects the transit time τ_0 in itself is modeled temperature-dependent as well. The temperature dependence of the transit times can be split in high current and low current range. In low current operation the time constants are dominated by the saturation velocity (v_S) which is increasing with temperature, effectively lowering the transit time. In contrast under high injection the mobility degradation in the collector region dominates the transit time and causes a significant

decrease with rising temperature. The second order formula to model the forward transit time as function of temperature is as follows

$$\tau_0(T) = \tau_0(T_0) \cdot [1 + \alpha_{\tau 0}\Delta T + k_{\tau 0}\Delta T^2]. \tag{4.119}$$

To adjust the observed decrease of T_0 under high injection the two parameters A_{LT0} and K_{T0} are optimized. However under high injection another effect has to be considered. The temperature dependence of the critical current is taken into account through the variation of the inner collector resistivity ($R_{CI}(T)$) modeled by the corresponding exponent factor Z_{ETACI}.

On a single geometry device, the characteristics at zero base-collector bias are optimized to reproduce the measured characteristics. In order to cancel out errors related to the modeling of the transfer current (avoid a corruption due to the dependence of I_C on T) the f_T and f_{max} verification is done as a function of V_{BE} (cf. Fig. 4.152). To draw a complete picture the usual plot as function of I_C (cf. Fig. 4.153) is verified alongside.

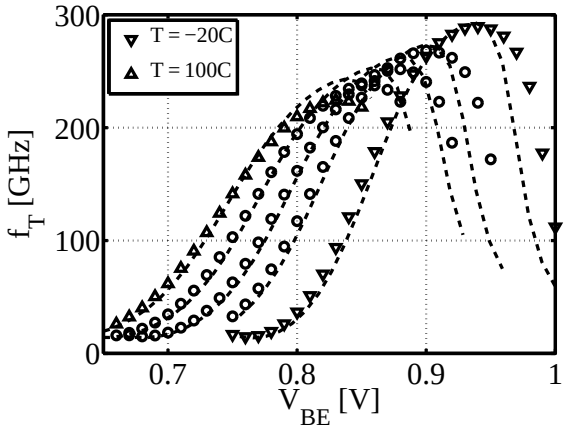


Figure 4.152: Temperature scaling of the transit frequency (f_T) as function of bias voltage V_{BE} in forward operation for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

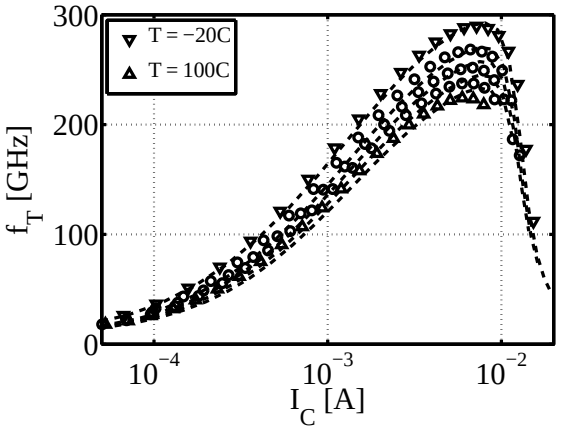


Figure 4.153: Temperature scaling of the transit frequency (f_T) as function of collector current I_C in forward operation for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

As pointed out before, the charge weighting factors are scaled temperature dependent, which also has an impact on the device speed. In general an exponential relation with the band-gap voltage of each region is assumed. Thus the definition of individual band-gap voltages in the following general form is used:

$$h(T) = h(T_0) \exp\left[\frac{\Delta V_G}{V_T}\left(\frac{T}{T_0} - 1\right)\right]. \tag{4.120}$$

For additional verification, the temperature progression of the small signal current gain (h_{21}) is simulated. For a fixed V_{BE} bias the peak f_T point is reached at lower collector currents with decreasing temperature. Likewise the scaling of the transit time at low injection can be verified using the plot as function of the inverse transfer current.

The transit time related weighting factor H_{F0} is scaled temperature dependent with the band-gap difference ΔV_{GBE} since the most part of minority charges at low injection is found in the BE SCR.

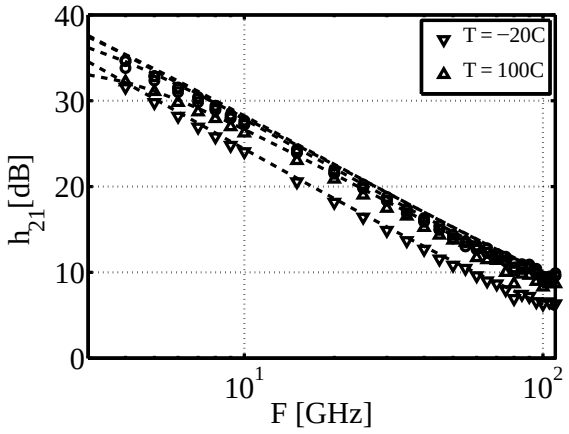


Figure 4.154: Temperature scaling of the small signal current gain (h_{21}) as function of RF frequency at fixed bias voltage $V_{BE} = 0.85V$ for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

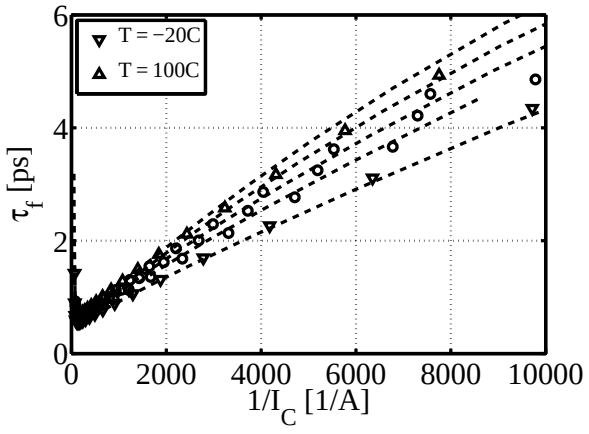


Figure 4.155: Temperature scaling of the transit time (τ_f) as function of the inverse collector current $1/I_C$ for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

4.8 Self Heating

Despite the favorable thermal properties of bulk silicon as base material for the BiCMOS process (compared to other technologies such as SOI or III-V), thermal issues (self heating) represent an emerging concern in SiGe:C HBT technologies for high frequency applications. In RF and high bit-rate applications transistors operate at maximum frequency. Highest operating frequencies however are reached under high current operation.

Yet one drawback under high current operation is localized heating under the active area of a HBT due to trapped heat. This effect is referred to as self heating and characterized by a thermal spreading impedance. The common adoption of advanced isolation techniques such the deep trench oxide isolation (DTI) to meet the demand for higher speed and integration density further degraded the heat spreading across the chip. Given the shrinking lateral device dimensions of latest technologies, the effect of self-heating is even more prominent and changing the internal operating temperature of a device far beyond the ambient temperature.

A serious phenomenon of concern in HBT devices is a thermal runaway. The increase of device temperature leads to higher current and power dissipation, which additionally increases the temperature further until the device is destroyed. As this imposes a serious design constraint thermal runaway and breakdown needs to be avoided by restricting the operating conditions of the device to a safe operating area (SOA).

For accurate device modeling the actual device temperature must be known due to its influence on the calculation of the thermal voltage (V_T), which is exponentially weightet in several model equations. For the electrical device characteristics self heating in HBTs has two major consequences: the emitter junction potential is lowered (visible in I_C increase with V_{CE} at constant V_{BE} in the output curve) and reverse hole injection from the base to the emitter is increased (visible in I_C decrease with V_{CE} at constant I_B injection). The predominant challenges for model parameter extraction related to self-heating is the feedback of device temperature being bias dependent and vice versa. Thus the variation of the device junction temperature can be summarized as follows:

- the collector current I_C is a function of temperature
- the small signal gain g_m is temperature dependent
- the critical current I_{CK} is shifted
- the transit time parameters (τ_f) depends on the device temperature

Accurate parameter extraction thus becomes very complex once the transistor operates under high-injection. In consequence the topic of self-heating has been studied extensively and several different characterization techniques have been presented [118–123] ranging from DC and time-domain measured data through application of low-frequency S-parameter measurement.

Given the strong analogy between thermal and electrical behavior, the calculation of a thermal

resistance is conducted as for a general resistance in semiconductors (using thermal capacitances and resistances). The dissipated power at a given operating point is calculated using bias information. With a dedicated thermal network (cf. Fig. 2.10), the difference to ambient temperature can be reproduced since voltage and temperature nodes are evaluated simultaneously.

The electro-thermal modeling is related to the most important parameter R_{TH}, usually extracted using DC measurements, whilst the thermal time constant C_{TH} cannot be obtained from DC measurements. However the accurate extraction of model parameters of the conventional single pole thermal network (as implemented in the HICUM model) are likewise possible with other techniques such as quasi-isothermal, pulsed measurements. One can thus distinguish the following techniques for self-heating effect characterization in order to obtain the desired parameters of the thermal network:

- CW and pulsed DC and RF measurements
- Low frequency S-parameter measurements (30kHz .. 3GHz)
- Transient electro-thermal simulation (TCAD)

4.8.1 Parameter Extraction

Due to the strong impact of self heating on the RF characteristics it is desirable to extract the respective parameters before assessment of the RF and high-current parameters. In measured DC data the presence of self heating can be visualized through two significant effects, that may be distinguished in the output characteristic or the respective normalized collector current curve (I_C vs. V_{CB}, cf. Fig 4.156), where the deviation of the collector current is compared to a reference value. Firstly the decrease of the BE junction built-in potential (V_{DBE}) causes I_C to increase with V_{CE} at constant V_{BE} bias. Secondly increased reverse hole injection from the base to the emitter causes an apparent I_C decrease under constant I_B condition. However for parameter determination another approach is preferred.

As a first parameter R_{TH} is assessed. The thermal resistance quantifies the devices capability to transfer heat within a given thermal path. For parameter extraction data is obtained from DC measurements (forward Early characteristic) carried out at different ambient temperatures (T_{amb}). Given the impact of self heating on the highly temperature sensitive base current I_B, the resistive parameters for the Z_{TH} network (R_{0TH_SI}, A_{LRTH}) of the devices are extracted as proposed in [124]. However given the new implementation in the recent model update to HICUM L2.32, the exponent factor (Z_{ETARTH}) can be adjusted as well

$$R_{TH}(T) = R_{TH}(T_0) \cdot \left[1 - A_{LRTH}\Delta T + \left(\frac{T}{T_0} \right)^{Z_{ETARTH}} \right]. \tag{4.121}$$

For parameter extraction, the transistor is measured at fixed BE bias under variation of V_{BC}. The relative change of the base current, calculated from the reference base current gives the $I_{B,n}$ curve as seen in Fig. 4.157 which is then related to the dissipated power in the base ($P_D(V_{BE}, V_{BC})$) as follows

$$i_{B,n} = \frac{I_B(V_{BE}, V_{BC}) - I_B(V_{BE}, V_{BC,0})}{I_B(V_{BE}, V_{BC,0})}, \text{ and} \tag{4.122}$$

$$T_j(P_D) = T_A + R_{TH} \cdot P_D, \quad R_{TH} \approx \frac{k \cdot T^2}{E_G - qV_{BE}} \left(\frac{i_B}{\Delta P_D} \right) \text{ [125].} \tag{4.123}$$

Using a single device geometry, the normalized base current (cf. Fig. 4.157) is suitable to extract the static thermal resistance. This method is suited for high V_{BE} bias, since self heating is causing a shift in thermal voltage ($V_T(T_A) \neq V_T(T_J)$) and an apparent increase of the normalized base current I_B/I_{B0}.

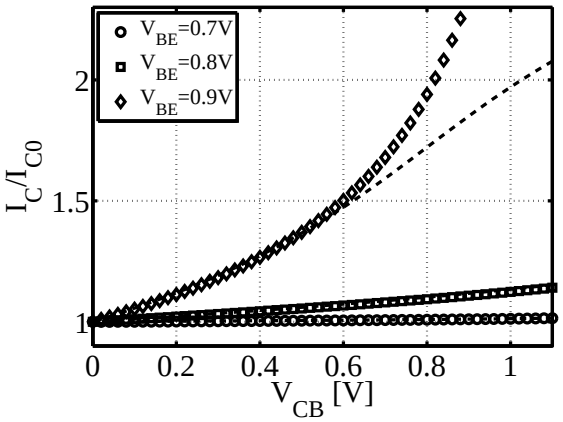


Figure 4.156: Normalize collector current indicating the deviation from a reference value under high forward bias, single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

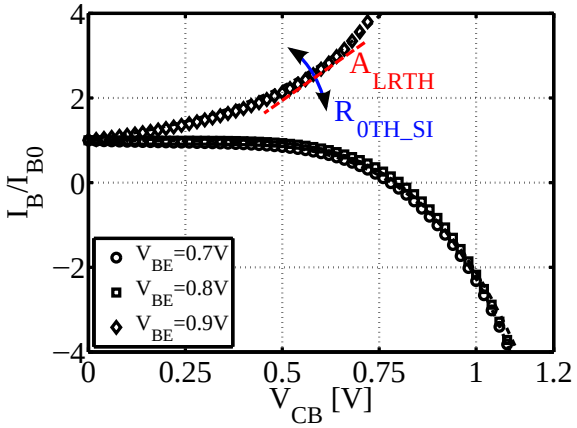


Figure 4.157: Impact of self heating on single device in Early characteristic as function of the BC bias V_{BC} at different V_{BE} ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

The thermal impedance (Z_{TH}) is defined as the thermal response of a given system to a induced power (P_D) as a function of time (t)

$$Z_{TH}(t) = \frac{\Delta T(t)}{P_D}. \tag{4.124}$$

For application in HBT modeling it provides information about the dynamic thermal properties of a semiconductor device. The key parameter related to the thermal network under dynamic power dissipation is the thermal capacitance, C_{TH}. This quantity is a measure of the capability of the device to absorb heat from the heat source and accumulate this heat, in analogy to a capacitor that

accumulates charge. The value of C_{TH} is best determined using measured data from dedicated low-frequency measurements. Small-signal Y-parameter measurements in a low frequency range (few kHz up to GHz) as well as according DC measurements are performed at biases close to the maximum f_T value, where self-heating has a significant impact.

At very low frequency the measured magnitude of Z_{TH} is dominated by R_{TH} allowing to verify extractions from DC measured data. Towards higher measurement frequencies, the decay of Z_{TH} due to heat trapping allows for extraction of C_{TH}. Since this quantity of C_{TH} is directly obtained, it is usually passed to the model without further modification.

After aligning the equivalent network parameters (notably R_{0TH_SI} and A_{LRTH}) using measured DC data of a single device, the geometry scaling may be verified analyzing different geometries at a sufficiently high fixed BE bias ($V_{BE} \geq 0.8V$, cf. Fig. 4.158).

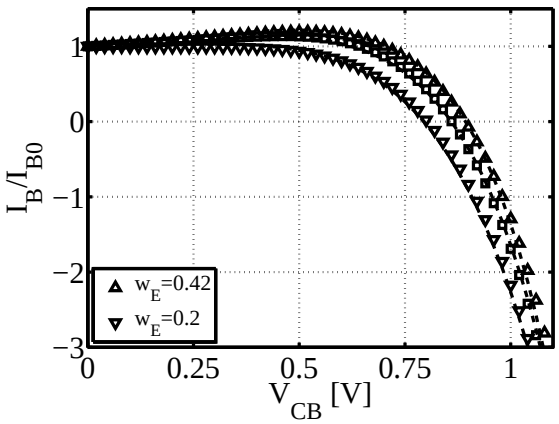


Figure 4.158: Scaling verification of self heating at fixed bias voltage $V_{BE} = 0.8V$ for constant emitter window length ($l_E = 5\mu m$)

4.8.2 Pulsed Measurements

Pulsed measurements are another solution to further investigate strong self-heating effects. In low current operation the device characteristics have to be measured with a high accuracy and measurement noise might not be tolerable. However for high current measurements, where apparent self-heating of the HBT has most influence, measured currents are high. Special pulsed measurement equipment may hence yield sufficient accuracy despite the usually increased noise floor (several nA). Steady state DC measurement data at low bias can be combined with pulsed measurement capabilities to efficiently eliminate the self heating influence resulting in quasi-isothermal measured data throughout the bias range of interest. By comparison with a reference (e.g. pure DC measurement) the model parameters for self-heating effects (notably R_{TH}) might be extracted.

A comparison is done extracting self heating by curve fitting.

The idea of a pulsed measurement is to set a pulse width that is smaller than thermal time constant of the transistor so that there is no significant change in internal device temperature during measurement. This method allows exploration electro-thermal effects as well as the safe operating area (SOA). Yet there are several linked constraints to be considered. One is that reliable measurement of several bias points in a series of measurement (e.g. output characteristic) can only be made once the DUT has been given sufficient time to cool before a new measurement can be started, thus system parameters (such as cycle times etc.) have to be adapted to the specific requirements in HBT characterization. These system parameters (listed below) in are best shown in a schematic pulse diagram for a single point DC measurement (cf. Fig. 4.159).

- Pulse width T_w
- Pulse period T_p
- Duty cycle $D \approx \frac{T_w}{T_p} \cdot 100$

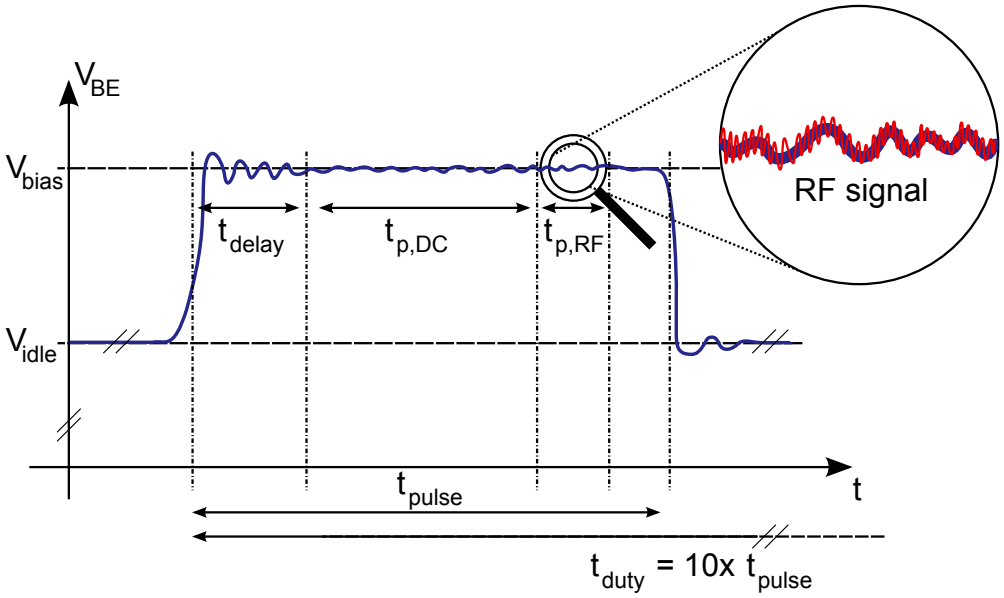


Figure 4.159: Schematic view of pulsed measurement with characteristic measurement times; *Pulse width: t_{pulse}, Delay time before measurement for stabilization: t_{delay}, DC measurement integration time: $t_{p,DC}$, RF measurement time: $t_{p,RF}$, Duty cycle t_{duty} (time between two msmts.)*

Due to a significantly reduced integration time towards small pulse width one has to carefully choose a trade-off between exploitation of isothermal behavior and measurement resolution or accuracy respectively. The current resolution is limited in the μA range. Hence accurate direct measurement of base current is not possible. The system characteristics to be respected have been experimentally determined in a previous assessment of system limitations [126] and can be summarized:

- Pulse width for DC measurement $T_{w,min} \geq 80ns$
- Pulse width for RF measurement $T_{w,min} \geq 100ns$
- Rise and fall time T_{rise} and $T_{fall} \geq 20ns$

In order to measure pulsed I-V and RF characteristics a complex system was set up. A R&S ZVA67 vector network analyzer (frequency range: $10MHz$ to $67GHz$) was paired with a 4200-SCS semiconductor characterization system by Keithley. This instrument comprises a 4225 PMU I-V source measurement unit (SMU) module to set and monitor bias, as well as a 4220-PGU pulse generator (PG) extension that provides voltage-sourcing for pulsed measurement. The PG unit is a dual-channel ultra-fast I-V module integrating voltage waveform generation with a minimum rise time of 10ns, a high current resolution ($100nA$ to $1A$) over a voltage range of 1mV to 40V. It is combined with the precision DC I-V mainframe (Model 4200-SMU). The high-speed voltage outputs of the instrument range from pulse widths of $60ns$ all the way up to a static DC signal. A schematic view of the system configuration used for pulsed on-wafer RF characterization is shwon in Fig. 4.161.

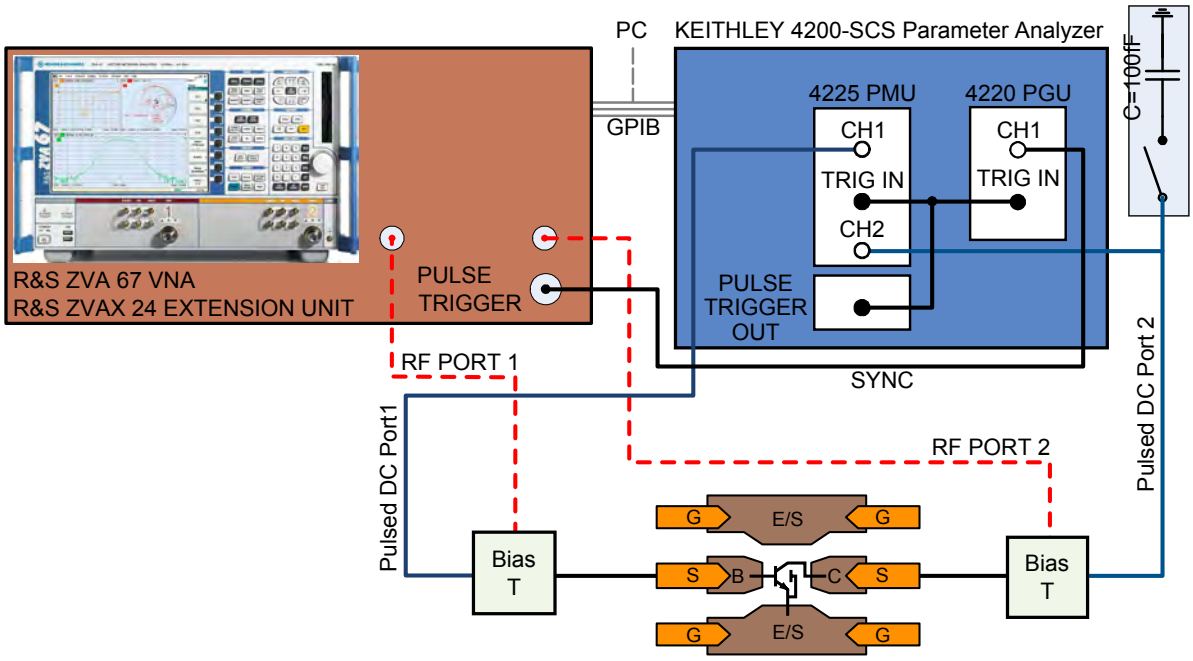


Figure 4.160: Measurement setup used for pulsed RF measurements; *VNA synchronized with parameter analyzer including pulsed DC measurement option*

The most aggressive timing for pulsed S-parameter measurement was at $t_{pulse} = 100ns$. For this setting the RF measurement window $t_{p,RF}$ was set at $10ns$. The idle voltage (V_{idle}) was kept unchanged at the default value $V_{BE} = 0.5V$ throughout the measurement campaign. However especially for pulsed I-V characterization several boundry conditions have to be respected. The

additional parasitics of wiring and bias T (cf. Fig. 4.161) has to be respected when the measured data is to be compared to device simulation for both cases of static (continous wave, CW) as well as transient behaviour.

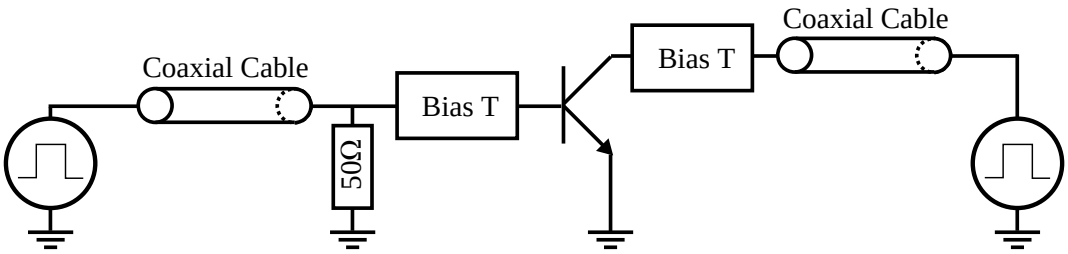


Figure 4.161: Simplified equivalent circuit of elements involved in pulsed measurement

Given the fact that a transistor is a dynamic system associated with charges, applying a pulse to the base yields current flow through the collector, leading to a voltage overshoot at the collector node upon rapid bias changes (cf. Fig. 4.162). Thus for proper device measurement the data sampling has to be performed in a stable region, where transient oscillation disappeared, using an optimized delay time and duty cycle.

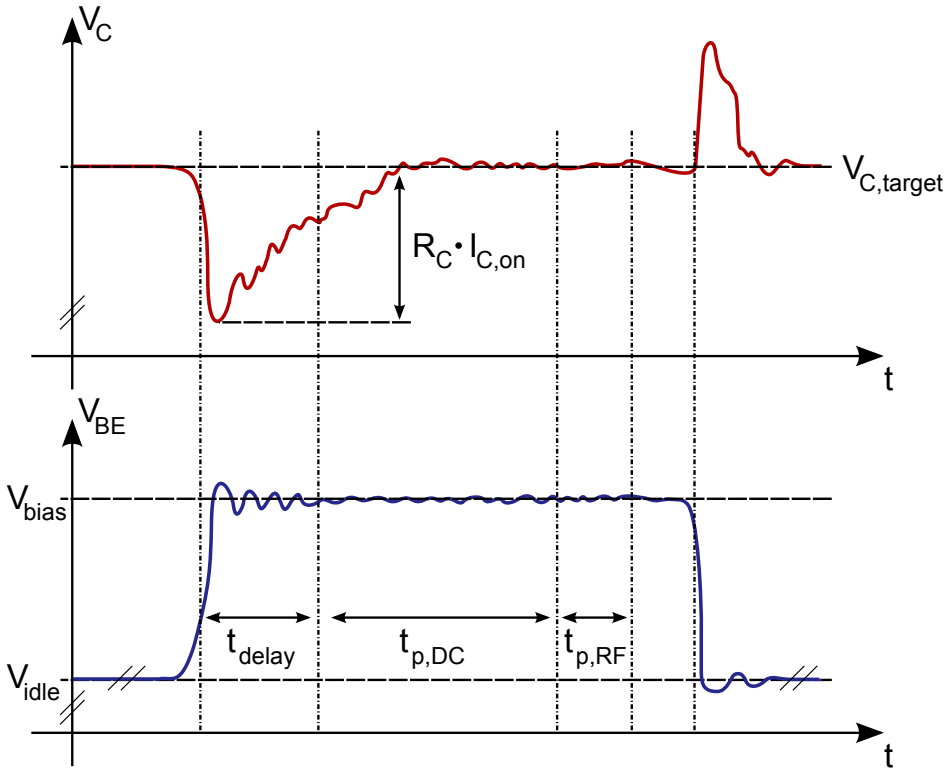


Figure 4.162: Reaction of collector bias to increased current (I_C) upon pulse applied to base

After system evaluation with the DUT these system parameters limits have been determined.

I-V measurements were performed at pulse widths as low as 60*ns* with reliable data from 80*ns* onward. However given the more complex measurement related to instrument synchronization and other timing constraints, high resolution ($\delta V_{BE} = 1mV$) RF measurements were obtained for larger pulse widths $t_{pulse} \geq 100ns$.

The transistor measured in pulsed mode compared to a conventional CW measurement shows significant deviation in the output characteristic (cf. Fig. 4.163). A measurement with a pulse width (t_{pulse}) as high as 1μs coincides with the static DC case and the collector current characteristic matches, whereas towards lower pulse width the noise floor is found to increase. Generally towards higher V_{BE} bias the measured current is higher and thus accuracy is very good (cf. Fig. 4.164).

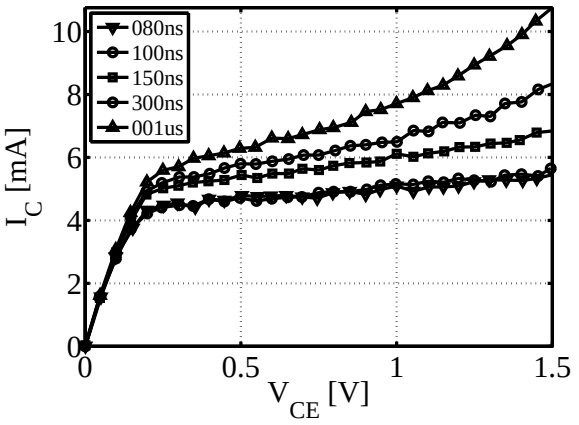


Figure 4.163: Forward output characteristic at fixed BE bias ($V_{BE} = 0.9V$) for variation of collector bias; *Device:* $w_E = 0.18\mu m$, $l_E = 5\mu m$

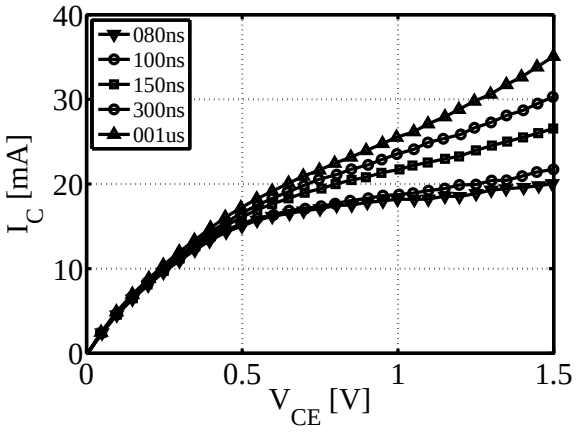


Figure 4.164: Forward output characteristic at fixed BE bias ($V_{BE} = 1.0V$) for variation of collector bias; *Device:* $w_E = 0.18\mu m$, $l_E = 5\mu m$

In order to correctly assess the current at evanescent pulses ($t \to 0$) one may use data extrapolation. Using measured data at different pulse widths with identical bias, one can extract a scaling with the pulse length (t_{pulse}). One approach is to express the measured collector current data (I_C) versus pulse width (t_{pulse}) through an empirically defined hyperbolic function of the form

$$I_C = a \cdot \left[\tanh(b \cdot t_{pulse} + c)\right] + d. \tag{4.125}$$

In the logarithmic scale the used *tanh* function is matching the measured data reasonably well (cf. Fig. 4.165). In order to estimate the influence of self heating on the output characteristic, data fitting at different bias points was be carried out giving a curve progression as shown in Fig. 4.166 and 4.166. With this method a theoretical, self-heating free collector current value ($I_{C,extrapol}$) can be estimated at each bias point through an analysis of the collector current versus pulse width (t_{pulse}). However for sufficiently high accuracy (wrt. system resolution limits at lower bias) this approach is limited to biases in the high injection range (close to and above peak f_T).

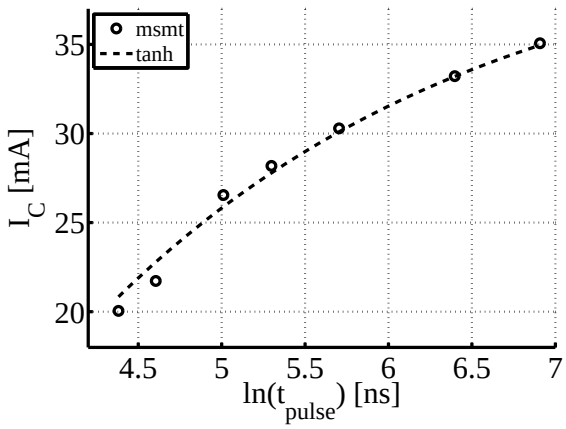


Figure 4.165: Forward output characteristic at fixed BE bias ($V_{BE} = 1.0V$) for variation of pulse width (t_{pulse}); *Device:* $w_E = 0.18\mu m$, $l_E = 5\mu m$

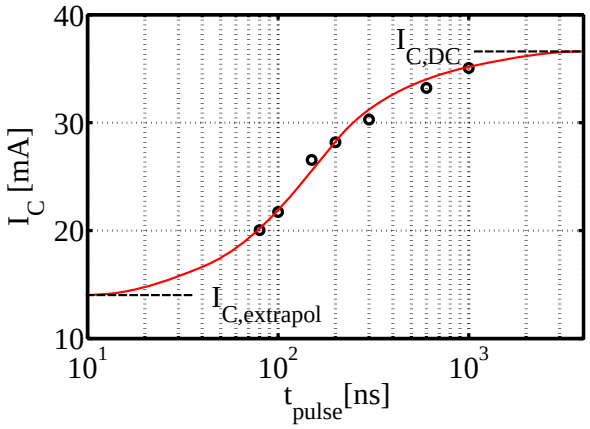


Figure 4.166: Hyperbolic tangent fitting and extrapolation at fixed BE bias ($V_{BE} = 1.0V$) for pulse variation; *Device:* $w_E = 0.18\mu m$, $l_E = 5\mu m$

Analyzing the RF characteristics under identical V_{CE} bias conditions one can clearly see the influence of self heating on the DC operating point under high current operation. The self heating has next to no impact under low current (cf. Fig. 4.167) whereas towards high current ($V_{CE} = 1.5V$) a significant shift of the operating point and little decay of the maximum transit frequency ($f_{T,peak}$) is seen (cf. Fig. 4.167).

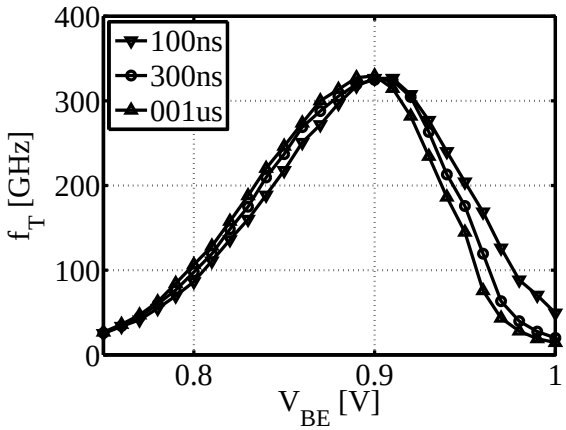


Figure 4.167: Pulsed S-Parameter measurement of transit frequency (f_T) as function of BE bias for constant $V_{CE} = 0.5V$; *Device:* $w_E = 0.18\mu m$, $l_E = 5\mu m$

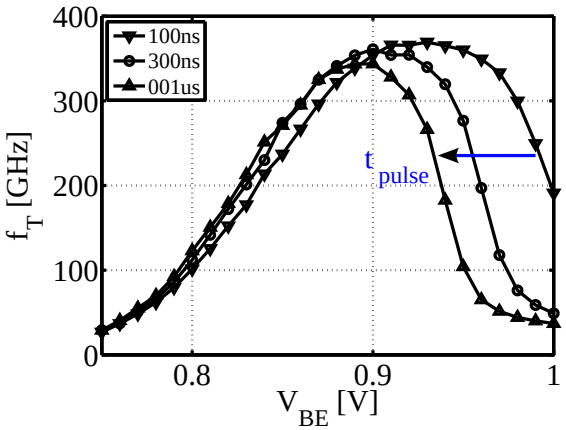


Figure 4.168: Pulsed S-Parameter measurement of transit frequency (f_T) as function of BE bias for constant $V_{CE} = 1.5V$; *Device:* $w_E = 0.18\mu m$, $l_E = 5\mu m$

The main problems for general application in the standard extraction procedure can be summarized as follows:

- Custom designed and cost intensive measurement equipment with synchronized DC and RF measurement solutions

- Measurement effort is highly increased in order to obtain reliable data (wrt. data verification and automated measurement)
- Large or multi-finger devices are more difficult to measure due to a significant V_C drop over series resistances for large I_C, requiring a higher delay time before DC and RF measurement is started (increasing self-heating)
- Re-simulation of device characteristics requires proper DC deembedding: transient simulations of dynamic behavior of external elements in the measurement setup are mandatory yet time-consuming

Due to this high complexity pulsed measurement setups are more suitable for experimental use. If the self heating impact is to be assessed from characteristics other than the classical base-current based DC methods one has the alternative of measurement with a RF network analyzer. The parameter Z_{TH} of the thermal network might be obtained using low frequency S-parameters measurements in a frequency range from 300kHz to 3GHz [127].

4.9 Transit Time Complex (τ_f)

The cutoff frequency (f_T) and the maximum oscillation frequency (f_{max}) are key characteristics of RF devices. For correct assessment of these FoMs, one of the challenges in RF modeling thus is the dependence of the transconductance (g_m) and f_T as functions of the transfer current. As seen in the DC current gain characteristic , the transconductance develops a distinct maximum at a defined bias before mobility degradation lowers g_m at high fields. Several additional effects come into play for the description of transit time complex. In order to reach highest accuracy, the extraction of parameters often-times requires iterative loops due to the strong link of AC and DC characteristics in the HICUM transistor model.

In the HICUM model bias dependence of the forward transit time (T_F) is governed by a total of 14 model parameters. This number shows the high degree of flexibility attributed to the modeling of dynamic characteristics, yet increases the model complexity and extraction effort. The parameter set can be sub-divided in two fundamental groups serving different purposes:

The first set of parameters is used under low and medium injection describing the bias dependence. Here the transit time (T_0) is used for modeling of the low current forward transit time at $V_{BC} = 0$. The additional parameters D_{T0H} and T_{BVL} account for the bias dependence of the low current transit time for varying V_{BC} bias. Delay times originating from charge storage in the neutral emitter are described by T_{EF0} and G_{TFE}.

Under high injection (around $f_{T,peak}$) the critical current $I_{CK}(V_{CE})$ is important for accurate modeling. The intrinsic collector resistance (R_{CI0}) as well as current spreading (L_{AT}) and characteristic voltages (V_{LIM}, V_{CES}, V_{PT}) describe the bias dependence of small signal characteristics around and above the minimum transit time. Being one of the major differentiators of the HICUM model formulation compared to less accurate compact models, these high-current parameters are critical for model accuracy. Lastly the base and collector region related parameters T_{HCS}, A_{HC} and F_{THC} allow for modeling of increasing transit time at high currents.

In the HICUM model, the various transit time contributions are modeled using their respective E and C related minority charge. Additionally the weighting factors attributed to emitter and collector (H_{FE} and H_{FC}) give the flexibility to adapt the equations to HBT physics.

$$T_F = H_{F0} \cdot T_{Flow} + \Delta T_{FB} + H_{FE} \cdot \Delta T_{FE} + H_{FC} \cdot \Delta T_{FC} \text{ , with} \tag{4.126}$$

$$\Delta T_{FB} = (1 - F_{THC}) \cdot \Delta T_{FH} \text{ and } \Delta T_{FC} = (F_{THC}) \cdot \Delta T_{FH} \tag{4.127}$$

4.9.1 Parameter Extraction Under Low Injection

Modeling the transit time is not directly possible since it presents an internal device quantity. However the transit time is the major factor of impact on the total transit frequency of the de-

vice (approximation: $f_T(I_C,V_{BC}) \simeq [2\pi T_F(I_C,V_{BC})]^{-1}$), whih can be measured. The analysis of measurement data of f_T for a sweep of the transfer current (I_C respectively) hence presents the best-suited approach for parameter extraction (cf. Fig. 4.172).

The fundamental governing equation is split into contributions originating from the low current transit-time (T_{Flow}), the neutral emitter transit time (ΔT_{FE}) and the contribution towards high current (ΔT_{FH}) related to charge storage in the base and collector region.

$$T_{Flow} = T_0 + D_{T0H} \cdot \left(\frac{C_{jCi0}}{C_{jCi}} - 1 \right) + T_{BVL} \cdot \left(\frac{C_{jCi}}{C_{jCi0}} - 1 \right) \tag{4.128}$$

$$\Delta T_{FE} = T_{EF0} \cdot \left(\frac{I_{TF}}{I_{CK}} \right)^{G_{TFE}} \tag{4.129}$$

$$\Delta T_{FH} = T_{HCS} \cdot w_i^2 \cdot \left[1 + \frac{2}{\frac{I_{TF}}{I_{CK}} \cdot \sqrt{\left[1 - \frac{I_{CK}}{I_{TF}}\right]^2 + A_{HC}}} \right]^{G_{TFE}} \quad \text{,with} \tag{4.130}$$

$$w_i = \frac{1 - \frac{I_{CK}}{I_{TF}} + \sqrt{\left[1 - \frac{I_{CK}}{I_{TF}}\right]^2 + A_{HC}}}{1 + \sqrt{1 + A_{HC}}} \text{ injection width} \tag{4.131}$$

$$T_F = T_{Flow} + \Delta T_{FE} + \Delta T_{FH} \tag{4.132}$$

The total transit time (T_F) of carriers traversing the vertical device profile is then defined as the sum of all involved elements (cf. eqn. (4.132)) and can be schematically represented as function of device bias as shown in Fig. 4.169.

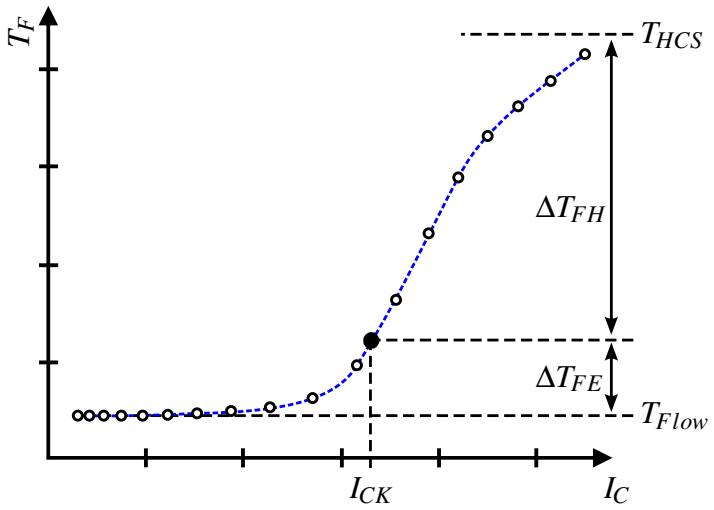


Figure 4.169: Schematic diagram of various contribution to the transit time in forward operation for a single geometry and bias

The most significant and characteristic data used for RF modeling is the forward transit time τ_f or T_F. This quantity is deduced from the extrapolated transit frequency ($\tau_f \propto 1/f_T$) as a function of the transistor bias (V_{BE} and transfer current I_C at various V_{BC}, network NW12). Electrical data is obtained from S-parameter measurement in the range of f_{meas} = [0.1 .. 110GHz]. The respective relation to measured S-parameter matrix is made using an extrapolation (cf. eqn. (C.3)). Usually one defines a single, so called spot frequency ($f_{meas} = f_{spot}$), allowing for easy treatment of the measured data. As best practice usually a medium frequency of about $f_{spot} \approx 20GHz$ is taken.

As in previous steps, there are two sets of model parameters to be determined. The first group is required for the proper modeling of bias dependence, using (mostly unitary) parameters related to the vertical device profile. For a fixed $V_{BC} = 0V$ bias condition, the BC junction can be disregarded and the principal model parameters can be determined. Those are T_0, the partitioning of the BE capacitance C_{BEpar} through adjustment of F_{BEpar} as well as A_{JEI} optimized under low through medium bias. A common approach to obtain the intrinsic transit time T_0 for each device is a linear regression of the total transit time plotted against the inverse transfer current as demonstrated in Fig. 4.170. In order to do so usually a single device is sufficient to extract the bias related model parameters. Afterwards the remaining parameters can be adjusted and the partitioning into areal ans perimetric contribution is done using different device geometries (cf. Fig. 4.171).

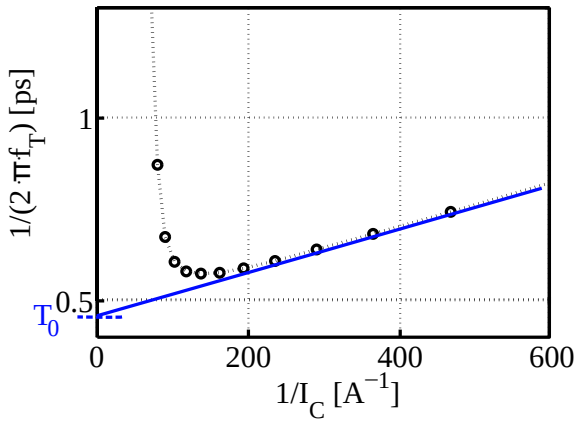


Figure 4.170: Transit time as as function of inverse collector current I_C for constant $V_{BC} = 0V$, linear extrapolation of low bias transit time T_0 (single device, $w_E = 0.2\mu m$, $l_E = 5\mu m$)

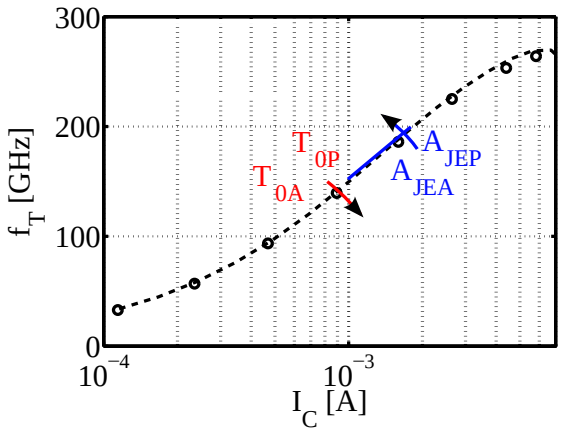


Figure 4.171: Transit frequency (f_T) as function of collector current I_C for constant $V_{BC} = 0V$ (single device, $w_E = 0.2\mu m$, $l_E = 5\mu m$)

The minimum value of the transit time T_0 is reached at high forward bias of the BE junction. Therefore one has to recall the general model equation for the junction capacitance (cf. sect. 3.3). The corresponding depletion charge is then obtained through integration of C_j. Once the low current transit time T_0 is known, the additional control parameter for capacitance limitation at forward bias (a_j) therefore is another important parameter for fine adjustment. The BE junction

capacitance related A_{JEI} can be altered from its default value (meaningful in a range of $A_{JEI} = [1.5 .. 2.5]$) to increase the model accuracy.

Along with the extrapolation for each geometry individually, the correct geometry dependence of the low bias transit time parameter T_0 can be determined. A geometry-related spread versus width and length respectively is assumed for constant bias (e.g. zero BC bias: $V_{BE} = 0$) as shown in Fig. 4.172.

Due to several physical effects, the device characteristics are specific for various device layouts. For small emitter structures (e.g. $l_{E,min}$) the parasitic external elements (especially capacitances) represent a significant contribution to the total values and limit device speed. For devices with large emitter area (e.g. $w_{E,max}$), intrinsic parts become large compared to the external parasitic elements (C_{par}). Thus devices with large total lengths usually have high cutoff frequencies at same number of fingers. However if the width is scaled, distributed effects under the emitter area come into play and non-ideal behavior decreases the device speed. Notably self-heating and distributed effects of devices with increased emitter area limit device performance. Thus geometry scaling of the transit time (τ_0) has to be taken into account in order to provide a fully predictable model over the full range of available process options.

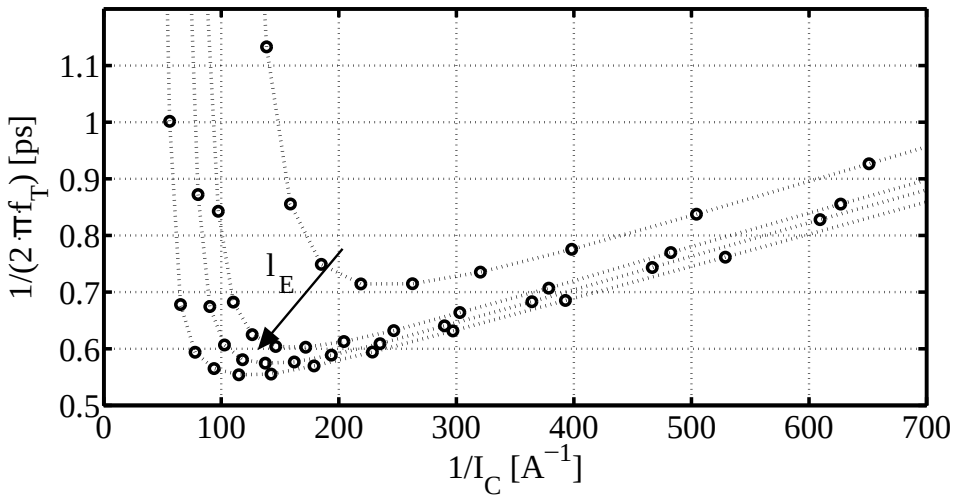


Figure 4.172: Geometry scaling of transit time τ_0 as extracted from f_T for a single, fixed emitter window width $w_E = w_{E,min}$

Under low injection the scaling equation, as presented in section 3.3.10, covers this through partitioning in surfasic and peripheral contribution by separation in the two parameter T_{0a} and T_{0p}, scaled with the effective emitter area A_{Eeff}. Multiple geometries are analyzed in order to find the correct partitioning of T_{0a} and T_{0p} as shown in Fig. 4.173 and 4.174.

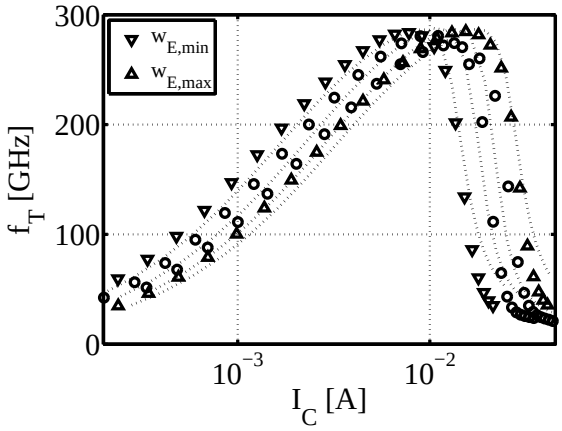


Figure 4.173: Geometry scaling of transit frequency f_T for a single, fixed emitter window length ($l_E = const.$)

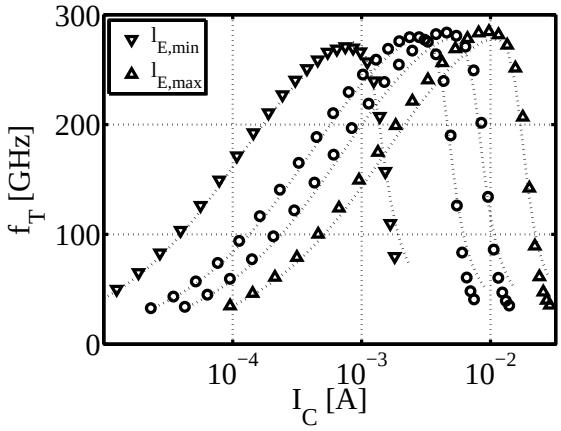


Figure 4.174: Geometry scaling of transit frequency f_T for a single, fixed emitter window width ($w_E = w_{E,min} = const.$)

For $V_{BC} \neq 0$ (esp. $V_{BC} > 0$) the additional collector charging times come into play and the parameters D_{T0H} and T_{BVL} can be adjusted. Given that the bias dependence is linked to the vertical profile rather than lateral device dimensions, adjustments for a single device geometry are sufficient while a second geometry might be used for verification purposes (cf. Fig. 4.175 and 4.176).

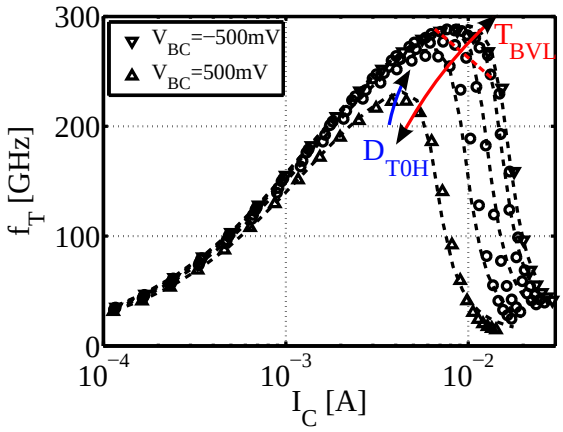


Figure 4.175: Transit frequency (f_T) as function of collector current I_C for different V_{BC} ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

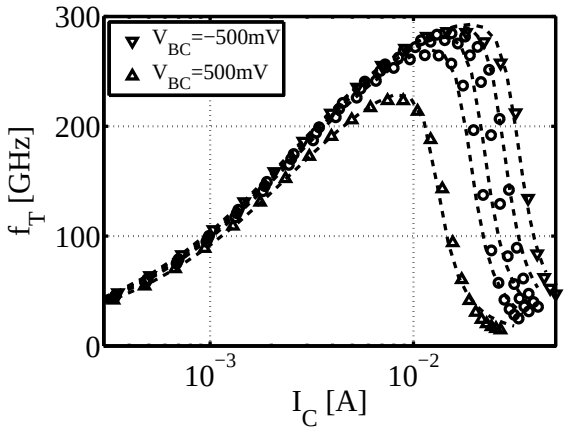


Figure 4.176: Transit frequency (f_T) as function of collector current I_C for different V_{BC} ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

In a last step regarding the transit time for modeling of the f_T characteristics, the measured data at constant CE bias (V_{CE}) is analyzed for correct modeling of the saturation region and the reverse transit time (τ_R, model parameter T_R). The reverse transit time (τ_r) accounts for the charge storage in the forward-biased base-collector junction. In analogy to the substrate current (for DC operation) this parameter is important for the simulation of the transient behavior of bipolar circuits

operated in saturation. Being a geometry-independent parameter related to the vertical device profile, τ_r can be determined using a single device geometry at very low, constant V_{CE} bias (cf. Fig. 4.177). Measured data from a single geometry provides sufficient information yet a second device dimension might be used to confirm the accuracy of the model (cf. Fig. 4.178).

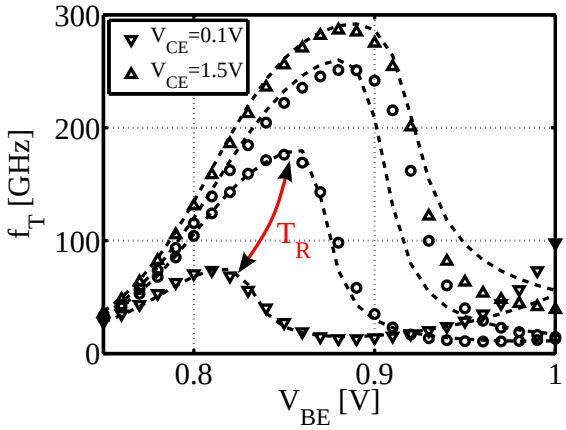


Figure 4.177: Transit frequency (f_T) as function of BE bias voltage V_{BE} for constant V_{CE} ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

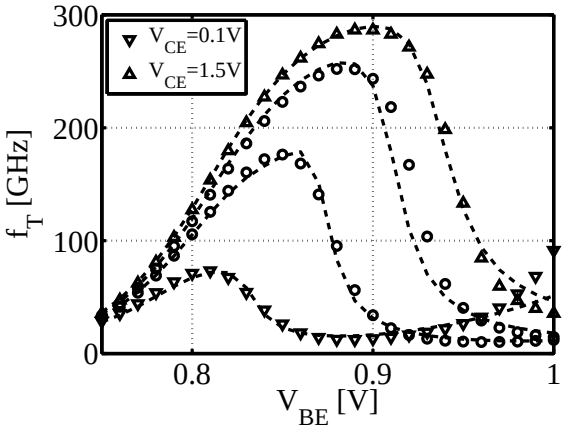


Figure 4.178: Transit frequency (f_T) as function of BE bias voltage V_{BE} for constant V_{CE} ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

For a final model verification of the low bias region a plot of all geometries (cf. Fig. 4.179) shows the progression of the f_T vs. I_C curves for different devices. However the scaling behavior of RF characteristics is very complex and not easy to predict. Additional parameters exist, that need to be adjusted under high current operation.

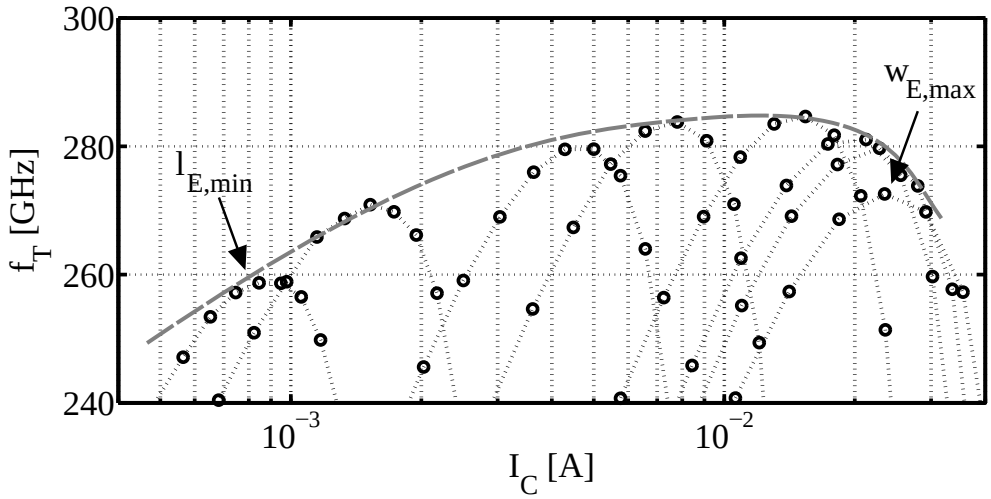


Figure 4.179: Detailed view of progression of peak f_T value for different devices indicating effects of geometry scaling on the total transit time τ_0, model verification for single BC bias

4.9.2 High Injection and Critical Current I_{CK}

Under high injection, another set of parameters is essential for description of the bias range beyond $f_{T,peak}$. Those are T_{EF0}, G_{TFE}, T_{HCS} and A_{HC} together with the set of model parameters describing the critical current I_{CK}: R_{CI0} (R_{KCI}), V_{LIM}, V_{CES} and V_{PT}.

Most of the parameters are defined as geometry-independent, yet the collector low-field resistance r_{Ci} covers many effects taking place in the epitaxial layer at high current densities. Therefore the geometry-scalable parameter R_{KCI0}, representing the surfasic resistance of the intrinsic collector, has to be extracted, taking into account its dependence on the device layout and configuration.

The concept of the critical current is used in HICUM for proper modeling of the high current region (I_{CK}). This boundary defines the current for the onset of high-injection related degradation of device performance (e.g. Kirk effect). The function depends on several parameters as shown in eqn. (4.133), where R_{CI0} represents the inner collector resistivity, scaling with device geometry.

$$I_{CK} = \frac{V_{Ceff}}{R_{Ci0}} \frac{1}{\sqrt{1+\left[\frac{V_{Ceff}}{V_{lim}}\right]^2}}. \tag{4.133}$$

A comprehensive extraction procedure for the high current related transit time parameters of the HICUM bipolar compact model was presented in [108]. It is understood that one of the fundamental steps is the assessment of the critical current I_{CK}. Taking the presented approach one step further, it is desirable to have a geometry-scalable approach without the requirement for iteration loops. The basic idea is that for a number of characteristics of ΔT, drawn versus normalized collector current (I_C/I_{CK}), one can find one value of I_{CK} that allows to superimpose ΔT for all measured biases values of V_{CB} in the Gummel characteristic.

For parameter extraction in the high current range ($I_C \geq I_{CK}$) one employs the concept of a differential transit time ΔT. With the minimum transit time reached at $I_C(f_{T,peak})$ one can define this characteristic transit time as reference for inferior current densities, allowing to calculate the additional forward delay time ΔT_f. In forward Gummel bias condition (NW12) for each set of curves (V_{BC} bias), T_{f0} is defined as the minimum of t_f (cf. Fig. 4.172).

Therefore the transit time and the respective f_T roll-off towards high forward bias is represented by the delay time ΔT defined as

$$T_F = T_{F0} + \Delta T_F, \tag{4.134}$$

where T_{F0} represents the low current transit time and ΔT_F represents the bias dependent rise in

transit time at medium and high current, which can be further sub-divided in the two contributions

$$T_{F0} = T_0 + D_{T0H} \cdot (c-1) + T_{BVL} \cdot \left[\frac{1}{c} - 1\right] \tag{4.135}$$

$$\Delta T_F = T_{EF0} \left[\frac{I_{TF}}{I_{CK}}\right]^{G_{TFE}} + T_{HCS} \cdot w^2 \left[1 + \frac{2 \cdot I_{CK}}{I_{TF}\sqrt{\left[1 - \frac{I_{CK}}{I_{TF}}\right]^2 + A_{LHC}}}\right], \tag{4.136}$$

Parameter extraction is done in a straight forward scheme (cf. Fig. 4.190). For each measured V_{BC} curve, a T_{F0} is defined (extracted) as the extrapolated minimum of the measured forward transit time (inverse transit frequency slope from f_T curves at low-medium injection). One can then define a threshold value for $\Delta T_{F,xtr}$ at which one extrapolates I_{CK} (usually $\Delta T_F = k \cdot T_{F0}$, with $k \approx 2$) for each bias value (cf. Fig. 4.180).

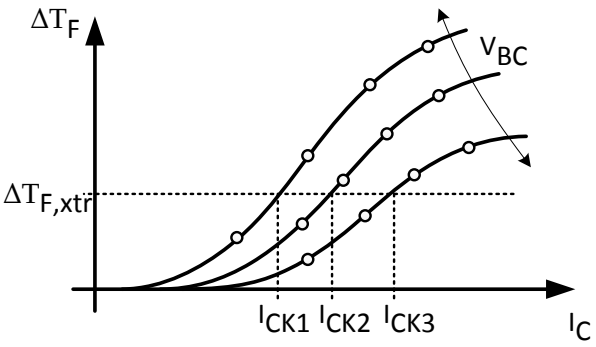


Figure 4.180: Schematic extrapolation of I_{CK} for different V_{BC} bias values of a single geometry

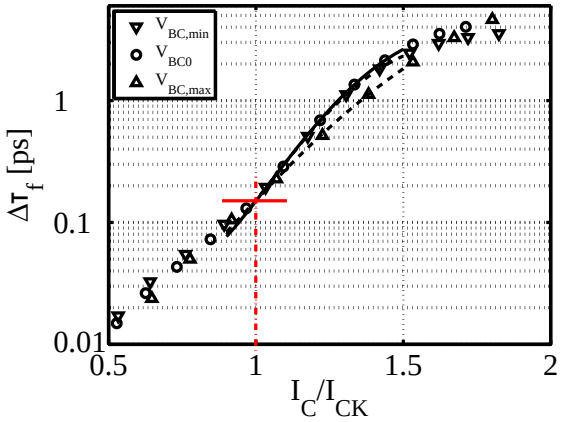


Figure 4.181: Superimposed ΔT_F curves after determination of $I_{CK}(V_{BC})$ for single geometry ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

Using the superimposed characteristics as function of the normalized collector current (I_C/I_{CK}) one can then perform numerical optimization of the related set of model parameters $T_{EF0}, G_{TFE}, T_{HCS}$ and A_{HC} (cf. Fig. 4.181).

The procedure of curve alignment can be automated using one fixed transit time offset (ΔT_F) for all devices and the repeatably applied to a selection of different geometries. Due to the steep fall-off at high current it is difficult to give a general recommendation that fits all use cases. A common value for $\Delta T_{F,xtr}$ is at half the peak transit frequency (50% $f_{T,peak}$) or $\Delta T_{F,xtr} \approx 2 \cdot T_{F,min}$.

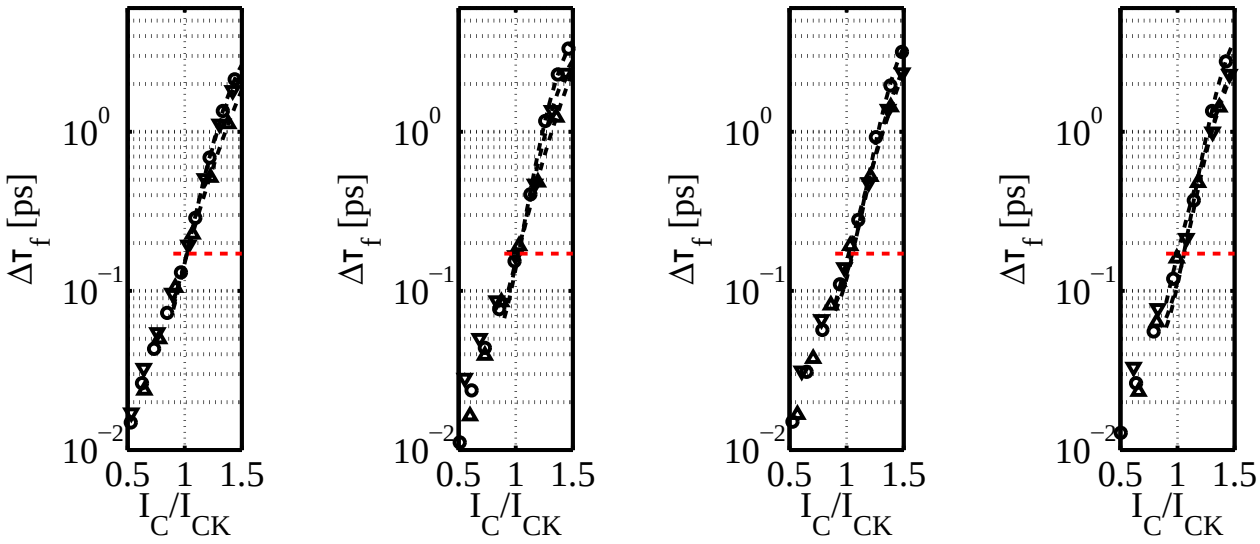


Figure 4.182: Extraction of high current parameters, superposition with normalized collector current for common threshold value $\Delta T_{F,xtr}$ (horizontal dashed line) for different device widths w_E at constant length $l_E = 5\mu m$ with the critical current I_{CK}

As shown in Fig. 4.182, one obtains one series of critical current values per device geometry as a function of collector bias. This result may then be normalized to obtain a common critical current density, summarized in one plot for all geometries as shown in Fig. 4.183. Given the apparent deviation of the current density per device geometry, one can then attempt to model the geometry dependence under high injection. The linear collector current spreading factor (f_{cs}, cf. eqn. (3.184)) used to correct for the three-dimensional current distribution in the epi-layer, is taking into account the related current spreading angle (γ_C). The spreading of the collector current is then expressed in the form

$$I_{CK,3D} = I_{CK} \cdot f_{cs} \propto r_{Ci0} \cdot f_{cs}. \tag{4.137}$$

The actual collector voltage (V_{CEi}) and the effective voltage across the collector region (V_{ceff}) used in the extraction are defined

$$V_{CEi} = V_{BC} + V_{BE} - ([R_{Cx} + R_E] \cdot I_{CK}) \text{ and} \tag{4.138}$$

$$V_{Ceff} = V_{DCi} - V_{BC}. \tag{4.139}$$

In order to see clearly which of the involved parameters influences what specific part of the critical current, several simulation studies were carried out (cf. Fig. 4.185 through 4.187). A general rule for high injection parameters is given by the physical meaning of the characteristic voltages: $V_{PT} > V_{LIM} > V_{CES}$. Starting from the smallest contribution one can set the collector-emitter satu-

ration voltage (V_{CES}) to rather small value (at less than 100mV) for standard RF technologies. The three characteristic threshold voltages are considered as unitary parameters and one common average value for each voltage is used for all geometries leaving R_{Ci} as the only geometry dependent value in the formulation of the critical current (cf. eqn. (4.133)).

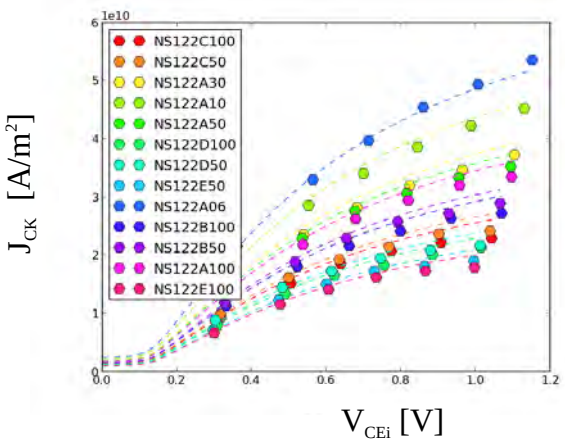


Figure 4.183: Scaling of critical current density for multiple different geometries, *dashed line: optimization of model parameters*

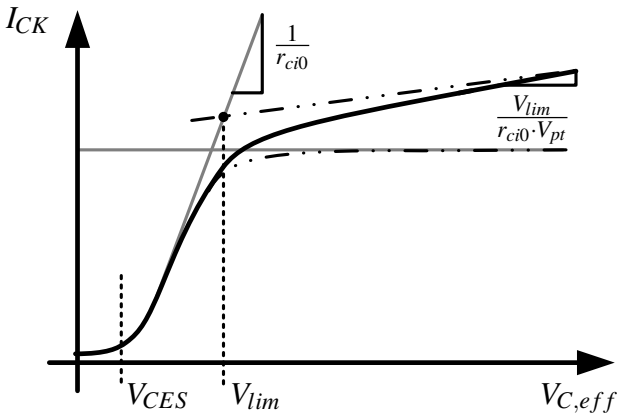


Figure 4.184: Influence of model parameters in the definition of the critical current (I_{CK}) as the onset of high-current effects versus CE bias

Given the values of I_{CK} as function of BC bias one can then use intelligent least square fitting to optimize the model equation (cf. eqn. (2.52)). From the curve slope at low V_{CB} bias ranges the resistivity R_{Ci} as well as the onset voltage V_{CES} is obtained, whereas towards high bias the remaining voltages V_{lim} and V_{PT} are optimized to fit the curve to the data (cf. Fig. 4.184). The newly introduced parameter δ_{CK} (from HICUM L2.3x) is kept at its default value due to the already satisfactory agreement of extracted parameters and model equation.

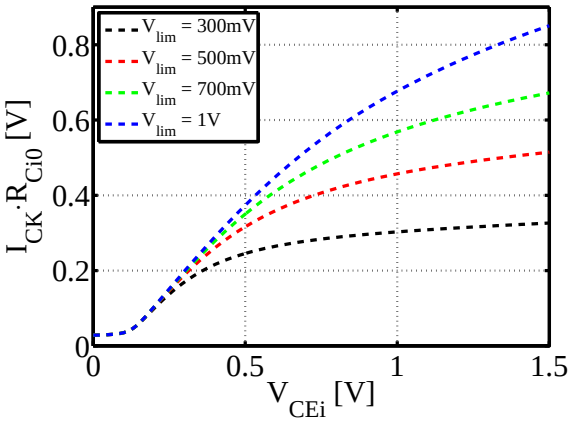


Figure 4.185: Variation of the critical current for different parameter values of V_{lim}

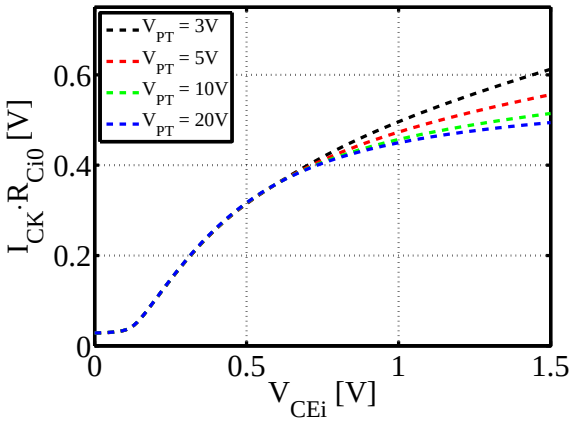


Figure 4.186: Variation of the critical current for different parameter values of V_{PT}

Thus the last step is the extraction of the geometry dependence under high injection. Here the total collector resistance is the key parameter due to its strong impact on the calculation of the transit time. The scalable model library uses a surfasic quantity per unit of epi-layer area, which scales with the collector area and the current spreading in the vertical collector (cf. eqn. (3.185)). The three-dimensional collector current spreading in intrinsic transistor, is taken into account through the model parameters L_{ATB} and L_{ATL} for spreading of the injection zone w_i in the respective spatial direction (width and length related). These parameters model a bias dependent collector current spreading (depending on device dimensions, cf. Fig. 4.188) by means of the following equation (giving the current spreading angle δ_C)

$$L_{ATB} = \frac{2w_c}{w_E}\tan\delta_C \quad \text{and} \quad L_{ATL} = \frac{2w_c}{l_E}\tan\delta_C. \tag{4.140}$$

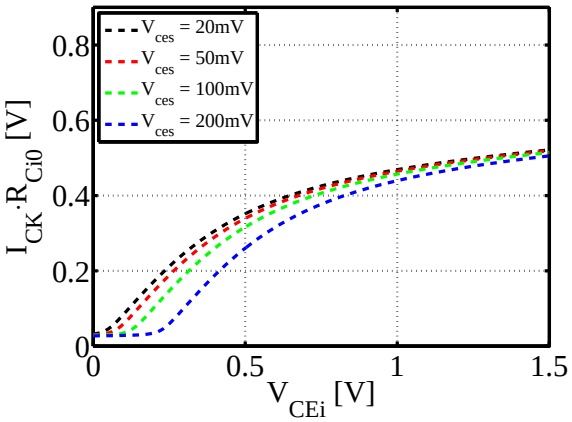


Figure 4.187: Variation of the critical current for different parameter values of V_{CES}

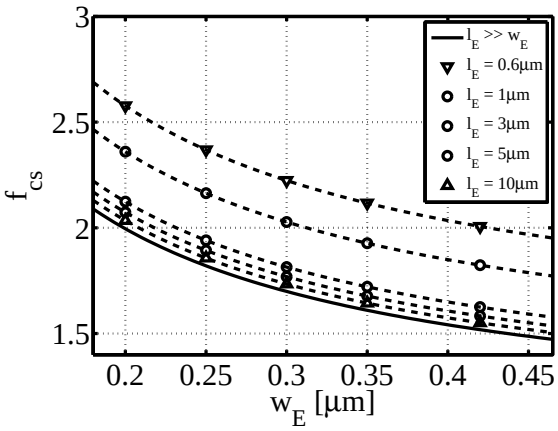


Figure 4.188: Scaling of current spreading factor f_{cs} for different geometries

As seen in the model formulation (cf. eqn. (4.133)), there exists a linear relation of the extracted current and the collector resistance R_{Ci0} ($I_{CK} \propto R_{Ci0}$). Given that the current spreading through the factor f_{cs} is likewise incorporated in this relation, the related parameters can be extracted by a curve optimization of the scaling equations (cf. sect. 3.3.10) to the extracted resistance values per geometry as shown in Fig. 4.189. The complete extraction flow for high-current related parameter with the individual extraction steps can be summarized in the schematic flow chart as seen in Fig. 4.190.

Eventually the quality and accuracy of the determined set of parameters is best seen in a plot, displaying the full scaling range for an important FoM such as f_T as shown in Fig. 4.179.

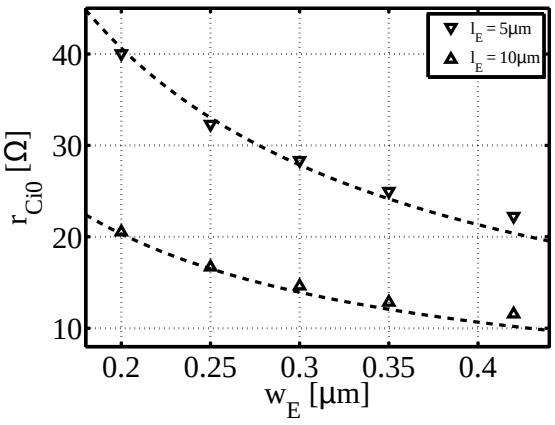


Figure 4.189: Optimization of current spreading parameters L_{AT} for extraction from r_{Ci0}

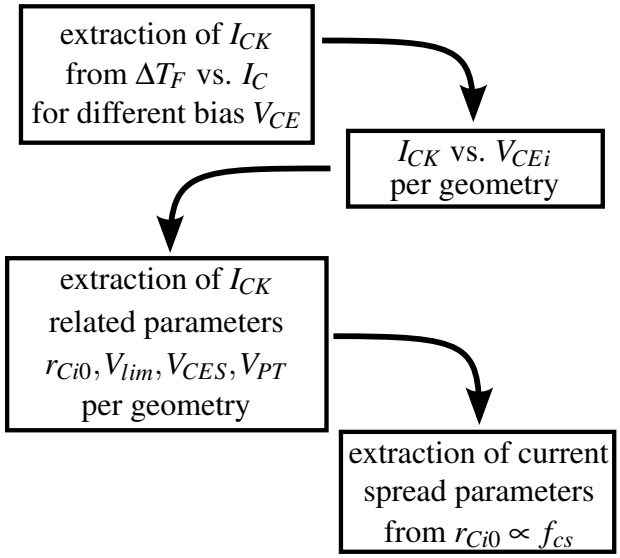


Figure 4.190: Schematic parameter extraction flow for high injection related parameter set

4.9.3 Polynomial Fit of the f_T vs. I_C Characteristic for Optimization

As for other parameter sets a direct extraction approach often-times is followed by a numerical optimization for re-adjustment and fine-tuning of parameters. Therefore, it is very important to have the capability to optimize the RF parameters using the most important RF characteristic, and consequently to simulate f_T versus I_C at different V_{BC} biases.

However as pointed out before (cf. sect. 2.2.4), HICUM employs iterative loops to solve the interrelation of AC and DC relation. Thus the formulation of the forward transfer current I_{TF} has an implicit formulation and in itself is influenced by the transit time through the hole charge Q_{pT}

$$I_{TF} = \frac{C_{10}\exp\left(\frac{V_{BE}}{V_T}\right)}{Q_{p0}+Q_j+Q_{pT}} \text{ with } Q_{pT} = \int_0^{I_{TF}} T_F \, di. \tag{4.141}$$

For application of any numerical data optimization algorithm, one needs optimization targets as well as simulated data to compare to. Usually the simulated characteristics with matching bias conditions are generated at the exact same bias points (terminal voltages) as the measured data and optimization is rather easy. A particular case for RF bipolar modeling is the f_T characteristic as a function of the transfer current I_C. It is of high importance that $I_{Csim,i}$ and $I_{Cmeas,i}$ are identical, otherwise data cannot be compared. Yet the simulated I_C in itself is a function of a multitude of model parameters and matching bias points are virtually impossible to achieve.

A polynomial fitting approach thus presents a workaround to re-generate arbitrary pairs of f_T, I_C data from a limited number of available measured data. A piecewise defined higher-order

polynomial function was found to give the desired flexibility for modeling. Thus at each bias point, the measured f_T is approximated with a third order polynomial equation of the following form

$$f_T = a_0 + a_1 \cdot I_C + a_2 \cdot I_C^2 + a_3 \cdot I_C^3. \tag{4.142}$$

The measured data is piecewise analyzed using four adjacent points (x_{-1} through x_2) for each individual curve section to be fitted (cf. Fig. 4.191).

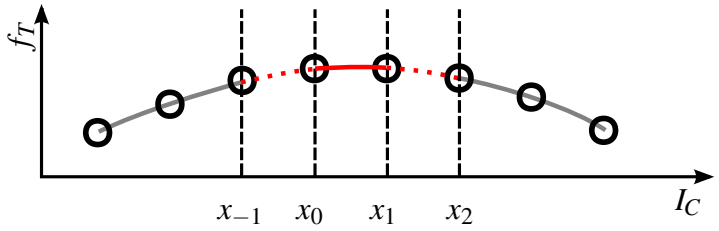


Figure 4.191: Polynomial curve fit of f_T vs. I_C characteristic with piecewise determination of curve segments

A common polynomial fitting routine available in optimization packages of object oriented programming languages is used to determine the polynomial coefficients. The resulting parameter matrix [**A**] with parameters a_0 through a_3 is then stored in an array for each bias point.

For re-generation of the required data, one simulates the f_T, I_C characteristic in a first step. In a second step the discrete simulated I_C matrix is passed to a dedicated routine that makes use of the determined [**A**] matrix and re-generates a 'pseudo'-measured pair of f_T, I_C data suitable for parameter optimization (cf. Fig. 4.192). This process might be implemented in optimization loops afterwards for automated parameter determination.

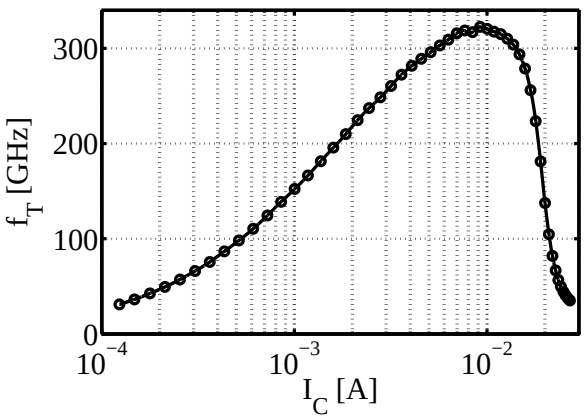


Figure 4.192: Polynomial curve fit of forward transit frequency f_T curve as function of transfer current I_C at V_{BC0}

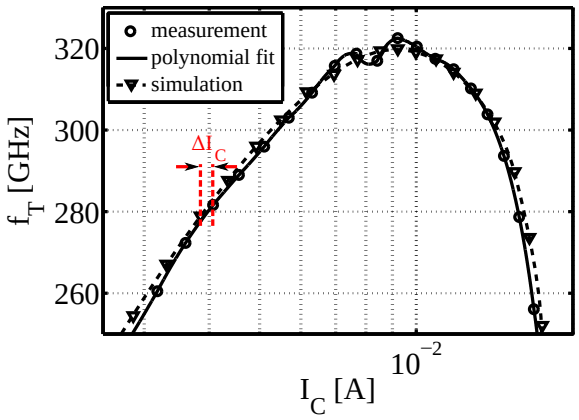


Figure 4.193: Detailed view of peak f_T region with segments of polynomial curve fit, *different discrete points of simulated and measured collector current (ΔI_C)*

4.9.4 Maximum Frequency of Oscillation f_{max}

Another RF characteristic that is important for circuit design is the maximum frequency of oscillation (f_{max}). The various series access resistances are an important parameter for the correct modeling of f_{max} characteristics. Especially the extrinsic base resistance R_{Bx} is having a significant influence on the maximum value ($f_{max,peak}$).

$$f_{max} = \sqrt{\frac{f_T}{8\pi R_B C_{BC}}} \text{ [128]} \tag{4.143}$$

Even though R_{Bx} is accurately determined in a direct extraction adjustments to its value might be necessary. Additionally, the correct scaling with device geometry can be adjusted with the highly sensitive capacitance C_{BC} through the partitioning factor (F_{BCpar}).

As for the transit frequency the bias dependence of f_{max} is to be regarded individually for selected geometries. In addition to the link resistance of the base, the second model parameter to be adjusted is the BC partitioning factor F_{BCpar}.

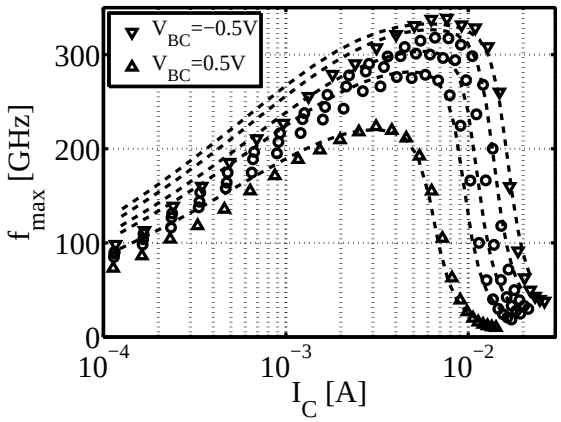


Figure 4.194: BC junction bias scaling of f_{max} over a range of V_{BE} bias voltage for single geometry ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

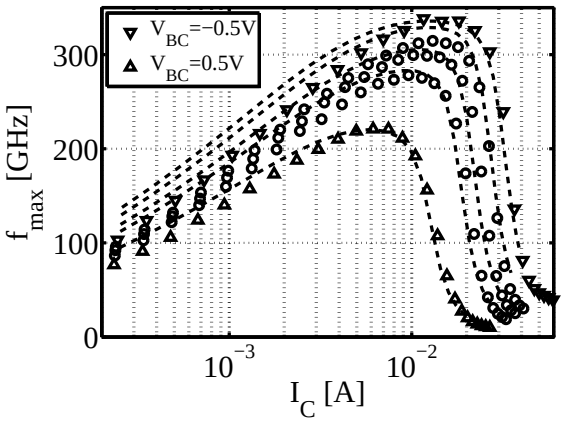


Figure 4.195: BC junction bias scaling of f_{max} over a range of V_{BE} bias voltage for single geometry ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

At the same time the progression of f_{max} relative to a variation of the BC junction bias can be used as a verification for the correct determination of the parameters describing BC bias dependence of the transit time f_T.

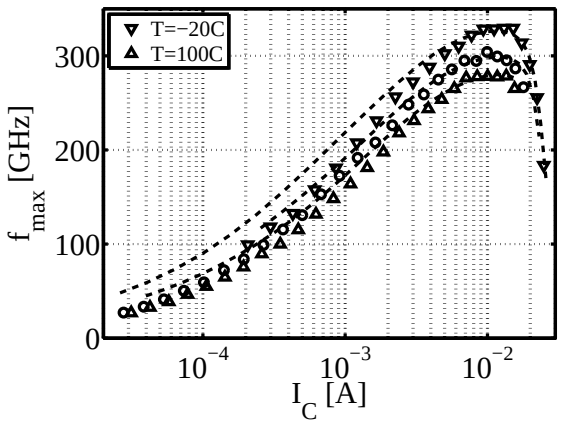


Figure 4.196: Temperature scaling of the maximum frequency of oscillation at fixed bias voltage $V_{BC} = 0V$ for single device ($w_E = 0.2\mu m$, $l_E = 5\mu m$)

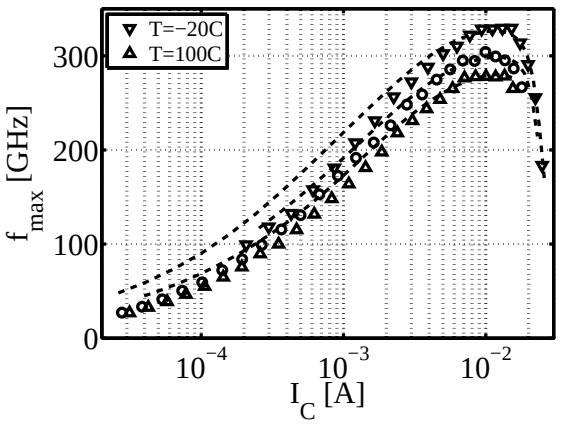


Figure 4.197: Temperature scaling of the maximum frequency of oscillation at fixed bias voltage $V_{BC} = 0V$ for single device ($w_E = 0.2\mu m$, $l_E = 10\mu m$)

4.10 Limitations to General Process Scalability

Some of the area-related quantities as the collector current density I_C, the capacities like C_{BE} or the transit frequency f_T exhibited a non-standard scaling behavior in recent technologies, differing from the general scaling behavior expected and predicted by scaling approaches. The reasons for these effects are not completely clear and topic of different research campaigns in different branches of device modeling [14, 129–131]. Different explanations exist, yet imaging analyses of recent technologies indicate a strong accumulation of the As impurities at the BE junction, limiting process precision towards smallest geometries. As a result for example a negative slope in the presented P/A method is observed which corresponds with the findings presented in studies of other process variants [67, 132, 133], showing that process scalability is a serious concern throughout modern technologies.

Most of the presented solution approaches presented so far compensate this problem thorugh the use of geometrical corrections for the respective lateral dimensions of the device. As demonstrated in section 3.1.2, the offset parameters incorporated in the scalable model library used in this work are represented through individual parameters dw_E and dl_E. However this simple solution using a linear offset may not provide satisfactory results for other process generations and non-linear scaling as well as process improvements need to be studied.

4.11 Model Validation

In a final step, a verification if the extracted model corresponds with the behaviour of real devices shows the reproducibility of measured characteristic through circuit simulation. The verification of the extracted compact model is accomplished by comparison under various scenarios and measured electrical characteristics. Thus model verification is performed covering the full range of variable operating conditions (e.g.: bias, frequency, temperature as well as geometry). Being most important for mixed-signal analog design with HBT devices, the accurate reproduction of DC characteristics as well as RF performance around and beyond the critical current (I_{CK}) is verified. For geometry scalable models simultaneously additional transistors with different configurations and geometries are used in order to confirm the agreement of all transistors in the model library.

In case the model does not correspond to the measured characteristics the linked extraction step may need to be repeated by iterative re-optimization of the respective parameter values. The resulting iteration loops for model verification may be used to (fine-)tune parameters that have not been explicitly determined in any of the previous extraction procedures. In addition to agreement of model (simulation) and measured data the smoothness and continuity of simulations is tested to avoid convergence issues and glitches.

As a best practice the verification should include the most important characteristics of the transistor in both static (DC) and dynamic (RF) operation for the intended operation as summarized:

- DC characteristics:
  1. Forward Gummel plot at different V_{CE}
  2. Transconductance g_m
  3. Output conductance g_o
- RF characteristics:
  1. f_T and f_{max} vs. J_C at different V_{CE}
  2. Y-parameters at constant spot frequency (f_{spot}) versus J_C at different V_{CE}
  3. Y-parameters vs. frequency for different bias conditions (e.g. $V_{BE}(f_{t,peak})$)
- Temperature dependence:
  1. DC Forward Gummel plot at V_{BC0} at different T_{amb}
  2. f_T and f_{max} vs. J_C at different T_{amb}
- Noise (low and high frequencies) and NQS

4.12 Conclusion

In this chapter the various extraction methodologies and strategies for parameter determination have been described. Based on measured data obtained from on-wafer measurements, the application of existing as well as new routines and procedures to a recent industry-leading SiGeC HBT technology was demonstrated.

It has been shown, that the basis for reliable, physics based device modeling is the accurate determination of the external parasitics such as series resistances (R_E,R_B and R_C) as well as peripheral capacitances (C_{BEx}, C_{BEx}). Existing and newly developed extraction routines for parasitic determination have been successfully applied to measured data. The influence of layout variations related to the substrate as well as the device periphery have been studied.

Some of the proven existing extraction methodologies had lost their basis due to a recent update of the model formulation. New approaches for the highly important collector current parameter and weighting factors have been studied and a reliable method was found so the model parameter extraction flow is gap-less for parameter extraction under low current operation.

Under high-injection however the parameter extraction is not as straight forward as for low bias. Given the high number of unknowns and the strong interaction of parameters several extraction steps are based on experience and best-practices rather than direct extraction. However a systematic, geometry-scalable solution approach for the determination of the critical current (I_{CK}) has been successfully implemented and tested.

Given the ever increasing importance of the self-heating problematic due to smaller lateral device dimension and concentration of heat within the device, an extensive analysis was performed. An innovative approach to eliminate the influence of self heating using pulsed measurements with a new experimental test setup was explored and the general application to SiGe HBT technology was analyzed. Even though the obtained results were promising the complexity of the required measurement and transient simulation make it difficult to integrate this technique into the general extraction flow.

In total, all described extraction steps and strategies have been executed and a fully scalable set of model parameters was extracted. The company proprietary, custom scalable model library based on the geometry scaling equations described in chapter 3, has shown very good agreement with the actual silicon technology. Even though difficulties related to shrinking device dimensions were seen, simulated characteristics show a considerably good agreement of the general agreement of the and the silicon-based data throughout a large range of bias and ambient temperature.

Chapter 5

Conclusion

5.1 General Conclusion

With advancing processes, today's RF HBT devices are more and more approaching physical limits of silicon manufacturing technology and consequently the accurate modeling of their behaviour has become more complex. Compact model equations of HICUM, applicable to earlier process generations, have been shown to loose their validity as operation frequency increased, calling for increased flexibility in the model formulation.

After a brief overview of the device architecture and process technologies used in the manufacturing of modern SiGe BiCMOS HBT devices, the physical origin of the new compact model formulations of HICUM L2.3x, suitable for physics-based simulation of semiconductor devices operating in the terahertz regime, have been presented. By means of numerical device simulation, it was shown how the apparent effects are related to the increasingly narrow vertical device profile of modern RF devices with a steeper slope of doping and alloy gradients ($\Delta N/\Delta x$). Especially the relation to the SiGe base profile was analyzed with respect to shape and position of the Ge alloy. Analyzing the underlying physics, special attention was paid to the reverse Early effect, being very sensitive to both bias and temperature changes for advanced HBTs.

The applicability of the model has been verified for a industry-leading RF BiCMOS technology in a combined $55nm$ node. The suitability of the new model formulations was validated with good agreement of measured and simulated characteristics of RF bipolar devices with cut-off frequencies up to $320GHz$ (f_T) and $370GHz$ (f_{max}).

Being a very important part of successful device modeling, a significant portion of the work carried out in this thesis was devoted to the development and refinement of a number of extraction procedures for the HICUM compact model. Some of the approaches are based on newly proposed test structures, making effects distinctly accessible through direct measurements, while other routines were tailored to enable direct extraction of model parameters as well as extrinsic parasitics from existing structures.

In particular the precise, geometry-scalable extraction of the emitter resistance (R_E) by means of unilateralization and the related improvement of the model value determination has been a valuable addition to the existing parameter extraction flow. The collector series resistance extraction of sheet resistance contributions (R_{Cx} with R_{BL} and R_{SK}) was improved by means of an advanced set of test structures, allowing for verification under active transistor operation.

Another significant part devoted to the extraction of parasitic network elements in the equivalent circuit focused on the substrate network and its influence on device performance. Trials with newly designed test structures and variation of the substrate ring were layouted, measured and evaluated for a state-of-the-art trench isolated technology. Furthermore numerical device simulation studies and measurement campaigns dedicated to the extrinsic overlap capacitances of the BE and BC junction provided valuable insight and allowed to confirm simulation results.

In the critical domain of the low bias forward operation of HBTs, new concepts for parameter determination of transfer current (i_T) and charge related (Q_B) model parameters were elaborated. Therefore the parameter determination had to be adapted to the latest model formulation. A comparison of a fully geometry-scalable iterative approach was done with a direct extraction approach, based on mathematical transformation of the charge weighting factor formulation (h_{jei}).

For the important high current operation, a smart algorithm to support parameter optimization by means of polynomial curve fit was made available, in order to enable the direct optimization of the f_T versus I_C characteristic as an important part of RF device modeling. Focusing on the measured data for the same characteristic of the transit frequency (f_T vs. I_C) a geometry-scalable approach for critical current (I_{CK}) parameter determination was implemented. The obtained extraction results were shown to give reliable and physics-based results for parameter initialization in this very sensitive operating region.

In addition, special efforts focused on the investigation of common challenges in on-wafer measurement of fast RF devices operating at high current densities with small spatial dimensions. Self-heating related issues were addressed by pulsed RF and DC measurement using a novel measurement system, available in the IMS research laboratory. Limitations regarding general applicability and industry-related modeling of HBTs were shown.

Given the close collaboration of research laboratory and industrial partners, the work presented in this thesis is strongly application-oriented. The research covered novelty and innovative new concepts for advanced technologies while at the same a focus was on the evaluation and assessment of routines with regards to general applicability in an industrial framework and the implementation of procedures in widely used software tools. Extraction routines and procedures were made available in a ICCAP based framework, presenting a commercial EDA solution dedicated to advanced device modeling. A common, high-level open-source programming language (Python 2.7) was used to implement and test the most part of the presented extraction strategies.

Keeping in mind the specific needs and expectations related to device modeling tasks, as a daily routine for modeling engineers in industrial companies, this thesis is also intended to be a guideline for successful parameter extraction. Key results of the findings presented in this work as well as the proper documentation produced at the conclusion of each phase should help modeling engineers using a unified custom parameter extraction flow. The full extraction procedure has been implemented in a commercial software solution by sequencing the presented extraction steps. The individual procedures have been evaluated with respect to the selection of bias ranges as well as sequential automation of the flow before their implementation.

In general, the initial expectation to improve, evaluate and describe a full model parameter extraction flow, applicable to the latest technology evolution of HBT devices manufactured in BiCMOS technology, has been met. A significant improvement was achieved for the asset of parasitics, being of increasing importance for device modeling as lateral dimensions are reduced. Furthermore the model was validated with regards to DC and RF characteristics over a large range of temperatures. A basic extraction flow for the extraction of application critical high current RF parameters was presented and realized through software automation yet future work has to improve the robustness as well as general applicability.

5.2 Future Works

Recent simulation studies have shown that the theoretical limits of HBT device technology is not reached yet. With the ever advancing reduction of feature size in silicon technology, SiGe HBTs may be integrated in more advanced lithography nodes. Yet the continuous advancement of silicon device technology towards higher operating frequencies and smaller feature sizes will surely bring along new challenges for device modeling.

In the near future SiGe may move to even more advanced lithography nodes such as 28*nm* or being integrated into silicon on insulator (FD-SOI) processes, originally intended for high-speed low-voltage applications, currently presenting new approaches for high-volume production of digital circuits. In addition, scaling and shrinking of the nominal feature size of the devices will have a even more significant impact on the ratio of extrinsic device parasitics to the intrinsic device. The accurate assessment of model parameters through dedicated test structures and extraction routines will therefore continue to be an important field for future research.

After a review of realizability and technical feasibility, preliminary models may be generated, based on experiences from the current process generation as well as predictions from TCAD simulation studies. Once a new process generation becomes available, future work in the field of fully geometry scalable compact modeling for HBT transistors will primarily be focused on any unforeseen physical effects related to new transistor architectures. Later on simulation results of predictive models will have to be verified and adapted, once measurement data from fabricated silicon becomes available.

With regards to the model formulations implemented in HICUM, this work confirmed the good agreement of the model implementation with bias and temperature. Yet this evaluation has to be exercised for every new process generation. More importantly the custom geometry scaling equations have to be continuously re-evaluated for upcoming processes. As used in this work, TCAD simulation based campaigns may aid this process.

Other branches not covered in this work are the field of statistical modeling. Given the strong physical basis of the model equations used in the HICUM formulation, there is a high chance that model parameters are strongly correlated to variations in the silicon process. However these inter-relations can only be exploited using more extensive process information and measured data. Relations have to be found between process variations and their impact on electrical characteristics by means of process variations and split lots.

The topic of corner modeling is another field, strongly related to the evaluation of robustness and variation in a given process. The estimation of process deviation is important, in order to correctly estimate process windows, enabling more aggressive designs. Yet this measure to improve model accuracy requires extensive studies of lot-to-lot and wafer-to-wafer variation.

Appendix

A List of Published Work

[1] *Extraction Procedure for Emitter Series Resistance Contributions in SiGeC BiCMOS Technologies*
Stein, F. ; Huszka, Z. ; Derrier, N. ; Maneux, C. ; Celi, D.
International Conference on Microelectronic Test Structures (ICMTS), 2014, IEEE

[2] *Selected Topics in Bipolar Modeling and Measurement*
Stein, F. ; Derrier, N. ; Celi, D. ; Maneux, C.
BIPOLAR Arbeitskreis (AKB), 2013

[3] *Investigation of the base resistance contributions in SiGe HBT devices*
Stein, F. ; Celi, D. ; Maneux, C. ; Derrier, N. ; Chevalier, P.
International Semiconductor Conference (CAS), 2013, IEEE

[4] *Advanced Extraction Procedure for Parasitic Collector Series Resistance Contributions in High-Speed BiCMOS Technologies*
Stein, F. ; Derrier, N. ; Maneux, C. ; Celi, D.
Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2013, IEEE

[5] *Robustness of the Base Resistance Extraction Method for SiGe HBT Devices*
Stein, F. ; Celi, D. ; Maneux, C. ; Derrier, N. ; Chevalier, P.
International Semiconductor Conference Dresden Grenoble (ISCDG), 2013, IEEE

[6] *Base Resistance Contributions in SiGe HBT Devices*
Stein, F. ; Derrier, N. ; Celi, D. ; Maneux, C.
13th HICUM Workshop, 2013

[7] *Device modeling for advanced SiGe HBT bipolar technologies*
Stein, F. ; Derrier, N. ; Celi, D. ; Maneux, C.
ST-IMS Workshop, 2013

[8] *Advanced SiGe HBT Modeling with HICUM/L0 (v1.3) for RF and mmW Applications*
Celi, D. ; Derrier, N. ; Stein, F.
Compact Modeling for RF/Microwave Applications (CMRF), 2012, IEEE

[9] *Extraction of the emitter related space charge weighting factor parameters of HICUM L2.30 using the Lambert W function*
Stein, F. ; Huszka, Z. ; Derrier, N. ; Maneux, C. ; Celi, D.
Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2012, IEEE

[10] *Device modeling for SiGe bipolar technologies with aggressively scaled vertical profiles*
Stein, F. ; Derrier, N. ; Celi, D. ; Maneux, C.
ST-IMS Workshop, 2012

B SiGe HBT Process Technology

B.1 The Front-End Process Flow of a Double Polysilicon Self-Aligned (DPSA) Architecture with Selective Epitaxial Growth (SEG)

Silicon nowadays is the material of choice when it comes to semiconductor mass production. Even though it is not the ideal material from an electronic stand point the ease of processing as well as the presence of a good native oxide made it the backbone of the semiconductor industry. The typical wafer diameters for todays BiCMOS technologies range from 150mm all the way up to 300mm. Before the description of the integration flow in a CMOS environment the principal manufacturing steps of a pure bipolar process are presented.

Ever since the introduction of the BiCMOS6 technology in 0.35μm lithography node in 1998 over a decade has passed. In this time-frame the subsequent of technology and design allowed to reach more than a tenfold of operating frequency. In general lateral scaling allows reducing all parasitic capacitances as well as resistances while vertical profile improvements reduce transit times to help increasing the intrinsic device speed.

The summary table C.1 shows most of the existing approaches for advanced bipolar device fabrication used today. Generally bipolar processes are divided in concepts depending on the base architecture. The first aspect is the process alignment. So called self-aligned processes are independent of overlay accuracy of lithography. Hence the fully self aligned HBT process is a industry standard solution these days.

Further differentiation is done depending on the base deposition. Single-poly (SP) architectures have exclusively the emitter junction deposition realized with polysilicon. The base link is realized through implantation allowing for best compatibility with the traditional MOS flow. For the double-polysilicon (DP) technology in contrast polysilicon deposition methods for both base and emitter junction are used. A lateral base link connection in DP technology is realized with poly-silicon allowing the placement of the connection of the base contact directly over the field oxide and thus significantly reducing the peripheral base-collector capacitance [134]. The emitter-to-base isolation is realized by means of a spacer allowing the extrinsic base to be fabricated in a separate step. Further downscaling is hence possible and heavy doping (or salicidation) provides excellent contact resistances. Thus it is the preferred option for high speed applications.

A last differentiator of technologies is the use of a non-selective base epitaxy (NSEG) or selective epitaxial growth (SEG) of the base. The latter referring to technologies with epitaxial growth inly on Si surfaces and no deposition on surrounding oxide or nitride material. Given its process control advantage it is the better-tested architecture for mass production providing easier implementation into a fully self-aligned (FSA) design. Below is a detailed schema of the most important manufacturing steps required to fabricate the core elements of the SiGe HBT device structure (cf.

Fig. B.1). For the Si processes analyzed in this work the Front-End-of-Line (FEOL) starts with a lightly pre-doped bulk silicon wafer substrate (p-substrate for the shown process) [7, 62].

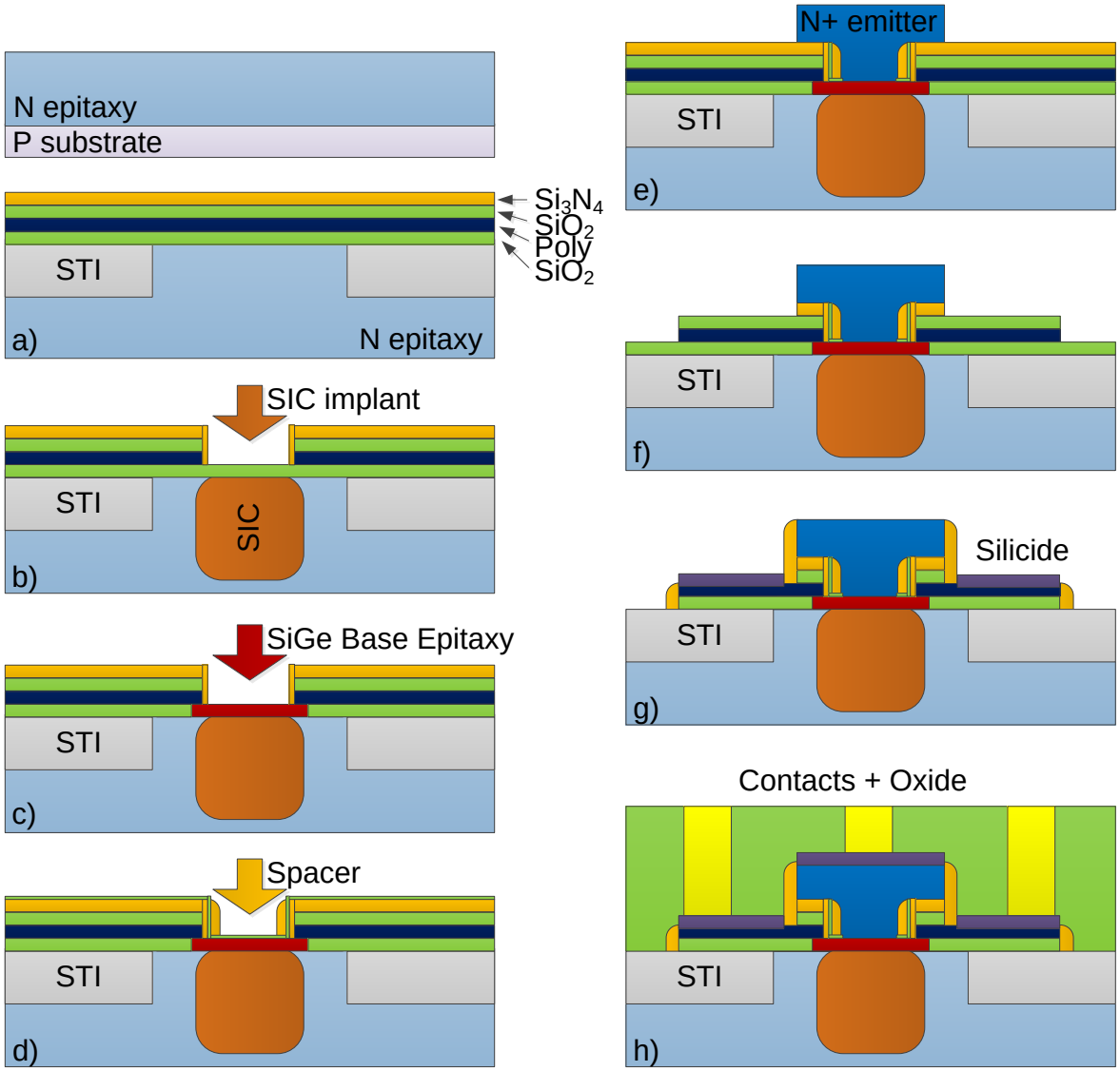


Figure B.1: HBT process flow of a DPSA-SEG architecture

The Collector Profile

Apart from the fundamental device architecture and the choice of process there is a multitude of process parameters that can be optimized in order to achieve desired characteristics. Most of them concentrate either on the reduction of external parasitic elements or adjustments to the vertical profile of the device.

The choice of the starting material is highly dependent on the specific application and characteristics of the fabricated devices. The semiconductor process used for bipolar as well as BiCMOS

device starts with a lightly P-doped wafer on which a N layer is grown epitactically. Epitaxy describes the growth of a single crystal film on top of a crystalline substrate, a crucial feature in semiconductor thin film manufacturing. In a SiGe HBT process homoepitaxy is employed where film and substrate are of the same material (Si on Si growth). The advantage of epitaxy is that grown layers are very pure and can be doped independently of the substrate material.

A base for processing a wafer for bipolar transistor integration is the availability of the LOCOS method (local oxidation of silicon). This procedure generally incorporates four basic layers or structures other than the silicon substrate to be structured by lithography. The used materials are highly compatible with the standard process route and used materials.

In a very first step the low-resistivity buried layer (BL) is formed by a high-dose implantation and a long high-temperature annealing where oxide serves as an implantation mask. Thereafter a shallow trench isolation (STI), being the primary technique for device isolation for advanced CMOS technologies, is formed by a local oxidation.

The surface is then polished using chemical mechanical polishing (CMP) in order to remove topography and maintain good uniformity. After the planarization step of the silicon surface, the wafer is covered with silicon oxide (SiO_2) or buffer oxide by chemical vapor deposition (CVD). A second layer of silicon nitride (Si_3N_4) is deposited on top followed by a third mask of SiO_2, the insulation oxide (typically by thermal oxidation) (a). In the following process steps selective etching is used to properly form the desired device structures using SiO_2 or Si_3N_4 as mask [135].

In a next step the active area of the high speed (HS) bipolar devices is opened and the selectively implanted collector (SIC) region is positioned directly under the intrinsic base formed later (b). A heavy collector doping will help to retard the onset of undesired effects as base push-out and the Kirk effect. Using an ion implantation process at high concentration and an adequate post implantation annealing the SIC is brought into the collector profile (of HS devices) in order to achieve better performance [136].

The Base Profile

The main target for optimization in the device profile is the base. Narrowing down the boron profile or base width w_B respectively, lowers the distance the carriers have to pass, lowering the base transit time τ_B. In addition the shape and concentration of the germanium profile is a parameter of optimization.

The vertical device profile directly affects the base width and thus the base transit time. Special attention is hence to be given to all process steps forming the base profile (step (c) in the process flow). After entirely opening the emitter window through SiO_2 removal, the Si surface for selective growth on the existing Si surface is available.

The main drawback of a DP structure (difficulty in maintaining a narrow vertical profile of the

intrinsic base) was effectively solved by the epitaxial base deposition. Introducing the epitaxial film growth technique, the base width w_B can be controlled with an accuracy of less than some hundreds of Ångsroms (few nm) providing good control of the vertical base profile.

The epitaxial SiGe film is deposited using rapid thermal chemical vapor deposition equipment (RTCVD), a thermally enhanced process carried out at high temperature yet for a very short time. The deposition is subdivided in three basic steps (cf. Fig. B.2). On top of the silicon substrate (Collector) a thin buffer layer of pure Si is deposited for the initial growth at the interface plane. This intermediate step helps forming the SiGe/Si interface plane that is deposited on top of the buffer. The in-situ p-doped SiGe base layer with the vertical base doping and germanium profile is combined with carbon in order to add compression stress (lattice strain respectively) to the layer and keep out-diffusion under control. On top of the SiGe layer another pure Si layer is added to the stack. The so called Si cap layer forms the Si/SiGe interface plane to have a homogeneous interface to the Emitter and help overall film stability.

Several process parameters can be adjusted in this step and add to the complexity of the base deposition. Through changes in the composition of the deposited material (in-situ) the factors stress (through carbon content), band-gap (through Ge composition) and doping (through boron incorporation) are defined depending on the vertical profile depth. The proper choice of the profile is not only important with respect to manufacturability constraints. As a key factor in SiGe transistor engineering the Ge content in the hetero-layer being a function of the profile depth in vertical direction and largely influences the physical operation of the HBT device.

Several concepts exist, namely a box shape with high Ge content to reduce the emitter transit time τ_E or a trapezoidal (or triangular resp.) shaped profile with a rising edge from the emitter into the base. This creates an additional electric drift field and thereby reduce τ_B. Since a trapezoidal Ge profile shape results in a desired accelerating electric field for electrons injected from emitter, the Ge mole fraction of the $Si_{1-x}Ge_x$ composition is varied in three discrete steps with increasing Ge amount towards the collector junction.

In several process steps a heat treatment (for annealing of implantation damages) may cause dopant outdiffusion causing the profile to reach into adjoining Si profiles which may cause undesired barriers and thereby degrade device performance. However a Gaussian doped base impurity profile with heavy doping is desirable, since it assures low base sheet resistance (r_{sBi}), which is important for reaching low transit times τ_B desired for high frequency operation.

Incorporating carbon with a strain-relieving defect formation prevents strain relaxation due to high-doping effects and thus minimizes base dopant out-diffusion of boron. This compensation technique maintaining critical thicknesses despite following annealing process steps enables a much higher base doping in a very thin SiGe base layer effectively lowering intrinsic base resistance whilst a narrow base profile is maintained [137].

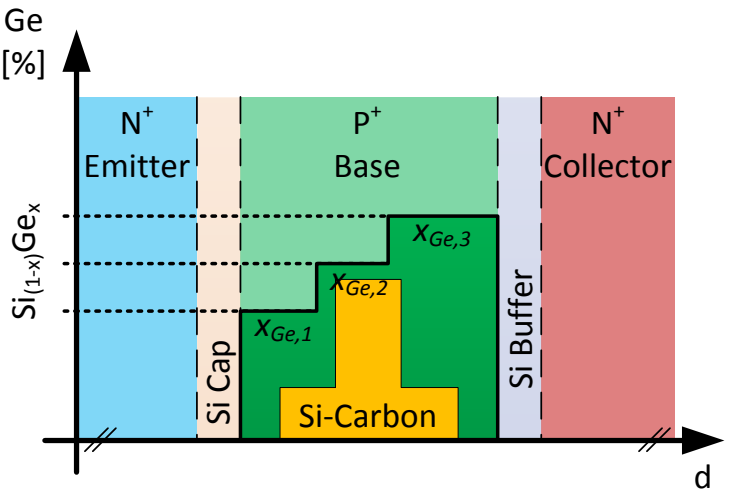


Figure B.2: Deposition steps and layers of the epitactically grown SiGe base layer in a selective epitaxial growth process

Process	B55	B5T	B9MW*
$x_{Ge,1}$[%]	20	20	10
$x_{Ge,2}$[%]	25	25	25
$x_{Ge,3}$[%]	30	30	-

Table B.1: Process parameters of Ge profile steps used in vertical SiGe profile of different HBT technology generations; **process using a two-step Ge profile*

The Emitter Architecture

The DP technology incorporates the formation of a L-shaped inside spacer providing separation of the extrinsic base link from the emitter and further reducing the effective emitter window width. After deposition of an oxide layer the required nitride spacers (SiN) are deposited all over the wafer. Thereafter the nitride (Si_3N_4) is removed using highly selective directional etch of nitride leaving the spacer within the emitter window (d). The effective junction width of the emitter window opened by lithography is now reduced by the thickness of the two spacers on each side. At last the etch-stop SiO_2 layer is removed providing access to the SiGe:C base.

For emitter formation heavily n-doped (in-situ) silicon is the deposited by another epitaxy of silicon by high temperature CVD (e). Etching the silicon emitter poly material leaves the n+ emitter on top of the SiGe base.

Extrinsic Device Periphery

The intrinsic device is now fully functional. The remaining process steps are dedicated to contact formation yet equally important for reasons of parasitics reduction affecting the device speed. In order to provide access to the SiO_2 mask the nitrite is removed. Photo-lithography and etch gives access to the base poly building the contact area for the base link to the base contact (f). The following salicidation step (deposition of a metal-silicon with low resistivity) on top of the base link (g) affects the extrinsic base resistance (R_B) of the SiGe HBT.

Tungsten plug fill has been the established method for filling contacts and vias in front- and back-end metallization to heavily doped silicon. The low-resistance contact plug metallization

deposited using a selective composite CVD of tungsten (W) for filling the via holes in the SiO_2 mask with high aspect ratio contact material and provide access to the first metal (M1) interconnect copper layer (h).

The Final Device

The device cross section of the HBT transistor after front-end processing is shown in Fig. B.3. The associated doping profile of the three individual transistor regions is shown thereafter (cf. Fig. B.4). The one-dimensional impurity dopant profile (N) of the npn bipolar transistor structure in vertical direction (x) is indicating the boundaries of the most important regions for the operation of the device (cf. Fig. B.4).

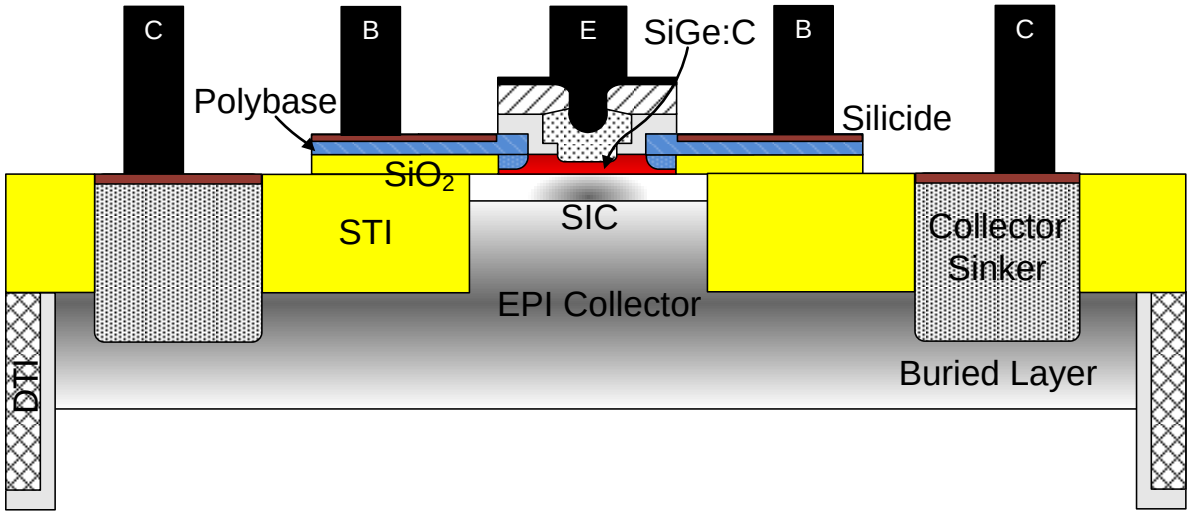


Figure B.3: HBT cross section after front end of line process for symmetrical CBEBC device

The three individual (neutral) regions are each separated by a space charge region (SCR) with the definition of the boundaries as function of the profile depth (x). Metallurgical junctions (denoted x_j) are defined as the point where the net impurity concentration is zero (acceptor N_A equals donor N_D concentration). A proper definition of the electrical junctions will be given in the chapter concerning device simulation (cf. Section 2.4).

- $[0, x_{jE}]$: n-doped emitter (E)
  - $[w_E, x_e]$: base-emitter (BE) space charge region
- $[x_{jE}, x_{jC}]$: p-doped base (B)
  - $[x_c, x_{c,C}]$: base-collector (BC) space charge region
- $[x_{jC}, x_{jS}]$: n-doped collector (C)

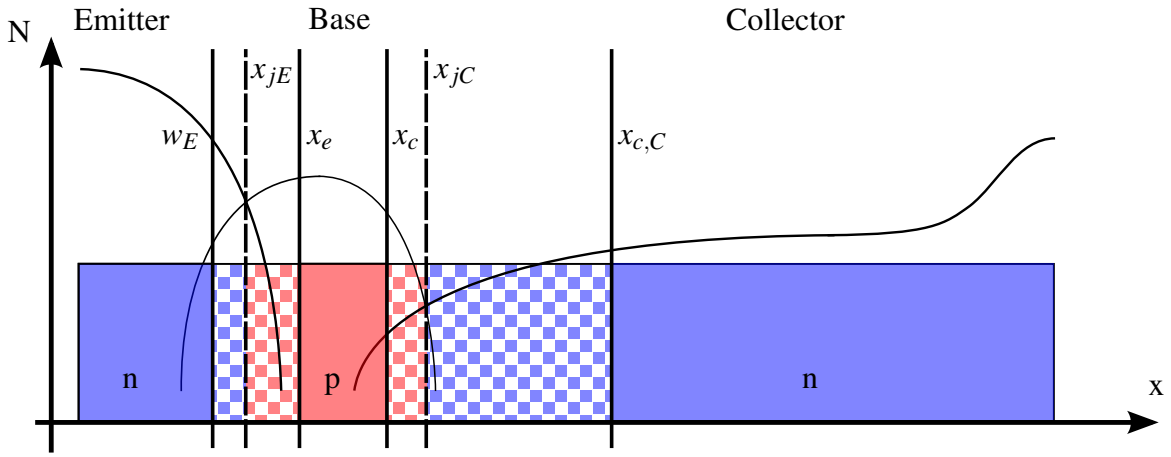


Figure B.4: Vertical impurity profile of npn BJT with illustration of space charge layers [*Checkered regions indicate space charge regions; changes in color indicate metallurgical junctions (x_j)*]

Variation of Device Performance

HBTs are selected by the needs for specific applications. However one figure of merit (FoM) alone, characterizing for example high speed properties (f_T and f_{max}), does not provide a complete technology information without the mention of breakdown characteristics (eg. BV_{CEO}). It is hence important that multiple device types (flavors respectively) can be co-integrated with the simple modification of implant masking steps. For definition of the most important FoMs the reader is referred to Appendix C.

Through (small) process changes a variety of devices is manufactured in the BiCMOS technology: The research focus of this work is on high-speed devices (HS). However making a modification of the collector profile (through omission of the SIC implant) a set of medium voltage (MV) devices (non-SIC transistor) with increased break-down limit yet lower cutoff frequency can be realized. In return one obtains a device with larger Base-to-Collector breakdown voltage (BV_{CBO}) as well as Emitter-to-Collector breakdown voltage (BV_{CEO}). Additionally a third high voltage (HV) device flavor may realized by modification (omission of the high doping implant) of the buried layer.

As seen in Tab. B.2, the device offer of the analyzed BiCMOS55 technology integrates various flavors of SiGe HBT devices at a technology node ranging from high performance type to a high breakdown type. It is important to note that the base-emitter architecture remains completely unchanged requiring no additional mask steps and thus keeping cost low. The offer given in the design environment ranges from nominal parameters of 1.5V BV_{CEO} and 320GHz f_t all the way to 3.5V BV_{CEO} and 60GHz f_t.

Parameter	HS device	MV device	HV device
f_T	320	180	60
f_{max}	370	370	300
BV_{EBO}	1.8	1.8	1.8
BV_{CBO}	5.2	6.7	13.5
BV_{CEO}	1.5	1.8	3.5

Table B.2: RF FoMs of SiGe:C devices supported in the design kit of 55nm technology (High Speed, Medium Voltage and High Voltage) characterized by their specification of f_T, f_{max} and BV_{CEO}

B.2 The BiCMOS Manufacturing Flow

At a given technology node (lithography node is defined through the smallest manufacturable lateral distance between two lines or contacts), CMOS typically requires less masking steps than the SiGe/BiCMOS as the bipolar process adds complexity to the root process. This generally results in higher wafer costs and faster manufacturing cycle times. However BiCMOS platforms offer high RF performance for analog and digital applications.

The basic idea of BiCMOS processes is an integration of a bipolar device with performance matching a bipolar only process (cf. Chapter B.1) into a final process including CMOS devices with characteristics unchanged from the original CMOS only process. The process lined out subsequently is embodied in a 55nm CMOS core process with a copper back-end-of-line [62, 138].

A key issue of modern BiCMOS manufacturing remains the thermal budget of annealing processes. The base epitaxy needs a high temperature. Proper placement of process steps for both thechnologies is therefore crucial for an adapted thermal budget. A so called 'HBT before gate' process as presented subsequently is favorable where the HBT module is inserted before the gate and entirely eliminates the impact of HBT thermal budget on CMOS whereas the HBT withstands the thermal budget of the CMOS specific process steps [139]. A schematic view of involved process steps is shown into Fig. B.7.

Front-End Manufacturing

Most of todays SiGe HBT technologies rely on a conventional collector structure meaning a heavily arsenic-implanted buried layer under lightly doped epitaxial silicon. The process for the manufacturing of the BiCMOS technology starts with the Buried Layer (BL) formation and collector epitaxy. The following step is a forming of deep and shallow oxide, the deep trench isolation (DTI) used by a vast majority of BiCMOS technologies. The DTI effectively isolates devices and reduces parasitic capacitances, allowing to significantly reduce the HBT layout area and thus improving packaging density [140]. All those steps being exclusive features of the bipolar transistors. In contrast the following STI formation is in common with the CMOS manufacturing.

After the implantation of the collector sinker connecting the BL and C contact further CMOS specific processes for well formation and first gate deposition steps are carried out. Respecting the thermal budget limitations the following steps are dedicated to the HBT integration is carried out before the final formation of CMOS gates. The Front-End BiCMOS flow therefore comprises a built of the Base/Emitter architecture of the HBT between gate deposition and gate patterning.

As described in chapter B.1 those steps include the SIC implant, the B doped polybase-link, epitactical growth of the Si/SiGe:C in-situ B doped base, the inside spacers as well as the deposition and patterning of the in-situ doped emitter followed by the patterning of the extrinsic base link (state-of-the-art BiCMOS manufacturing using FSA DPSA-SEG process steps). The base flow therein is comprised of Si buffer layer followed by the SiGe base layer (with half-graded Ge profile) and a final Si cap layer deposited in a single process step. The substitutional C background is directly incorporated whilst the boron (B) doping is adjusted to obtain low internal base sheet resistance (R_{SBi}).

Except for contact formation and device connection all bipolar-dedicated steps are done at this point, leaving the gate patterning as well as Source/Drain formation of the CMOS process as well anneal followed by a commonly used silicide deposition. The Front-End processing is finalized by the formation of the common contact metalization to the first metal layer. This electrical connection is either realized with via contacts or stripe contacts, the latter being better suited for the high current densities encountered in high speed HBT devices in order to avoid electromigration induced damages. A summary of contact configuration for the used technologies is given in Tab. B.3.

Contact	B55	B5T	B9MW
Emitter	Stripe	Stripe	Stripe
Base	Single Via Row	Single Via Row	Single Via Row
Collector	Triple Via Row	Stripe	Double Via Row

Table B.3: Contact configuration for different SiGe HBT technology generations

Back-End Manufacturing

After all the transistors are manufactured in the Front-End process the Back-End-Of-Line (BEOL) integration technique provides the inter-device connection as well as the connection to bond pads on the chip surface. With increasing number of interconnect levels a good control of planarization in the previously manufactured layers is required. It is hence desirable to use standard metallization layers of CMOS in the BiCMOS process as well building existing experience as well as models for parasitics. The more advanced metallization and planarization techniques of modern CMOS fabrication also permit vias to be stacked on top of lower-level vias and contacts [141]. The increased number of transistors per chip and reduction in die size favors higher stacking whilst overall capacitance of passives (line to line or line to substrate) can be decreased at the same

time by increasing the number of interconnect layers.

For the used technologies up to eight metal layers (BiCMOS55 technology) allow stacking of contacts and vias in mutiple metal layers to connect different types of devices, passives and to reach the pad connection (cf. Fig. B.5). The cross-sectional TEM analysis of dielectrics, metals, and vias shows the full metal stack as well as critical dimensions (cf. Fig. B.6).

The first metal film stack is substantially different from the following layers since it provides the tightest possible lateral dimensions offered in the process technology as well as the best aspect ratio. The contact is made directly to the junctions of the devices and contact windows are very deep and narrow. Subsequent layers have to keep stress low and respect a thermal budget in order to avoid damage to underlying layers.

In order to reduce resistivity the formerly used aluminum has been replaced by cooper whereas the classical SiO_2, providing electrical isolation between two metal layers is often replaced with other materials, providing lower dielectric constant [142]. The electrical coupling between metal 1 and metal 2 is achieved by a via 1 module. As for the contact fill in the device connection, a conducting material is filling up the vias followed by a polishing step to planarize the surface. Those steps are repeated with increasing critical dimensions with increasing number of metallization layer. A feature not necessarily available in pure CMOS environments is the analog Back-End. For RF applications the BEOL with its multiple structured metal and high-k dielectric layers holds most of the analog passive components (e.g. low-tolerance resistors, junction varactors, inductors and high-density metal-insulator-metal (MIM) capacitors). The final passivation layer on top of the chip fulfills the role of environmental protection.

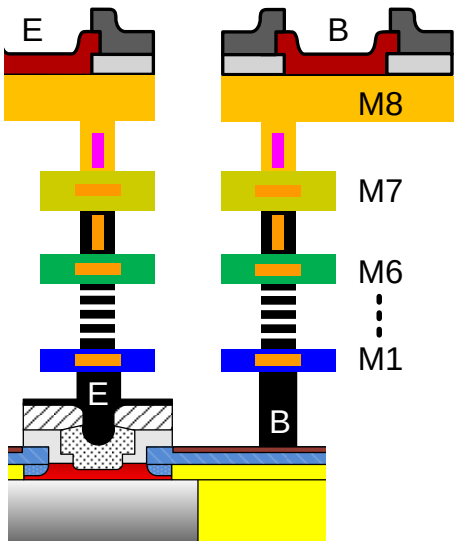


Figure B.5: Schematic view of Back-End-of-Line metallization

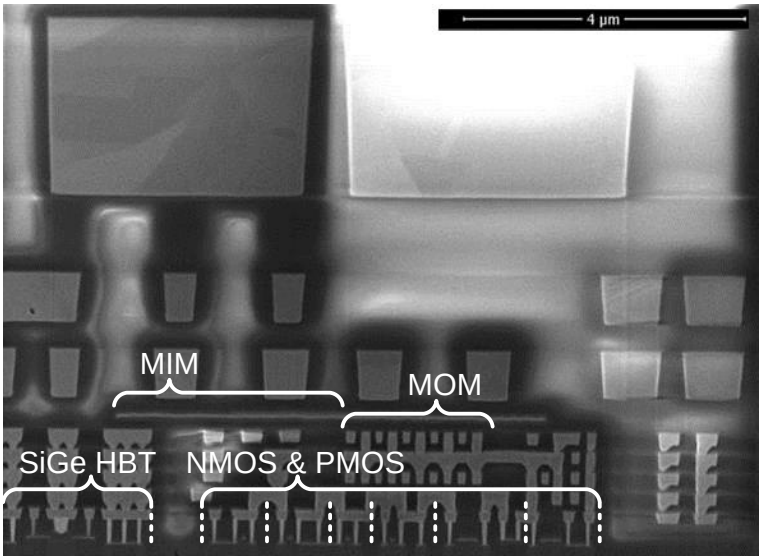


Figure B.6: TEM cross-section of Back-End-of-Line for SiGe BiCMOS process with multiple devices

B.3 Schematic View of the BiCMOS Manufacturing Flow

The flowchart below shows schematically the process steps for integration of bipolar HBT technology into a CMOS process as described in section B.2.

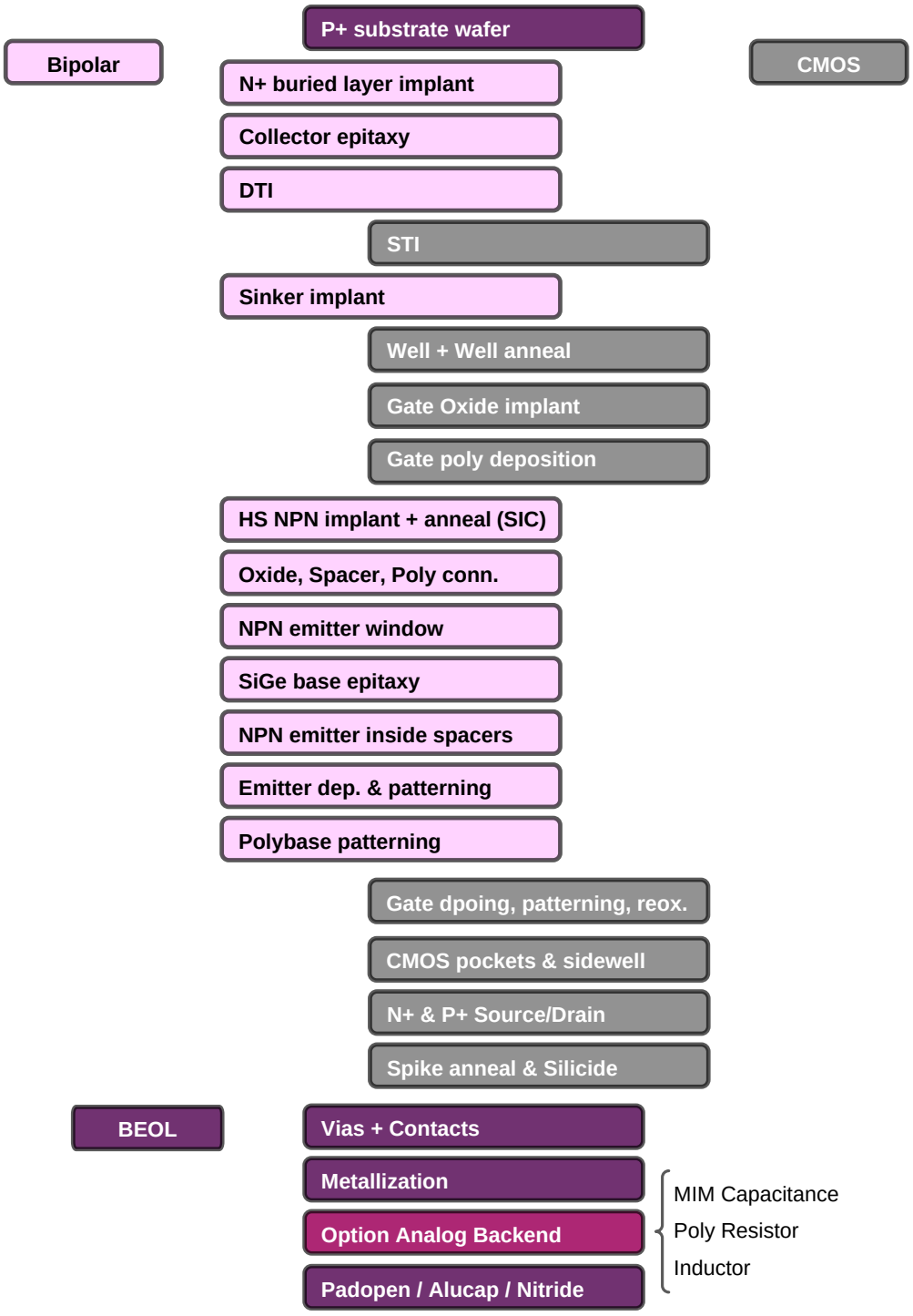


Figure B.7: SiGe compatible HBT integration to CMOS process, schematic view of the BiCMOS process flow for fabrication of analog mixed signal circuits

C Device Performance Measures

C.1 Figures of Merit of HBT Transistors

Key aspects for technology selection considering micro and millimeter-wave designs are noise (RF and 1/f), speed of operation (f_T, f_{max}), current gain at low current and a variety of different devices (wide and short emitter stripes) as well as reasonable operating voltages (BV_{CEO}) and device linearity.

The proper measure of the quality of transistors is crucial to compare various types of Si-based bipolar transistors and technologies. In oder to assess the advancement of technologies there are a few important measures that allow to characterize device performance. The proper definition of frequently used abbreviations and quantities and *F*igures *o*f *M*erit (FoMs) is hence important.

The bipolar circuits of todays industry available solutions targeting the area of high-speed applications are commonly measured by their performance in radio-frequency (RF) circuits operating at alternating currents (AC). Amongst others the most critical FoMs for todays high-speed devices are the current gain cutoff frequency f_t , the maximum frequency of oscillation f_{max} and the collector-emitter breakdown voltage BV_{CEO} [143].

Short-Circuit Current-Gain Cut-Off Frequency

The maximum current gain cutoff frequency (peak f_T or $f_{T,peak}$ respectively) is a strong function of the vertical device profile and transistor performance and thus reflects nicely the advancement and sophistication of technology. f_T is directly related to emitter to collector transit time (c.f. C.1) where τ_E is the emitter charging time, τ_B the time required to discharge the excess electrons in the base through the collector junction, τ_{RC} the collector charging time and $\tau_{C,SCR}$ the collector space charge transit time.

$$\tau_{tot} = \tau_{EC} = \frac{1}{2\pi f_t} = \tau_E + \tau_B + \tau_{RC} + \tau_{C,SCR} \tag{C.1}$$

Due to capacitances in the device equivalent circuit the small-signal current gain (h_{21}[1]) rolls off towards high frequency. The cutoff frequency f_t is defined as frequency where h_{21} equals unity[144]. The value of f_t gives an idea of the intrinsic delay of the transistor as a first-order FoM for its frequency response

$$h_{21}(f_t) = \frac{i_c}{i_b} = 0 \tag{C.2}$$

giving the simple formulation

$$f_t = \frac{f_{meas}}{2\pi \cdot imag(1/h_{21})}. \tag{C.3}$$

[1]Transistor current gain is defined as ratio of the small signal output current to input current of the transistor with the output short-circuited.

Maximum Frequency of Oscillation

High speed SiGe HBTs are reaching maximum frequency of oscillation (f_{max}) in a range from about 300GHz all the way up towards 500GHz. Though f_{max} presents one of the important FOM in circuit design representing the frequency limit, at which the unilateral power gain (U) rolls off to 1 (0dB respectively). Mason's invariant or the unilateral power gain for a linear two-port devices (U) can be used as a figure of merit to compare any three-terminal, active device. It is defined as

$$U = \frac{|y_{21} - y_{12}|^2}{4 \cdot [\Re(y_{11})\Re(y_{22}) - \Re(y_{12})\Re(y_{21})]} \text{ [145, 146].} \tag{C.4}$$

Even though other (simple) formulas exist[147], the accurate definition of the maximum frequency of oscillation (f_{max}) in a circuit where only one active device is present is defined as the frequency f where the unilateral gain U equals one ($U = 1$). Current high-frequency transistors those frequencies reach values in excess of 300GHz. Due to limitations of the measurement equipment and techniques (notably the applicable measurement frequency range f_{meas}), the unilateral power gain could not be determined directly up to U=1. Hence a common practice is to trace the values of the cutoff frequency f_max by extrapolating at -20 dB/decade the unilateral power gain U measured at lower frequencies (up to 110GHz).

A common simplified relation developed by Pritchard [148], and refined later [149] shows the most important interdependence factors of f_{max} on device parameters with R_B being the base resistance and C_{BC} being the collector-base capacitance.

$$f_{max} = \sqrt{\frac{f_t}{8\pi R_B C_{BC}}} \tag{C.5}$$

Breakdown Voltage

The open-base breakdown voltage BV_{CEO} is a third critical parameter. There are circuits operating at supply voltages close or beyond the breakdown voltage constraint. In high voltage design these voltages play a crucial role both for functionality and reliability. As the reverse bias potential across the BC junction (V_{BC}) increases, the probability that a carrier in the depletion region will undergo an impact ionization increases and impact ionization may generate an electron-hole pair causing the avalanche multiplication and eventually leading to junction breakdown.

Due to contradicting profile optimization constraints of the collector BV_{CEO} is closely related to f_t through the Johnson limit [56] where the product of f_t and BV_{CEO} is said to be constant.

C.2 Evolution of RF Bipolar and BiCMOS Technologies Manufactured by STMicroelectronics

In order to show the technology evolution of RF BiCMOS technologies at STMicroelectronics Tab. C.1 summarizes the development history of production and prototyping technologies in 200mm (8") and 300mm (12") wafer fabrication in Crolles, France. The maximum transit frequency of devices was continuously increased in order to provide faster circuits or operate circuits at lower power levels.

Technology Generation	Technology Node	Device Architecture	Performance [f_T / f_{max}]
BiCMOS6	0.35μm	Si BJT[1]	25GHz / 40GHz
BiCMOS6G	0.35μm	SiGe HBT, NSEG[1]	45GHz / 60GHz
BiCMOS7	0.25μm	SiGe HBT, NSEG[2]	70GHz / 90GHz
BiCMOS7RF	0.25μm	SiGeC HBT, NSEG[2]	60GHz / 90GHz
BiCMOS9	0.13μm	SiGeC HBT, NSEG[2]	160GHz / 160GHz
BiCMOS9MW	0.13μm	SiGeC HBT, SEG[2]	220GHz / 280GHz
*B3T**	0.13μm	SiGeC HBT, SEG[2]	260GHz / 330GHz
*B4T**	0.13μm	SiGeC HBT, SEG[2]	270GHz / 370GHz
*B5T**	0.13μm	SiGeC HBT, SEG[2]	300GHz / 400GHz
BiCMOS55**	55nm	SiGeC HBT, SEG[2]	320GHz / 370GHz

Table C.1: development history of RF BiCMOS technologies at STMicroelectronics with corresponding figures of merit and technology; technologies in 200mm Fabrication; [**bipolar only R&D technologies developed in the framework of the DOTFIVE project*; ** technology in 300mm fabrication; [1] single poly, polyemitter Technology; [2] double poly, mono-emitter technology]

D Cross Section and Top View of Vertical SiGe HBT Device

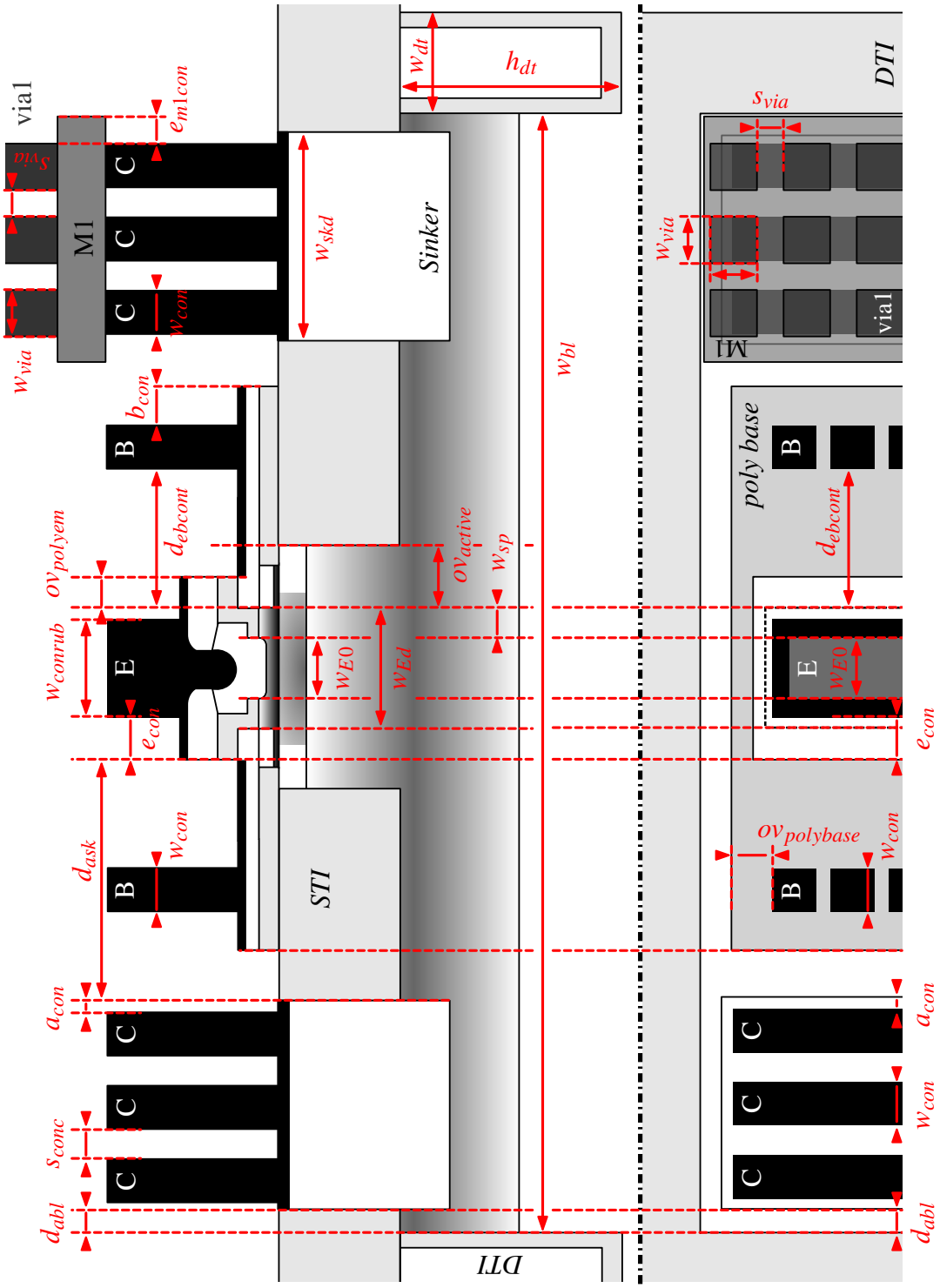


Figure D.1: Detailed cross section and top view of a vertical SiGe HBT device with definition of global geometry parameters for scalable modeling (cf. section 3.2.1)

E The Mass Action Law and the Implication for SiGe Devices

The impact of the SiGe composition on transistor engineering is predominantly seen in the density of states of the carriers effectively reducing the effective mass of the carriers. To make the link between the density of states (n_i) and the collector current of the bipolar transistor it is useful to derive the fundamental equations for carrier transport providing the basic equations used in most bipolar transistor models that are commercially available.

A quantum-mechanics-based analysis of the semiconductor band structure results in the carrier diffusivity $D_C(E)$ and $D_V(E)$ with the effective density of states in the conduction and valence band (N_C and N_V) gives both governing equations (E.1) and (E.2). The electron density of a semiconductor can be obtained from

$$n = N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) \qquad \text{with} \qquad N_C \equiv 2\left(\frac{2\pi m_n^* k_B T}{h^2}\right)^{3/2} \tag{E.1}$$

where N_C is the effective density of states in the conduction band and m_n^* is the electron effective mass. In the same way, the hole density p in the valence band is defined as

$$p = N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right) \qquad \text{with} \qquad N_V \equiv 2\left(\frac{2\pi m_p^* k_B \cdot T}{h^2}\right)^{3/2} \tag{E.2}$$

where N_V is the effective density of states in the valence band (E_V) and m_p^* is the hole effective mass. Here h is Planck's constant, k_B is Boltzmann's constant, T stands for the absolute temperature, and m_e^* and m_h^* are the electron and hole effective masses respectively[1]. By definition the two concentrations of free electrons and holes in equilibrium are independent of the Fermi energy. Hence using the the Fermi-Dirac-statistics for the occupation of states[2] one obtains the energy distribution of electrons (n_0) and holes (p_0) [the index $_0$ indicates concentrations at equilibrium]. However in a biased junction, the pn-product is increased above the equilibrium value due to application of a voltage. For an undoped semiconductor, the number of electrons per unit volume in the conduction band equals the number of holes per unit volume in the valence band, i.e. $n = p = n_i$. The electron density equals the hole density since the thermal activation of an electron from the valence band to the conduction band yields a free electron in the conduction band as well as a free hole in the valence band.

The intrinsic carrier density denominated with n_i plays an important role in the analysis of SiGe HBTs. Intrinsic semiconductors are semiconductors which do not contain impurities. The

[1] The concept of mass of carriers plays a central role in solid-state electronics. The effective mass (m^*) is different from the free carrier mass accounting for the effects of crystalline force and the quantum mechanical properties. The effective carrier mass along a particular direction (m^*) is defined through the momentum of the carrier in the direction of interest divided by the group velocity (v) in the same direction

[2] The Fermi-Dirac function shows the probability that an available state with energy E is occupied by a carrier

electron and hole density in any non-degenerate semiconductor is always equal to the square of the intrinsic carrier density. This relation is known as the mass action law and allows to calculate the hole density if the electron density is known or vice versa known as the Law of Mass-Action [115]. It defines the product of concentrations of electrons and holes as constant, irrespective of individual concentrations

$$np = n_i^2 \tag{E.3}$$

Using the relations (E.1) and (E.2) one can substitute the carrier density. The physical interpretation of N_C is a effective number density of accessible states at the conduction band bottom whereas N_V represents the effective number density of accessible states at the valence band top. E_c and E_v are the energies of the conduction and valence band edges. The energy gap in between both bands is defined as $E_g = E_c - E_v$ resulting in

$$np = n_i^2 = N_C N_V \exp\left(\frac{E_V - E_C}{k_B T}\right) \simeq N_C N_V \exp\left(-\frac{E_g}{k_B T}\right) \tag{E.4}$$

known as the law of mass action, defining the product of concentrations of electrons and holes in equilibrium at a certain temperature as constant. The intrinsic carrier density is then obtained by

$$n_i \quad = \quad \sqrt{N_C N_V} \exp\left(-\frac{E_g}{2k_B T}\right). \tag{E.5}$$

This intrinsic density n_i is a typical material parameter with a strong temperature dependency. In silicon as well as SiGe technology the intrinsic carrier concentration n_i may be changing through the apparent band-gap narrowing (BGN) that is caused by heavy doping effects or heterojunctions by a certain absolute value ΔE_g.

The *effective* intrinsic carrier concentration n_{ie} is a normalized factor using the relative band-gap ΔE_g change to obtain the ratio to the reference concentration n_i. For silicon-germanium alloys ($Si_x Ge_{(1-x)}$) typically used in SiGe HBTs (fractions up to 30%) the effective density of states is calculated by the simple formula

$$n_{ie}^2 = n_i^2 \exp\left(\frac{\Delta E_g}{k_B T}\right) \text{[69]}. \tag{E.6}$$

The reason for the obviously increased impact of the BE junction on the device characteristics is found in the germanium profile incorporated in fast bipolar transistors. There is a significant difference in the energy band-gap (E_G) or the intrinsic carrier concentration (n_i) respectively, when comparing pure silicon ($Si_1 Ge_0$) with pure germanium ($Si_0 Ge_1$).

The according material constants for the two materials of interest to calculate the effective density of states in conduction and valence band at room temperature are shown in table E.1

Material	Minimum indirect energy gap at 300K (E_g)	Effective density of states in the conduction band (N_C)	Effective valence band density of states (N_V)	Intrinsic carrier concentration (n_i)
Si	1.12 eV	$2.9 \cdot 10^{19} cm^{-3}$	$1.8 \cdot 10^{19} cm^{-3}$	$1.45 \cdot 10^{10} cm^{-3}$
Ge	0.66 eV	$1.0 \cdot 10^{19} cm^{-3}$	$5.4 \cdot 10^{18} cm^{-3}$	$2.4 \cdot 10^{13} cm^{-3}$

Table E.1: Properties of the SiGe material system at ambient temperature of 300K [69, 150]

The relation of band-gap and Ge fraction is described by the empirical second degree polynomial given in equation (E.7) that is valid up to a Ge fraction of 85%.

$$E_g(x) = (1.155 - 0.43 \cdot x + 0.206 \cdot x^2) eV \quad \text{for } Si_{1-x}Ge_x \text{ with } 0\% \leq x \leq 80\% \quad [151] \quad \text{(E.7)}$$

As a consequence of the different band-gap values of pure Si and pure Ge the band-gap in a gradually deposited SiGe alloy (cf. Section B.1) is a function of the germanium fraction thus the position in the vertical profile x. With the band-gap being dependent on the germanium fraction in the alloy, the intrinsic carrier concentration at room temperature differs significantly between both materials as shown in table E.1. Given the fact that the band-gap difference is weighted exponentially in the calculation of n_i, it is feasible to neglect band-gap narrowing due to high-doping and other effects, taking exclusively the band-gap offset as the main impact factor on n_i. This results in a plot of $r_{ni,eff} = n_{i,SiGe}/n_{i,Si}$ versus Ge fraction (cf. eqn. E.6) as shown in Fig. E.2. Since the germanium concentrations currently employed in SiGe technologies rarely exceed values of 30% the curve progression are plotted up to a fraction of 50%.

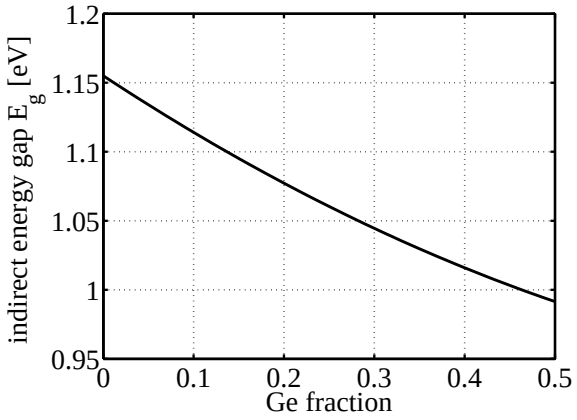


Figure E.1: Indirect bandgap (E_g) of silicon-germanium alloys versus Ge fraction

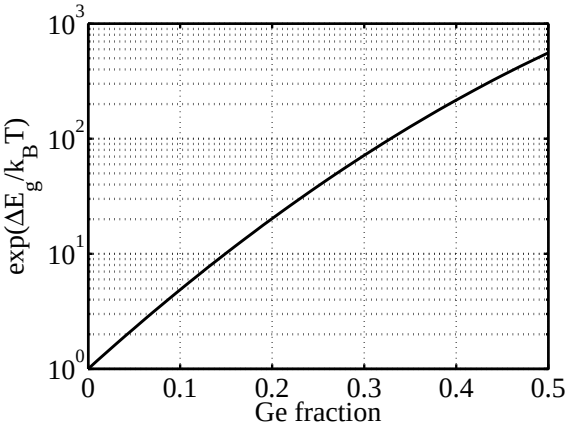


Figure E.2: Factor of the intrinsic carrier density (n_i) of a silicon-germanium alloy [150, 151]

F On-wafer RF Measurement

F.1 Measurement Setup

Characterizing a unknown device using RF measurement gives a linear behavioral model of the device at specific measurement conditions (e.g. operating point defined by voltage and current) versus frequency under different source and load conditions (e.g. short and open circuits). The behavioral model allows compute device parameters from measured data and obtain information about performance. However since accurate on-wafer RF measurements of HBTs depend on a number of factors, the fundamental basics for successful measurement are outlined below. A complete system for RF characterization of HBT devices usually consists of the following components:

- Vector Network Analyzer (VNA)
- connecting cables (transition) and bias supply (bias T)
- RF probes *with calibration substrate*
- probe station (wafer prober): *with thermally stabilized chuck, positioners, microscope*

A simplified setup schematic is shown in Fig. F.2. For high frequency measurements of high-speed devices a Vector Network Analyzer (VNA) is used for uncorrected measurements. The VNA unit itself measures a reference signal provided by an (internal or external) signal generator that supplies the required for the system usually swept over frequency. By signal separation the receiver monitors incident (r), reflected (a) and transmitted (b) travelling waves (cf. Fig. F.1) with complex amplitude and phase. The signal generator has an internal impedance R. In order to have the returning pulse completely absorbed in the generator the transmission line shall be terminated in its characteristic impedance ($R = Z_0$).

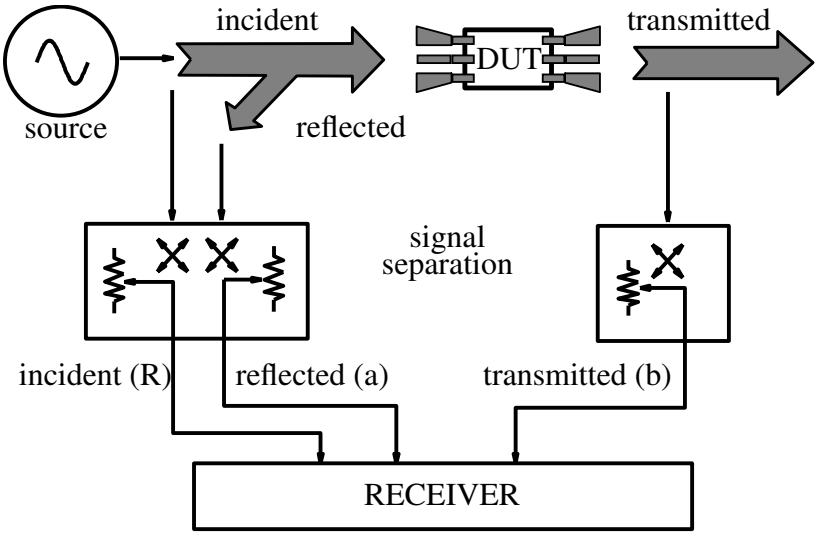


Figure F.1: Schematic view of Vector Network Analyzer components

The measured S-parameters (cf. section G.5) are calculated by the CPU of the VNA using the normalized transmitted and reflected waves and in general monitored and stored by a computer connected via GPIB interface.

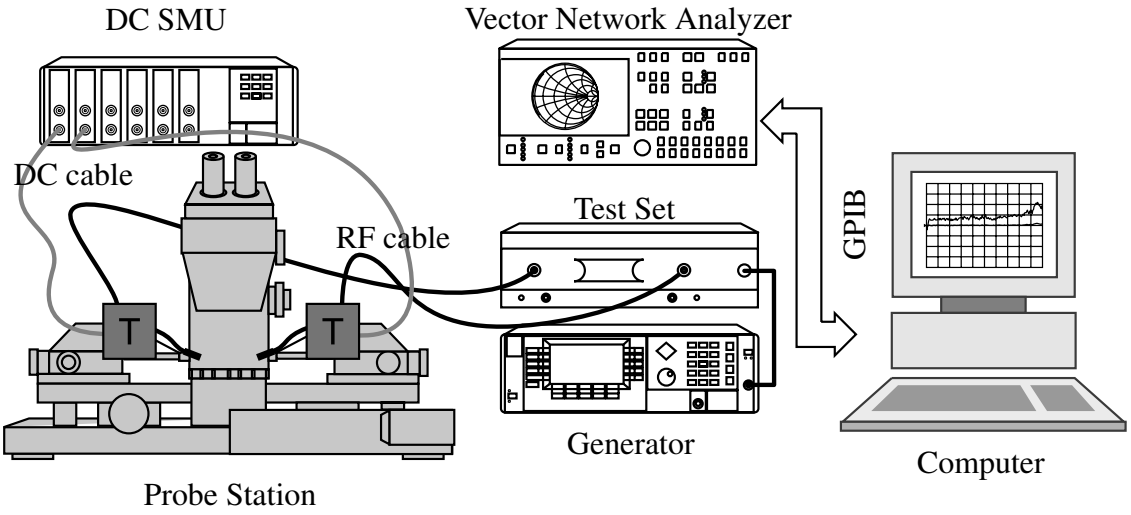


Figure F.2: Schematic view of measurement setup for RF on-wafer measurement

A DC source measure unit (SMU) being a precision voltage and/or current source and measuring instrument provides biasing for the desired operating point. A diplexer, the so called Bias Tees (T), is used to inject/remove the DC signal (currents or voltages) into the RF signal provided by the VNA without affecting the RF signal. They are passive components with a low frequency port (set the DC bias) and a high frequency port (passing the RF signal while blocking bias). The combined port is then connected to the device with both DC bias and RF signal. The bias T is normally located near the device in order to achieve lowest losses and distortion.

To connect the instruments with the device generally any measurement equipment requires transmission lines (transitions) from their ports to the individual ports of the device under test (DUT). In case of RF measurement these transitions are usually realized using coaxial connection cables or waveguides. Matching them to the characteristic impedance (Z_0) is very important for low reflection and maximum power transfer.

$$Z_0 = \frac{1}{2\pi}\sqrt{\frac{\mu_0}{\varepsilon_0\varepsilon_r}}\ln\left(\frac{D}{d}\right) \approx \frac{60}{\sqrt{\varepsilon_r}}\ln\left(\frac{D}{d}\right) \tag{F.1}$$

In this formulation D is the diameter of the outer conductor, and d represents the diameter of the inner respectively. The constants in the square root yields approx. 377Ω resulting in a factor 60. Usually coaxial cables provide electrical contact up to the frequency range measured for model parameter extraction[1]. The used 50Ω standard here is a compromise between the power handling

[1] up to approx. 67 or 110GHz

capacity peak at 30Ω and lowest signal attenuation at 77Ω in coaxial connections.

For the wiring of the tester setup flexible coaxial RF cables with low loss and good phase stability provide the connection between RF ports of the VNA and the bias T as well as bias T and probe while simpler coaxial and triaxial cables were used for LF signals. A typical coaxial cable has a dielectric insulation of the two signal paths with relative dielectric constant ε_r between inner and outer conductor.

Here one usually assumes $\varepsilon_r = 1$ for ideal vacuum and $\varepsilon_r \approx 2.29$ for polyethylene-insulated cables. The impedance is then determined by the ratio of the electric field E between the conductors, and the induced magnetic induction H caused by the current flow. To obtain the desired 50Ω termination in air-dielectrics the ratio D/d hence is ≈ 2.3.

The final contact to the wafer surface is provided by dedicated RF probes. Infinity probes of Cascade Microtech and Z-probes from SUSS MicroTec were used for measurements shown in this work.

Figure F.3: GSG probes on B5T HBT multi-project wafer for pulsed RF measurement on PA200 prober, IMS Bordeaux

F.2 Error Correction to Network Analyzer Measurements

All used system components add uncertainty and change system characteristics due to non-ideal behavior as well as changing environmental conditions in the measurement lab. Two-port vector calibrations are therefore required in order to create a reference plane with known standards. A proper calibration plays a key role in reducing system related errors. The procedure eliminates parasitics related to instrument related errors and cabling as well as probes, reducing error sources between the on-wafer device and the instrument.

De-embedding is defined as the process of mathematically removing the influences of transitions from the measured results. A full two-port error correction gives highest accuracy through removal of most of the effects related to errors from: directivity, source and load match, reflection and transmission tracking as well as crosstalk between the ports [152].

To a large extent the accuracy of a calibrated VNA measurement is determined by the techniques and completeness of the error model used to account for non-ideality of the system. Different techniques have been proposed and used for many years and the filed of de-embedding has been subject to a number of refinements to improve accuracy and applicability.

Properly defined and manufactured calibration standards are used in order to make two-port calibration fast and less prone to operator errors. Several calibration standards and procedures exist named after the standards incorporated used in the deembedding technique [153–156]:

- SOLT: Short Open Line Through
- LRM: Line Reflect Match
- TRM: Through Reflect Match
- TRL: Through Reflect Line
- LRL: Line Reflect Line

A general overview of calibration accuracy and quality versus ambiguity and ease of use is shown in Fig. F.4. As a result the full two-port calibration through SOLT has developed as a popular and widely accepted standard.

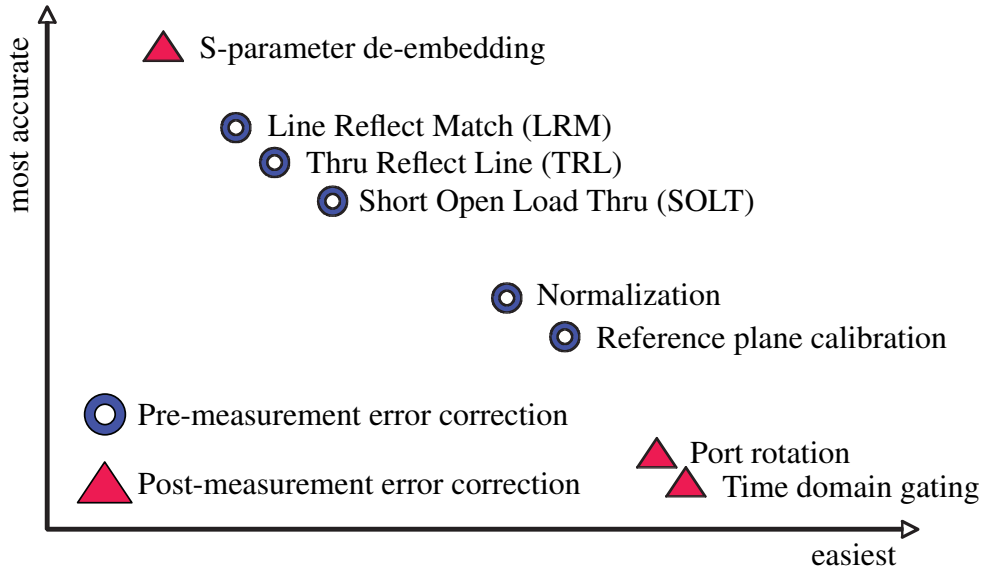


Figure F.4: Calibration standards evaluated in terms of their accuracy and simplicity [157]

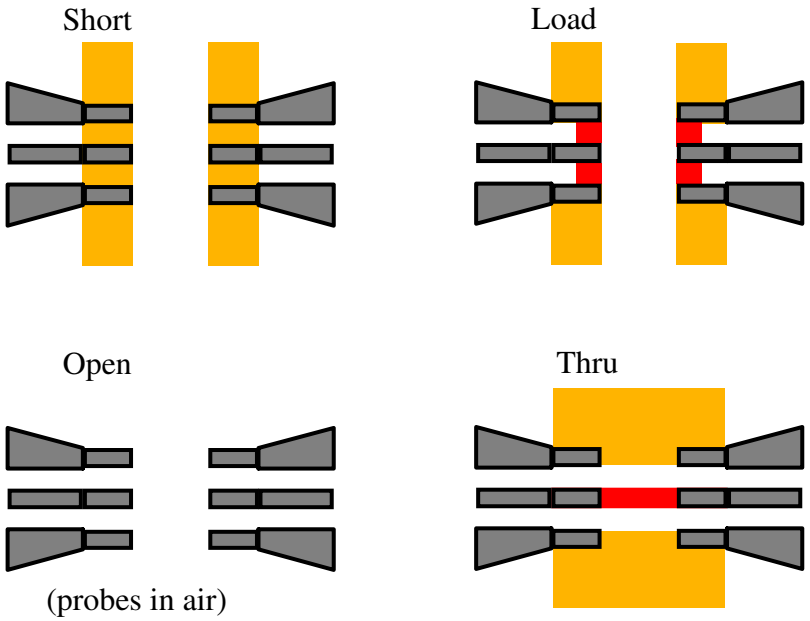
For on-wafer characterization generally a two-step de-embedding technique is used. In a first step the measurement system is corrected up to a defined reference plane. The VNA analyzes vector ratios of reflected and transmitted energy relative to incident upon the DUT in a stimulus-response measurement and thereby determines the properties of the device.

Within this first step the instrument de-embedding is realized by subtracting the network of the test fixture (error terms) from the measured result giving an effective measurement plane at the device connections (including internal VNA errors after sampler, cables and probes) [158].

The second part of the two-step de-embedding technique makes use of the dedicated complete-open and complete-short dummy structures attributed to each device on the test mask in order to correct the silicon backend parasitics using wafer-embedded test structures [159]. Even though the short de-embedding is generally known to have no significant influence on f_T characteristics it is especially important for correct modeling of the f_{max} characteristics.

Probe Tip Calibration

The commonly used calibration plane for on-wafer measurement is defined at the wafer probe tips. The impedance standard substrate (ISS) or calibration substrate positioned on a auxiliary chucks provides the required highly accurate standards for the first step of the de-ebedding procedure the so called probe tip calibration [160]. These standards are manufactured on a Impedance Standard Substrate (ISS) as high-precision thin film resistors, short-circuit connections as well as 50 Ω transmission lines as shown in Fig. F.6 for a SOLT correction.



available on virtually every VNA

- measured standards must be perfectly known
- non-ideality:

  open has capacitance

  short and load have inductance
- sensitive to probe placement

Figure F.5: Calibration standards on impedance standard substrate used for on-wafer calibration

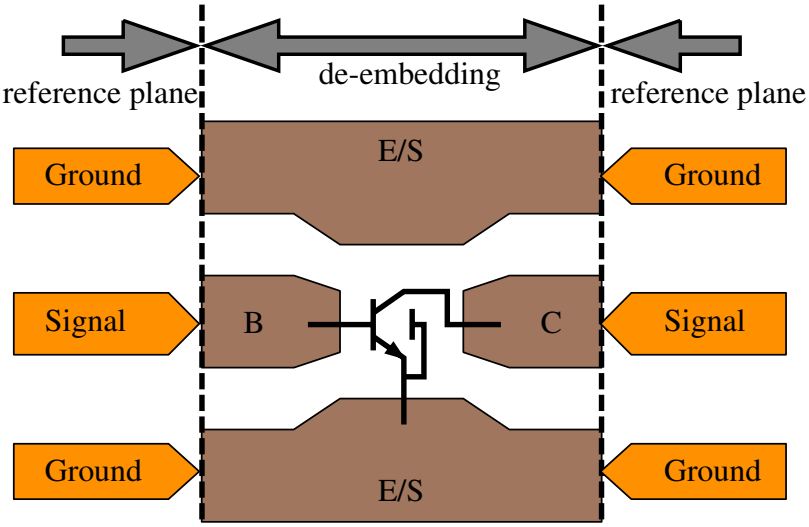
The OPEN standard is often implemented by simply raising the probe tips in the air above the wafer. A real THRU standard for on-wafer measurement is approximated by a transmission line between the signal pads since probes cannot be connected directly to each other.

Issues related to over-temperature RF measurements are related to system drift of probes and cables over temperature (predominantly phase error due to probe and cables expansion). It is hence recommended to repeat calibration after probes and cables have stabilized upon changing temperatures.

Configuration of the system depends on RF pad configuration and probe pitch (distance between contact pads). Since a significant amount of inaccuracy in microwave probing is associated with parasitic coupling at the probe tip, the preferred pad configuration approach for high-speed RF BiCMOS technologies is the ground-signal-ground (GSG) interface which effectively terminates field lines to both sides of the centered signal pad [161].

Further general rules for successful RF measurements include:

- RF probes should have $\geq 200\mu$m separation to avoid cross-talk
- grounds should be connected together
- adjacent devices should be far away to reduce interference
- oxidation of aluminium pads increases contact resistance [use tungsten tips, multiple touch-down, self-cleaning]



Full 2-port calibration gives all error parameters to account for:

- *directivity*
- *source and load match*
- *reflection and transmission tracking*
- *crosstalk*

Figure F.6: Reference plane for RF on-wafer calibration

De-embedding of the Silicon Backend Parasitics

With the reference at the probe tips measured data is the response of the DUT and the parasitics associated with the pads and contact lines. The importance of de-embedding the pad capacitance and metal interconnections of the BEOL in order to obtain intrinsic device characteristics is well known.

The simple OPEN de-embedding is widely used because of its simplicity by assumption that the parasitics leading to the DUT are parallel admittances y_p. The OPEN dummy structure predominantly accounts for parasitic capacitances (parallel elements, importance for high impedance devices).

The additional SHORT however is important to subtract series resistances (series elements to the DUT, important towards high frequency). All parasitics associated with probe pads and interconnect-metal lines can be represented and subtracted from the measurement.

The classical OPEN-only de-embedding is hence improved for high frequency measurement by a joint OPEN-SHORT de-embedding approach using the following formulas (cf. eq. F.2) in

order to shift the reference plane to the first metal layer (M1) [162].

$$Y_{HBT,O} = Y_{DUT} - Y_{OPEN}$$
$$Y_{HBT,OS} = \left[(Y_{DUT} - Y_O)^{-1} - (Y_S - Y_O)^{-1}\right]^{-1} \tag{F.2}$$

with Y_O: [y] matrix of OPEN standard and Y_S: [y] matrix of SHORT.

Yet it is evident that the pad OPEN and complete SHORT structures use a lot of Si surface and a scalable solution for de-embedding may be favorable with increasing number of devices.

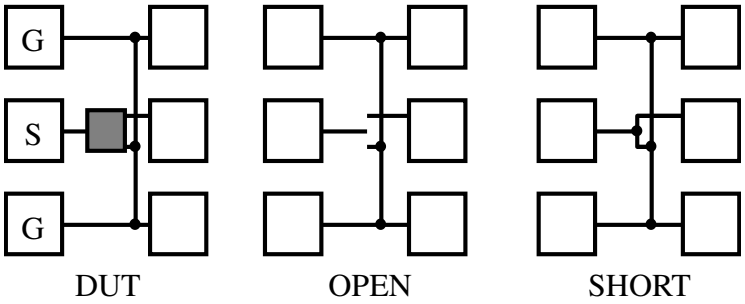


Figure F.7: DUT with dedicated complete-OPEN and complete-SHORT structure for on-wafer RF measurement

- pad capacitance between signal pad and ground (dielectric loss)
- series inductance and loss
- capacitance to ground
- contact inductance (depending on distance)
- contact resistance originating from non-ideal contact between probe tip and aluminum pad

G Transistor Two-Port Parameters

The intention of a small signal model is to use equations which relate small variations in currents and voltages to each other linearly creating a linear equivalent circuit through differentiation. A popular circuit representation widely used is the hybrid-pi model [163].

A two-port network (a kind of four-terminal network or quadripole) is an electrical network with two separate ports for input and output [164, 165]. Two-port-parameters in general (except for S-parameters) are defined through voltage and current with open or short in- or output ports. The most common representations are impedance parameters ([**z**]-parameters) and admittance parameters ([**y**]-parameters).

G.1 [z]-parameters

The [**z**]-parameters are also known as short-circuit impedance parameters as they are calculated under short circuit conditions:

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \tag{G.1}$$

resulting in:

$$\begin{aligned} Z_{11} &= \left.\frac{V_1}{I_1}\right|_{I_2=0} & Z_{12} &= \left.\frac{V_1}{I_2}\right|_{I_1=0} \\ Z_{21} &= \left.\frac{V_2}{I_1}\right|_{I_2=0} & Z_{22} &= \left.\frac{V_2}{I_2}\right|_{I_1=0} \end{aligned} \tag{G.2}$$

G.2 [y]-parameters

The Y-parameters are also known as open-circuit impedance parameters as they are calculated under open circuit conditions:

$$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \tag{G.3}$$

resulting in:

$$\begin{aligned} Y_{11} &= \left.\frac{I_1}{V_1}\right|_{V_2=0} & Y_{12} &= \left.\frac{I_1}{V_2}\right|_{V_1=0} \\ Y_{21} &= \left.\frac{I_2}{V_1}\right|_{V_2=0} & Y_{22} &= \left.\frac{I_2}{V_2}\right|_{V_1=0} \end{aligned} \tag{G.4}$$

Especially under cold or non-active bias conditions the π-equivalent representation of the measured [**y**]-parameters is important as it provides the representation of capacitances as defined below.

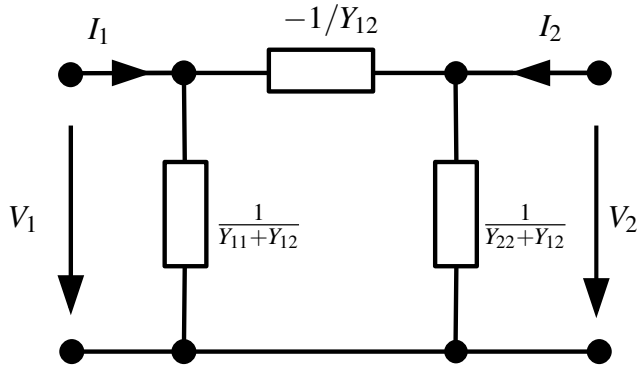


Figure G.1: Simplified π-model of measured [**y**]-parameters for bipolar transistor in off-state (cold)

$$C_{BE} = \frac{Im(Y_{11}+Y_{12})}{\omega} \tag{G.5}$$

$$C_{BE} = \frac{-Im(Y_{12})}{\omega} \tag{G.6}$$

$$C_{CS} = \frac{Im(Y_{12}+Y_{12})}{\omega}. \tag{G.7}$$

G.3 Transistor Hybrid Parameters

Basically every linear circuit having input and output terminals can be analyzed by four parameters called hybrid or h-Parameters. Hybrid represents a set of mixed parameters since they are of different dimensions (impedance, admittance and dimensionless) typically well suited to transistor circuit modeling. Those parameters are very convenient since they allow for simple development of formulas for input impedance, voltage gain etc..

This circuit has input voltage and current labeled V_1 and I_1 and the respective output voltage and current labeled V_2 and I_2. Per definition both input and output currents I_1 and I_2 are assumed to flow into the linear two-port.

$$\begin{aligned} V_1 &= h_{11}I_1 + h_{12}V_2 \\ I_2 &= h_{21}I_1 + h_{22}V_2 \end{aligned} \tag{G.8}$$

$$\begin{aligned} h_{11} &= \left.\frac{V_1}{I_1}\right|_{V_2=0} & h_{12} &= \left.\frac{V_1}{V_2}\right|_{I_1=0} \\ h_{21} &= \left.\frac{I_2}{I_1}\right|_{V_2=0} & h_{22} &= \left.\frac{I_2}{V_2}\right|_{I_1=0} \end{aligned} \tag{G.9}$$

The matrix representation can be interpreted as follows:

- h_{11} ⇒ Input impedance with output short circuited
- h_{12} ⇒ Reverse voltage transfer ratio with input open circuited
- h_{21} ⇒ Forward current gain with output short circuited
- h_{22} ⇒ Output admittance with input open circuited

The values of the h-parameter model are complex numbers varying as a function of the applied frequency and bias conditions. The simple equivalent circuit of the h-matrix is shown in Fig. G.2.

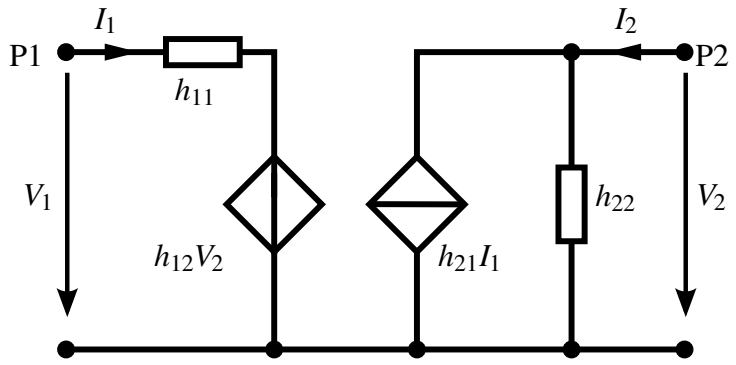
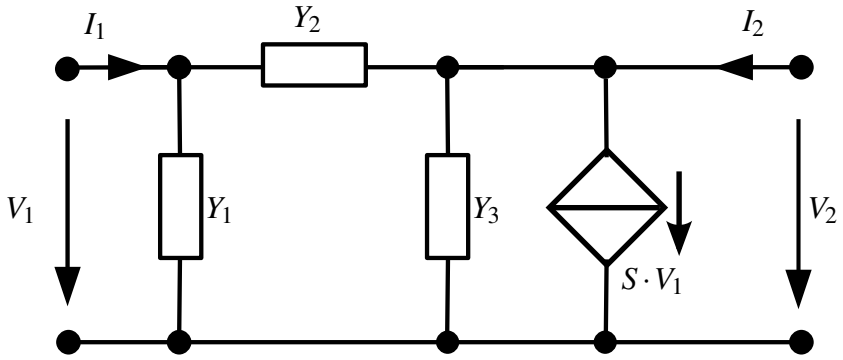


Figure G.2: Two port h-Parameter representation of a simple transistor in amplifier configuration

In forward active mode a more sophisticated representation of the BJT for small-signal operation is the extended π equivalent circuit (hybrid-π) as shown below.



- $Y_{11} = Y_1 + Y_2$
- $Y_{12} = -Y_2$
- $Y_{22} = Y_2 + Y_3$
- $Y_{21} = S - Y_2$

Figure G.3: Simple π-model for bipolar transistor in fwd. active

The other way around one obtains:

- $Y_1 = Y_{11} - Y_2$
- $Y_2 = -Y_{12}$
- $Y_3 = Y_{22} + Y_{12}$
- $S = Y_{21} - Y_{12}$

With the given boundary conditions for each parameter it is possible to perform measurements of the hybrid parameters on a BJT in active mode. However for application to transistor measurements the hybrid π-model (Giacoletto model [163]) is preferred due to its suitability for a higher frequency range. This is due to the fact that it comprises more of the frequency sensitive components present in a BJT transistor than the simple h-matrix. It is sufficiently accurate for low-frequency operation and may easily be adapted towards higher frequency.

The hybrid π-model is also a basic circuit representation for several parameter extraction strate-

gies from small-signal measurements. Through modification by addition of appropriate capacitances and other parasitic elements it may be adapted towards higher frequencies as well.

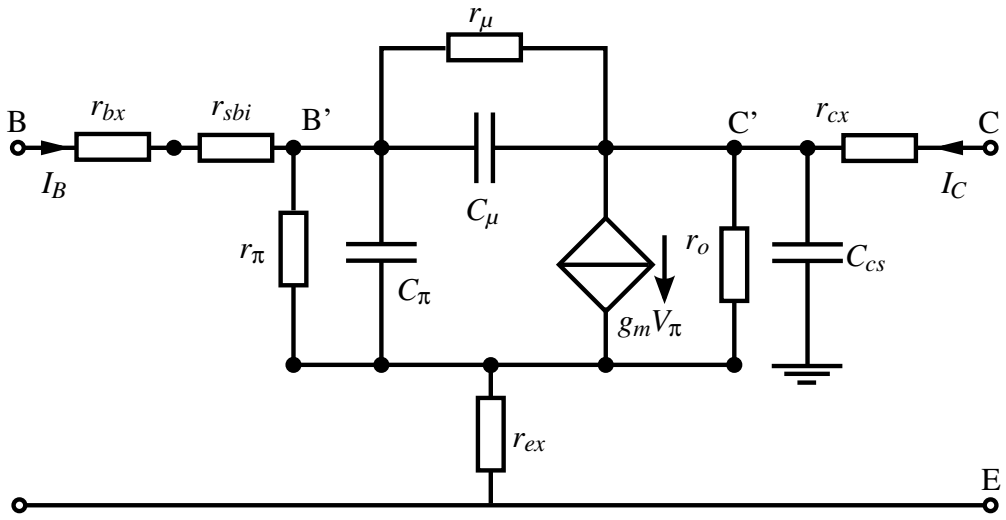


Figure G.4: Extended hybrid π-model for bipolar transistor modeling up to medium frequency

Hybrid π model parameters can be interpreted as actual transistor parameters as follows:

- $r_{bx} = r_{b'b} \Rightarrow$ extrinsic base resistance
- $r_{ex} \Rightarrow$ emitter series resistance
- $r_{cx} \Rightarrow$ extrinsic collector resistance
- $r_\pi = r_{b'e} \Rightarrow$ dynamic base-emitter resistance
- $r_o = 1/g_{ce}$ output resistance due to the Early effect, reciprocal output conductance
- $r_\mu = r_{b'c} \Rightarrow$ internal collector-base resistance (change in recombination component of I_B)
- $C_\pi = C_{b'e} \Rightarrow$ dynamic base-emitter capacitance (BE stored charge)
- $C_\mu = C_{b'c} \Rightarrow$ collector base transition plus diffusion capacitance (base width modulation)
- $C_\mu = C_{c's} \Rightarrow$ collector-substrate coupling capacitance
- $g_m V_\pi = g_m V_{b'e} \Rightarrow$ equivalent current generator
- $g_m \Rightarrow$ transconductance (reciprocal of g_m is the intrinsic resistance r_E)

The transfer characteristic is defined as:

$$i_C = I_S \left(1 + \frac{v_{CE}}{V_A}\right) \exp\left(\frac{v_{BE}}{V_T}\right) \tag{G.10}$$

The output characteristics is defined as:

$$i_C = \beta \left(1 + \frac{v_{CE}}{V_A}\right) i_B \tag{G.11}$$

The hybrid-π small-signal representation of the transconductance g_m, collector-to-emitter resistance r_0 and base-emitter resistance r_π are defined as follows:

$$g_m = \frac{\partial I_C}{\partial V_{BE}} = \frac{I_S}{V_T}\exp\left(\frac{V_{BE}}{V_T}\right) = \frac{I_C}{V_T} \tag{G.12}$$

$$r_0 = \left(\frac{\partial I_C}{\partial V_{CE}}\right)^{-1} = \left[\frac{I_S}{V_A}\exp\left(\frac{V_{BE}}{V_T}\right)\right]^{-1} = \frac{V_A + V_{CE}}{I_C} \tag{G.13}$$

$$r_\pi = \left(\frac{\partial I_B}{\partial V_{BE}}\right)^{-1} = \left[\frac{I_S}{\beta V_T}\exp\left(\frac{V_{BE}}{V_T}\right)\right]^{-1} = \frac{V_T}{I_B} \tag{G.14}$$

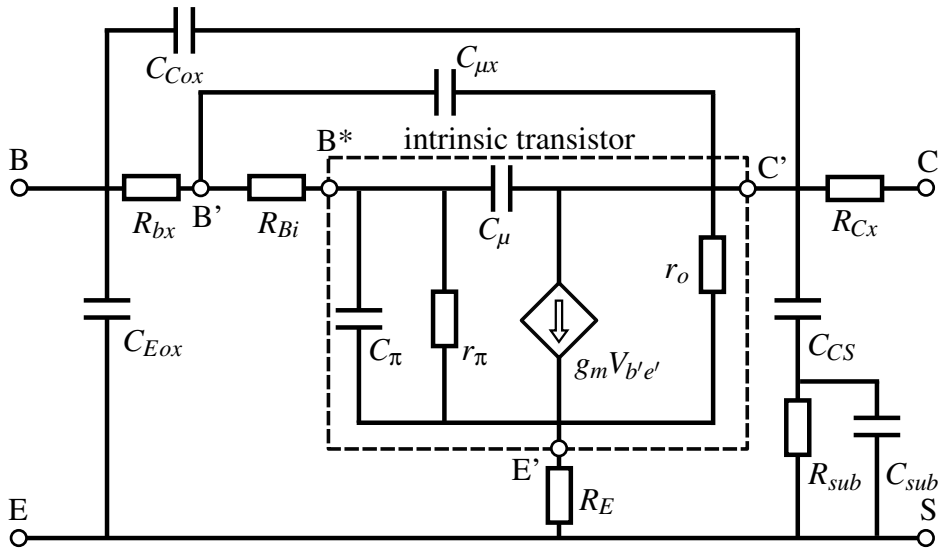


Figure G.5: Simplified equivalent circuit up to medium frequency range

G.4 Capacitance Extraction

For the particular case of BJT device related extraction one uses the hybrid-π circuit below.

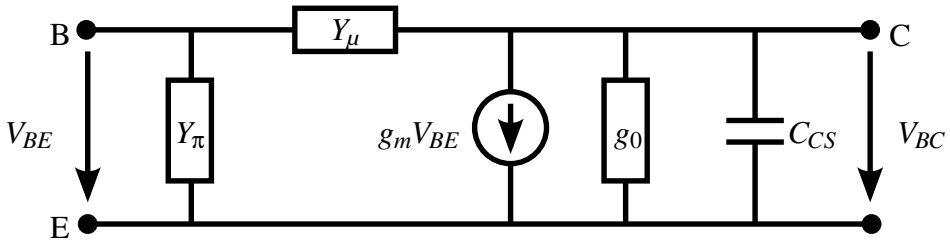


Figure G.6: Small signal representation of the BJT used for capacitances calculation

The according matrix representation are obtained from the equivalent circuit with their respec-

tive depletion (C_j) and diffusion (C_d) capacitance contribution:

$$\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} Y_\mu + Y_\pi & -Y_\mu \\ g_m - Y_\mu & Y_\mu + g_0 + j\omega C_{CS} \end{bmatrix} \tag{G.15}$$

$$Y_\mu = g_\mu + j\omega C_\mu, \quad \text{with } C_\mu = C_{jBC} + C_{dBC} \tag{G.16}$$

$$Y_\pi = g_\pi + j\omega C_\pi, \quad \text{with } C_\pi = C_{jBE} + C_{dBE}. \tag{G.17}$$

In off-state (reverse biased BE junction) the contributions g_m, g_μ, g_π and g_0 may be neglected and the capacitances are computed directly from [**y**]-parameters (cf. eqn. (G.6) through (G.7))

G.5 Scattering Parameters (S-Parameters)

Scattering parameters or two-port S-parameters defined through incident and reflected signal waves at the respective ports are best suited for measurement of high-frequency characteristics (MHz and GHz range) and hence very common in microelectronics. They are commonly employed when direct measurement of voltages and currents is not applicable and only incident and reflected power is available [166]. For a linear circuit, one can define a scattering matrix for the circuit in terms of the incident and reflected voltages at each of the circuit nodes.

S-parameters can be directly used to express electrical properties such as gain, insertion loss, return loss, the voltage standing wave ratio (VSWR, ratio of the maximum to the minimum amplitude of the standing wave), reflection coefficient ρ (absolute value of the magnitude of Γ) as well as amplifier stability. One defines waves travelling towards the two-port as a_i whereas waves travelling away from the two-port are denoted b_i. By definition of two-ports currents going into two-port networks are counted positively and currents flowing out of the n-port negatively.

The impedance Z_o is called the characteristic impedance of the network. Z_o may in general be any arbitrary reference impedance. In practice its value for most RF measurement devices is $Z_o = 50\Omega$. In general the S-parameters are complex and frequency dependent. The entries of the S-matrix may have a different formats such as real and imaginary part but also length and phase.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{G.18}$$

- S_{11} $\Rightarrow$ Forward Reflection (input match - impedance)
- S_{22} $\Rightarrow$ Reverse Reflection (output match - impedance)
  S_{11} and S_{22} are best viewed on a Smith chart.
- S_{21} $\Rightarrow$ Forward Transmission (gain or loss)
- S_{22} $\Rightarrow$ Reverse Transmission (leakage or isolation)

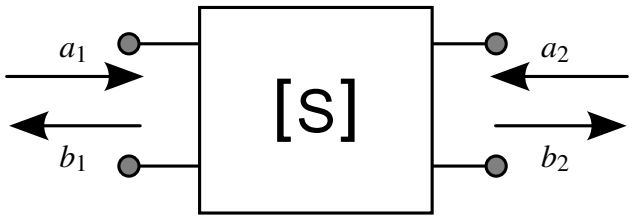


Figure G.7: Two Port network in S-Parameter representation

Scattering parameters or S-parameters are commonly used to describe a two-port network operating at high frequencies (RF and microwave frequency range). Compared to other representations the difference of the S-parameters is their description through normalized power waves when input and output ports are properly terminated [167]. The wave variables a1, b1 and a2, b2 are normalized forward and backward traveling waves. a_1 thus represents the incident wave at port 1 with b_1 being the corresponding reected wave. The same applies for the second port. For real measurement devices this means that a network analyzer measures the waves a1, b1 and a2, b2 at the generator and load where-from the waves at the inputs of the two-port can be determined.

Definitions of S_{ii}:

- $S_{11} = \frac{b_1}{a_1}\Big|_{a_2=0} \Rightarrow$ input Γ for output terminated with Z_o
- $S_{12} = \frac{b_2}{a_1}\Big|_{a_2=0} \Rightarrow$ forward transmission ratio with Z_o load
- $S_{21} = \frac{b_1}{a_2}\Big|_{a_1=0} \Rightarrow$ reverse transmission ratio with Z_o source
- $S_{22} = \frac{b_2}{a_2}\Big|_{a_1=0} \Rightarrow$ output Γ for input terminated with Z_o
- $|S_{21}|^2 \Rightarrow$ Transducer power gain with Z_o source and load

Definitions of Γ_L, Γ_s, Γ_{in} and Γ_{out}:

- $\Gamma_L = \frac{Z_L - Z_o}{Z_L + Z_o} \Rightarrow$ the reflection coefficient of the load
- $\Gamma_s = \frac{Z_s - Z_o}{Z_s + Z_o} \Rightarrow$ the reflection coefficient of the source
- $\Gamma_{in} = \frac{Z_{in} - Z_o}{Z_{in} + Z_o} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \Rightarrow$ the input reflection coefficient
- $\Gamma_{out} = \frac{Z_{out} - Z_o}{Z_{out} + Z_o} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \Rightarrow$ the output reflection coefficient

For a unilateral network one defines $S_{12} = 0$ with the consequences for in- and output reflection:

- $\Gamma_{in} = S_{11}$
- $\Gamma_{out} = S_{22}$

At a given frequency, the maximum gain that an amplifier can deliver is limited by either its G_{max}, $G_{T,max}$ or by a stability limit G_{MSG}. If the device is unilateral (a linear, lossless, reciprocal four-port network), or the S_{12} contribution is small enough to be ignored, the unilateral transducer gain (ratio between the magnitude of output and input signals) G_{TU} is simplified.

The unilateral figure of merit u then reads:

$$U = \frac{|S_{12}||S_{21}||S_{11}||S_{22}|}{(1-|S_{11}|^2)\cdot(1-|S_{22}|^2)} \tag{G.19}$$

$$= \frac{|Y_{21}-Y_{12}|^2}{4(\mathrm{Re}[Y_{11}]\,\mathrm{Re}[Y_{22}]-\mathrm{Re}[Y_{12}]\,\mathrm{Re}[Y_{21}])}. \tag{G.20}$$

G.6 Conversion between parameters

The following table summarizes the conversion between various forms of electrical 2-port network parameters as described in [168].

	$[\mathbf{z}]$	$[\mathbf{y}]$	$[\mathbf{h}]$
$[\mathbf{z}]$	$\begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix}$	$\begin{bmatrix} \frac{y_{22}}{\Delta[\mathbf{y}]} & \frac{-y_{12}}{\Delta[\mathbf{y}]} \\ \frac{-y_{21}}{\Delta[\mathbf{y}]} & \frac{y_{11}}{\Delta[\mathbf{y}]} \end{bmatrix}$	$\begin{bmatrix} \frac{\Delta[\mathbf{h}]}{h_{22}} & \frac{h_{12}}{h_{22}} \\ \frac{-h_{21}}{h_{22}} & \frac{1}{h_{22}} \end{bmatrix}$
$[\mathbf{y}]$	$\begin{bmatrix} \frac{z_{22}}{\Delta[\mathbf{z}]} & \frac{-z_{12}}{\Delta[\mathbf{z}]} \\ \frac{-z_{21}}{\Delta[\mathbf{z}]} & \frac{z_{11}}{\Delta[\mathbf{z}]} \end{bmatrix}$	$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{h_{11}} & \frac{-h_{12}}{h_{11}} \\ \frac{h_{21}}{h_{11}} & \frac{\Delta[\mathbf{h}]}{h_{11}} \end{bmatrix}$
$[\mathbf{h}]$	$\begin{bmatrix} \frac{\Delta[\mathbf{z}]}{z_{22}} & \frac{z_{12}}{z_{22}} \\ \frac{-z_{21}}{z_{22}} & \frac{1}{z_{22}} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{y_{11}} & \frac{-y_{12}}{Y_{11}} \\ \frac{y_{21}}{y_{11}} & \frac{\Delta[\mathbf{y}]}{y_{11}} \end{bmatrix}$	$\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$

Table G.1: Complex quadripole parameter conversion table for two-port network parameters at the same frequency into one another: Y (admittance), Z (impedance), h (hybrid)

For all matrix conversions given above Δ represents the determinant of the two-port network. In case of a 2x2 matrix this gives:

- $\Delta_Z = Z_{11}Z_{22} - Z_{12}Z_{21}$ being the determinant of the $[\mathbf{z}]$-parameter matrix,
- $\Delta_Y = Y_{11}Y_{22} - Y_{12}Y_{21}$ being the determinant of the $[\mathbf{y}]$-parameter matrix and
- $\Delta_h = h_{11}h_{22} - h_{12}h_{21}$ being the determinant of the $[\mathbf{h}]$-parameter matrix.

The conversion of S-parameters has to take into account the matching conditions of the characteristic impedance $Z_0 = 1/Y_0$, by which the S parameters are defined. The transformation of $[\mathbf{S}]$

matrix to Y-parameter representation is hence a little more complicated. The individual elements for the [**y**] matrix are obtained using the following formulations:

- $Y_{11} = \frac{((1-S_{11})(1+S_{22})+S_{12}S_{21})}{\Delta_S} Y_0$
- $Y_{12} = \frac{-2S_{12}}{\Delta_S} Y_0$
- $Y_{21} = \frac{-2S_{21}}{\Delta_S} Y_0$
- $Y_{22} = \frac{((1+S_{11})(1-S_{22})+S_{12}S_{21})}{\Delta_S} Y_0$

  with the determinant $\Delta_S = (1+S_{11})(1+S_{22}) - S_{12}S_{21}$

The other way around one might transform the Y matrix into S-parameter representation as follows:

- $S_{11} = \frac{(1-Z_0Y_{11})(1+Z_0Y_{22})+Z_0^2Y_{12}Y_{21}}{\Delta}$
- $S_{12} = \frac{-2Z_0Y_{12}}{\Delta}$
- $S_{21} = \frac{-2Z_0Y_{21}}{\Delta}$
- $S_{22} = \frac{(1+Z_0Y_{11})(1-Z_0Y_{22})+Z_0^2Y_{12}Y_{21}}{\Delta}$

  with the determinant $\Delta = (1+Z_0Y_{11})(1+Z_0Y_{22}) - Z_0^2Y_{12}Y_{21}$

G.7 R_B Extraction from RF Measurement Using the Circle Impedance Method

The approach presented in section 4.4.2 is suited for scalable extraction of sheet resistances using dedicated DC tetrode structures. However those are not actual transistor configurations used in circuits it is desirable to have a second method based on normal modeling devices in order to verify the obtained results. The modified circle impedance method using small-signal AC data presents such an approach. In contrast to the highly accurate parameter determination from direct resistivity measurement an interpolation to determine the value of R_B is used [98, 100].

The basic idea of the circle impedance method is a regression using the small-signal hybrid parameter (unit of impedance) h_{11}^* used to analyze small signal behavior of a transistor in the real and imaginary plane. h_{11}^* here is the input impedance with the output short circuited (cf. G.9).

$$Y_{11} = \frac{g_B(g_\pi + j\omega(C_\mu + C_\pi))}{g_B + g_\pi + j\omega(C_\mu + C_\pi)} + j\omega C_{BCx} \tag{G.21}$$

$$Y_{12} = -\frac{g_B \cdot j\omega C_{BCx})}{g_B + g_\pi + j\omega(C_\mu + C_\pi} - j\omega C_{BCx} \tag{G.22}$$

$$Y_{11} + Y_{12} = \frac{g_B(g_\pi + j\omega C_\pi)}{g_B + g_\pi + j\omega(C_\mu + C_\pi)} \tag{G.23}$$

$$h_{11}^* = \frac{1}{y_{11} + y_{12}} = (R_B + r_\pi)\left[\frac{1 + j\omega\frac{C_\mu + C_\pi}{g_B + g_\pi}}{1 + j\omega\frac{C_\pi}{g_\pi}}\right] = \frac{1}{g_{BE} + j\omega C_{BE}} + \frac{1}{g_B} \tag{G.24}$$

Per definition one obtains the total resistance of the two-port networks equivalent circuit including the emitter resistance R_E from interpolation towards infinite frequency ($f \to \infty$ canceling out the $1/j\omega C_{BE}$ term).

The progression of the modified hybrid parameter h_{11}^* versus frequency ($f \to \infty$) with sufficiently high collector current yields a circle with given center and radius

$$x_0 = R_B + \frac{r_\pi}{2} \text{ circle center} \tag{G.25}$$

$$r = \frac{r_\pi}{2} \text{ circle radius.} \tag{G.26}$$

The circle representation $(x - x_0)^2 + y^2 = r^2$ is resolved using imaginary and real part of the input impedance for

$$\Im\{h_{11}^*\} = \pm\sqrt{\left[\frac{1}{2 \cdot g_{BE}}\right]^2 - \left[\Re\{h_{11}^*\} - \left(\frac{1}{g_B} - \frac{1}{2 \cdot g_{BE}}\right)\right]^2}. \tag{G.27}$$

H The P-N Junction

A donor impurity atom is defined as donating an electron (n) to the conduction band (E_C) without creating a hole in the valence band. The resulting nomination is n-type material with its impurity atom concentration N_D. Similarly an acceptor impurity atom generates a hole in the valence band without generating an electron in the conduction band with material nomination p-type and impurity concentration N_A. semiconductor

The abrupt silicon p-n junction is created when a p and n layer are brought together. The electrostatic potential varies with the x-coordinate within a small region around the junction (x_j) from the p- to n-type semiconductor. Electrons and holes around the metallurgical junction diffuse across the junction into the inversely doped region where in consequence hardly any free electrons or holes are present, leaving the ionized atoms behind.

This creates a region around the junction, which is depleted of mobile carriers, the so called space charge region (SCR). The boundries are defined as x_p and x_n. The charge due to the ionized carriers causes an electric field E, which in turn causes a drift of carriers in the opposite direction. The diffusion of carriers continues until the drift current balances the diffusion current reaching the thermal equilibrium. Moreover, caused by this electric feld an internal (built-in) potential difference V_{bi} (qΦ_i respectively) occurs which equals the potential across the SCR in thermal equilibrium. Contrarily the electric field outside the depletion region is assumed to be zero.

An electrostatic analysis of a pn junction is of interest since it provides knowledge about the involved charges and the electric field in the depletion region. In addition it allows to derive the fundamental equation for capacitance-voltage characteristics of a diode.

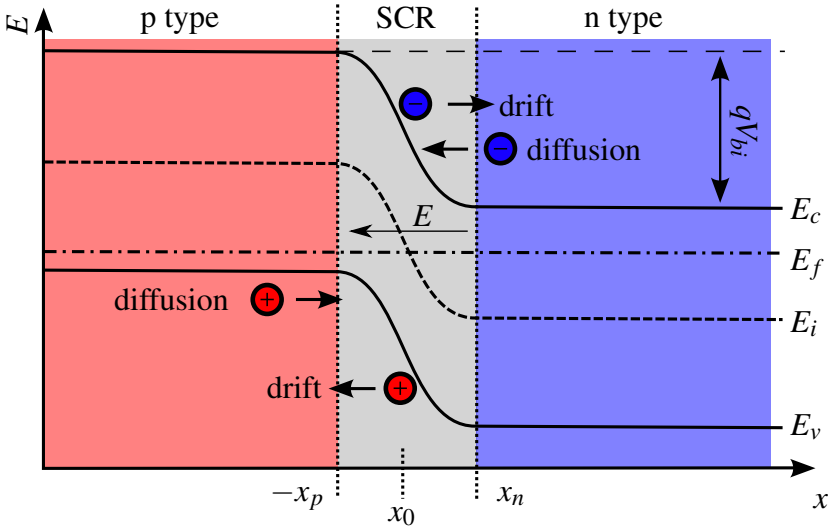


Figure H.1: Band diagram of a pn junction, *band bending as result of the different Fermi levels in the p and n region resulting in a depleted space charge region with built-in potential V_{bi}*

The total electrostatic potential difference between the p-side and the n-side neutral regions at thermal equilibrium is V_{bi} for an abrupt junction with $p = N_A$ from the formulation of the quasi-Fermi-level ψ_p as well as the respective formulation for the n-type region $n = N_D$ and ψ_n

$$\psi_p \equiv -\frac{1}{q}(E_i - E_F)\Big|_{x \leq -x_p} = -\frac{k_B T}{q} \ln \frac{N_A}{n_i} \text{ and} \tag{H.1}$$

$$\psi_n \equiv -\frac{1}{q}(E_i - E_F)\Big|_{x \geq x_n} = k_B T \ln \frac{N_D}{n_i} \tag{H.2}$$

$$V_{bi} = \psi_n - \psi_p = V_T \ln \frac{N_A N_D}{n_i^2} \text{ , with } V_T = \frac{k_B T}{q} \tag{H.3}$$

The pn-diode formed by the joint junctions can be biased by an external tension applied to the two terminals. A forward bias is defined as a positive voltage to the anode (p-type region) relative to the cathode (n-type region). In turn a reverse bias corresponds to a negative voltage. The applied voltage is added to the built-in voltage ($q\Phi_i$) and in consequence proportional to the difference between the Fermi energy in the neutral regions. The total potential across the junction equals the built-in potential minus the applied voltage

$$\Phi = \Phi_i - V_a. \tag{H.4}$$

A positive applied external voltage (V_a¿0) hence lowers the energy barrier whereas a negative voltage (V_a¡0) increases the barrier encountered by carriers.

For the calculation of junction capacitances the width of the SCR is critical. In unbiased thermal equilibrium, the total space charge must be zero and in turn the total p and n space charge on each side must be equal:

$$N_A x_p = N_D x_n \tag{H.5}$$

The total depletion region width ($d = x_n + x_p$) is then given by calculating the electric field of the SCR by

$$V_{bi} = \frac{q N_A x_p^2}{2\varepsilon_S} + \frac{q N_D x_n^2}{2\varepsilon_{Si}} \tag{H.6}$$

and from this, the depletion width d is given by

$$d = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right) V_{bi}}, \tag{H.7}$$

whereas for a biased junction one obtains in a similar manner

$$d = \sqrt{\frac{2\varepsilon_{Si}}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)(V_{bi} + V_a)}. \tag{H.8}$$

The capacitance C associated with the separation of carriers in the pn junction is obtained by calculating the change in charge for a change in applied voltage from

$$C = \frac{dQ}{dV}. \tag{H.9}$$

The total junction capacitance is calculated using the expression for the parallel plate capacitance $C_j = \varepsilon \cdot A/d$ as

$$C_j = \frac{\varepsilon \cdot A}{d} = \sqrt{\frac{q\varepsilon_{Si}}{2(\Phi_i - V_a)}\frac{N_A \cdot N_D}{N_A + N_D}}. \tag{H.10}$$

For modeling purposes the capacitance of a pn diode is frequently expressed as a function of its zero bias capacitance ($C_{j,0}$). Modifications to this equation might be necessary due to different material compositions (heterojunction) resulting in different permittivity (ε) requiring to account for series connection of the capacitance of each layer individually.

A common expression implemented in device modeling is

$$C_j = \frac{C_{j0}}{\left[1 - \frac{V}{V_{bi}}\right]^m} \tag{H.11}$$

where m is the grading coefficient with a default value for 0.5 for the case of the ideal abrupt pn junction.

I The Drift-Diffusion Model

The drift-diffusion model is applicable to a variety of different devices and is thus frequently employed to describe the operation principles of semiconductor devices. The classical drift-diffusion (DD) model is derived from Boltzmanns transport equation (BTE) for steady state conditions and is one of the cornerstones of modern bipolar device modeling.

It is the simplest current transport model and became the backbone of semiconductor device simulation expressing the electron current density as consisting of two components: the *drift* component is driven by the electric field (E) and the *diffusion* component by the electron density gradient [169, 170].

The underlying assumptions of the simplified drift-diffusion model are [26]:

- complete ionization[1]
- non-degenerate semiconductor (moderate doping levels)[2]
- steady-state (time-independent)[3]
- temperature is constant throughout the device

All current commercially available bipolar models are based on a set of equations derived for one-dimensional (1D) carrier transport. The system of fundamental equations may hence be limited to the 1D case with the following set of classical device equations tailored to describe the whole simulation domain of a semiconductor device. With the Poisson equation and the continuity equations for electrons and holes as well as the drift-diffusion current relations for electron- and hole-current a complete set of equations which can be seen as fundamental for the simulation of semiconductors.

Current equations

There are two effects leading to current flow in silicon. First, the drift of charged carriers due to the influence of an electric field, and second, the diffusion current due to a concentration gradient. Combining the current contributions of the drift and the diffusion effect we get the drift-diffusion

[1] dopants are assumed to be ionized; formation of ions as an atom loses the associated electron (for n type, negative charge), the atom is ionized and is positively charged; consequently a doped semiconductor contains free carriers if the impurities are ionized

[2] The Fermi energy E_f is at least $3k_BT$ from the band edge (E_V or E_C); The product of the electron (n) and hole (p) density of a non-degenerate semiconductor is always equal to the square of its intrinsic carrier density (n_i), whether the semiconductor is intrinsic (contains no impurities, $E_F = E_i$) or extrinsic (doped); allows the Fermi-dirac statistics (describing probability of occupancy versus energy) to be replaced by MaxwellBoltzmann statistics can be used as an approximation resulting in a more simple exponential function where the carrier density integral can then be solved analytically

[3] when no external source is connected to the junction, diffusion and drift balance each other out for both holes and electrons and D is a time-independent background charge density

(DD) current relations [171]:

$$J_n = qn(x)\mu_n E(x) + qD_n \frac{dn}{dx} = qn\mu_n \xi + qD_n \nabla n \tag{I.1}$$

$$J_p = qp(x)\mu_p E(x) - qD_p \frac{dn}{dx} = qp\mu_p \xi - qD_p \nabla p \tag{I.2}$$

where

$$D_n = \frac{k_B T}{q}\mu_n \text{ and } D_p = \frac{k_B T}{q}\mu_p \tag{I.3}$$

Continuity Equations

The Ampere-Maxwell law reflects how an electric current ($\vec{J}$) and the change in the electric field ($\vec{E}$) produce a magnetic field ($\vec{H}$) by $\vec{\nabla} \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}$, where $\partial/\partial t$ is the partial derivative w.r.t. time. Using the divergence operator ($\vec{\nabla}$) one can separate the total current density $\vec{J}$ into hole and electron current densities ($\vec{J} = \vec{J}_p + \vec{J}_n$) and obtains the continuity equations for electrons and holes with U representing net recombination rate ($U = R - G$ generation and recombination)[1].

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot J_n + U_n \tag{I.4}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot J_p + U_p \tag{I.5}$$

Poisson's equation

The Poisson equation relates the electrostatic potential Φ to a given charge distribution (ρ is the electric charge density). It can be derived from Maxwell's equation ($\vec{\nabla} \cdot \vec{D} = \rho$) using the relation between the electric displacement vector ($\vec{D}$) and the electric field vector($\vec{E}$). The permittivity is obtained from the relative ε_r and the vacuum permittivity ε_0 as $\varepsilon = \varepsilon_r \cdot \varepsilon_0$.

$$\vec{D} = \hat{\varepsilon} \cdot \vec{E} \tag{I.6}$$

$$\nabla \cdot \varepsilon \nabla V = -\rho = -q(p - n + N_D^+ - N_A^-) \tag{I.7}$$

[1] In semiconductors supplied with an appropriate impulse, valence band electrons will be lifted into the conduction band thus generating pairs of electrons and holes. The rate G describes the number of generated electron-hole pairs per unit time and volume. The opposite process to generation is called recombination with the characteristic rate R that indicates the number recombined electron-hole pairs per unit time and volume.

J The HICUM model

HICUM (HIgh CUrrent bipolar compact transistor Model) is an advanced transistor model for bipolar transistors with main emphasis on circuit design for high-speed applications. The model has been developed and continuously improved more than 20 years. Work started about 1982 at Ruhr-University Bochum, Germany, with major emphasis on high-speed ECL type circuits for fibre-optic applications. The model has been extended to high-speed small-signal applications since 1989 as well as to SiGe HBTs since 1993. HICUM development resulted from the experience that the SPICE Gummel-Poon model (SGPM) is not accurate enough for high-speed large-signal transient applications and the required high collector current densities. Other major disadvantages of the SGPM are lack of sufficient physical background, poor description of base resistance and (quasi-)saturation. Major features of HICUM (and advantages over the SGPM) are [172]:

- Accurate description of the high-current operating region (including quasi-saturation and saturation)
- Distributed modelling of external base-collector region
- Emitter periphery injection and charge storage is taken into account
- Internal base resistance is dependent on operating point (conductivity modulation AND emitter current crowding), as well as emitter geometry
- Sufficiently physical model equations allowing predictions of temperature and process variations as well as scalability even at high current densities
- Parasitic capacitances, independent on operating point, are available in the equivalent circuit, representing base-emitter and base-collector oxide overlaps, that become significant for small-size transistors
- Weak avalanche breakdown is available
- Self-heating effects can be simulated for d.c., AC and transient operation
- Non-quasi-static effects, resulting in a delay of collector current and stored minority charge, are modelled consistently as function of bias for small-signal and large-signal operation
- Collector current spreading is included in minority charge and collector current formulation
- Extensions for graded-base SiGe HBTs have been derived using the Generalized Integral Charge-Control Relation (GICCR); the GICCR also permits modelling of HBTs with bandgap difference within the junction, as long as thermionic emission can be neglected
- Base-emitter tunneling model is available (used for reverse bias leakage current simulation)
- Simple parasitic substrate transistor is included in the equivalent circuit
- Sufficiently simple parallel RC network taking into account the frequency dependent coupling between collector and substrate terminal
- Parameter extraction is closely related to the process enabling parametric yield simulation;

parameter extraction procedure and list of test structures are available; HICUM parameters can be determined using standard measurement equipment and mostly simple extraction procedures
- Simple equivalent circuit and numerical formulation of model equations result in easy implementation and relatively fast execution time

These features together with the choice of easily measurable capacitances and transit time yield high accuracy, compared to the GP model, ranging from DC, small-signal and, high-speed large-signal transient simulation. Given the strong physical background HICUM is laterally scalable over a range of transistor configurations up to very high collector current densities. To make the model applicable to HBT device physics, the major changes from version L2.1x to L2.2x included (but were not limited to):

- Temperature dependent band-gap voltage ($V_g(T)$) for devices fabricated in different types of materials
- Improved transfer current equations (reach-through, exponent factor of the temperature dependence (ζ_{CT}))
- Base current components such as recombination at the BC barrier and temperature dependent junction current
- Smoothing of depletion capacitances and charges and temperature dependence of associated built-in voltages ($V_{Dj}(T)$)
- Minority charge formulation with effective collector voltage (v_{ceff}), critical current (i_{CK}) and current spreading
- Base-emitter tunneling (i_{BEt})
- Base-emitter capacitance partitioning for parasitics (C_{BEpar})
- Substrate transistor temperature dependence ($I_{TS}(T)$)

The latest release of HICUM/L2 version L2.30 has been developed within the framework of the DotFIVE project. The model improvement over the former release version L2.24 primarily focused on the applications for emerging mmW markets [116].

K The Terahertz Gap

The so called Terahertz gap describes the band of frequencies in the electromagnetic spectrum that lies between radio waves (Electronics) and the infrared light (Photonics). For this application range practical technologies for generation and detection of the radiation did not exist for a long time (frequencies generated by transistors and lasers did not overlap, hence electrical power could not be converted into electromagnetism in that range) [173].

However driven by possible biological and medical applications[1] during the last years, research devoted an important part of their activity to solve problems related to THz generation and detection in semiconductor systems. In consequence the cut off frequency of transistors was pushed up to a few hundreds of GHz and the THz limit became realistic for electronic devices [174] [175].

Given this advancement with technology the range is defined differently and precise values vary depending on literature sources. A reasonable range today is narrowed down the limits of 0.3 to 3 THz, which is the equivalent of a wavelength spectrum (λ) of $1mm$ to $100\mu m$ (under vacuum condition $v = c$).

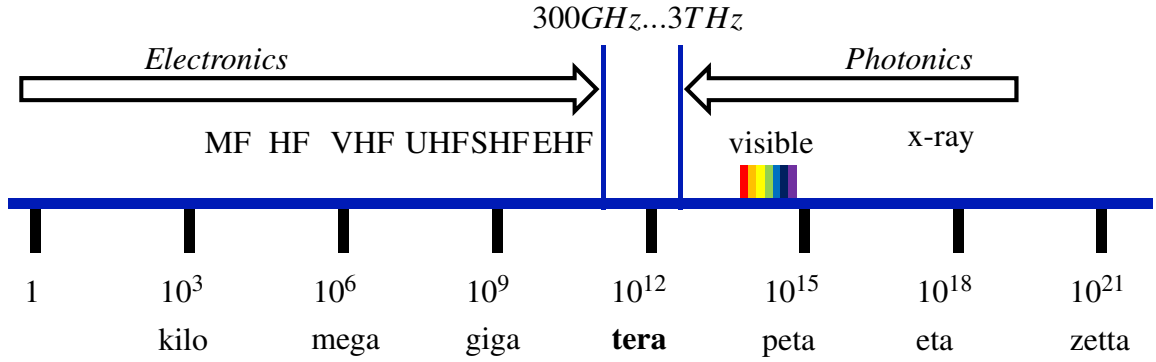


Figure K.1: Graphical illustration of the terahertz gap

[1] Terahertz waves are safe to use because they are non-ionizing and can pass through skin, plastic, wood and ceramics as well as fog and clouds, but cannot penetrate metal, concrete or water

L Sample Input File for Numerical Device Simulation with DEVICE

```
NAME 'B9MW adapted'
&SETTINGS lang_vers='1.90' num_dig=12 max_exp=300 tcpu_lim=1e9 /
&TEMP_SWEEP T_chip=300.0 T_par=300.0 /
STRUcture
 &GEN_INFO spat_dim=2 stru_file=' ' /
 &REGION_DEF reg_mat='SEMI' mod_name='SILI' low_xyz=0.0     0.0
    upp_xyz=0.3     1.0 /
 &REGION_DEF reg_mat='CONT' mod_name='CON0' low_xyz=0.0     0.0
    upp_xyz=0.0     1.0  cont_name='E ' /
 &REGION_DEF reg_mat='SUPP' mod_name='SUPO' low_xyz=0.0284  0.0
    upp_xyz=0.0286  0.0  cont_name='BL' /
 &REGION_DEF reg_mat='SUPP' mod_name='SUPO' low_xyz=0.0286  1.0
    upp_xyz=0.0286  1.0  cont_name='BR' /
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz=0.0     0.0
    upp_xyz=0.0284  0.0 /
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz=0.0286  0.0
    upp_xyz=0.3     0.0 /
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz=0.0     1.0
    upp_xyz=0.0284  1.0 /
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz=0.0286  1.0
    upp_xyz=0.3     1.0 /
 &REGION_DEF reg_mat='CONT' mod_name='CON0' low_xyz=0.3     0.0
    upp_xyz=0.3     1.0  cont_name='C ' /
.STR
DISC
&AUTO_GRID x_pnts=390 y_pnts=2 gen_type='keep' int_pol='spline'
   pot_ratio=1.0 /
&RANGE_GRID disc_dir='x' intv_pnts=
0.0         0.008          30
0.0080001   0.08       250
0.0800001   0.14           80
0.1400001   0.3        30  /
.DIS
```

```
 MODELs
&CONTACT mod_name='CON0' v_con=0.0    zeta_vcon=0.5 /
&SUPPLY  mod_name='SUPO' cont_dist=0.0 cont_curr=0.0 /
&SEMICON mod_name='SILI' bgap_vg0=1.1241 bgap_at1=0 bgap_at2=0
    bgap_fvgt=0
bgap_vhd0=6.92e-3 bgap_alhd=1.43e-3 bgap_chd=0.5 bgap_dophd=1.3
    e17 bgap_gamhd=0.5
rmn_eff=1.0957 rmp_eff=0.91973 dens_stat=2.509e19
srhn_tmin=0 srhn_tmax=9e-6 srhn_dop=2.5e15 srhn_beta=0.5
    srhn_zeta=0
srhp_tmin=0 srhp_tmax=3e-6 srhp_dop=2.5e15 srhp_beta=0.5
    srhp_zeta=0
augn_c=2.2e-31 augn_zeta=0 augp_c=0.99e-31 augp_zeta=0 user_reco
    =0
srhp_js=0 srhp_j0=0 srhp_ca=0 srhp_c0=0
mobn_0=80 mobn_max=1430 mobn_dop=8.6e16 mobn_beta=0.77 mobn_alph
    =-2.5 mobn_bett=0.35 mobn_betf=1.4
mobn_hd=0.0 mobn_doph=3.43e20 mobn_beth=2.0 mobn_rmin=3.1
    mobn_ecr=0 user_mobn=0
mobp_0=47 mobp_max=480 mobp_dop=1.7e17 mobp_beta=0.79 mobp_alph
    =-2.5 mobp_bett=0.35 mobp_betf=1.0
mobp_hd=0 mobp_doph=6.1e20 mobp_beth=2.0 mobp_rmin=2.6021
    mobp_ecr=0 user_mobp=0
vsn_0=1.07e7 vsn_al=0 vsn_zeta=0 vsp_0=0.837e7 vsp_al=0 vsp_zeta
    =0 ehsc_al=1
elaf_0=4.05 elaf_al=0 eps_rel=11.7 tun_fbtb=0 tun_bbtb=0
    tun_gbtb=0
avaln_al=0 avaln_E=1.231e6 avaln_ebet=0 avaln_hdam=0 avaln_hdbw
    =0 avaln_hdct=0
avalp_al=0 avalp_E=2.036e6 avalp_ebet=0 avalp_hdam=0 avalp_hdbw
    =0 avalp_hdct=0
bgap_avg=0 /
```

```
&MAT_COMP mat_type='GE' bulk_mat='SILI'
 bgap_a1=-0.896 bgap_a2=0.396 mefn_a1=-3.3274 mefn_a2=10.963
    mefp_a1=-3.9185 mefp_a2=8.8025
 srhn_mna1=0 srhn_mna2=0 srhn_mxa1=0 srhn_mxa2=0 augn_a1=0
    augn_a2=0
 srhp_mna1=0 srhp_mna2=0 srhp_mxa1=0 srhp_mxa2=0 augp_a1=0
    augp_a2=0
 elaf_a1=-0.15 elaf_a2=-0.02 epsr_a1=3.133344 epsr_a2=1.1665
 mobn_0a1=-20 mobn_0a2=-80 mobn_mxa1=-1879 mobn_mxa2=-400 vsn_0a1
    =-3.5e7 vsn_0a2=6.8e7
 mobp_0a1=-20 mobp_0a2=300 mobp_mxa1=-780 mobp_mxa2=4200 vsp_0a1
    =0 vsp_0a2=0
 mefn_c=0 mefp_c=0 srhn_tc=0 srhp_tc=0 augn_cc=0 augp_cc=0
 mobn_latt=0 mobn_zetal=0 mobp_latt=0 mobp_zetal=0 /
.MOD

PROFile
&GEN_INFO don_solub=4E20 acc_solub=4E20 /
* Emitter
&EXP_PROD d_max=-1.767e20 xyz_max=0.0 0   a_xyz=0.0099 0
  beta_xyz=2.02 0  xyz_low=0.0 -0.1       xyz_upp=0.0105 2.0 /
&EXP_PROD d_max=-0.5e20 xyz_max=0.0105 0  a_xyz=0.00262 0
  beta_xyz=1.95 0  xyz_low=0.0105001 -0.1 xyz_upp=2.0 2.0 /
* Neutral Collector
&EXP_PROD d_max=-3.45e17 xyz_max=0.038 0  a_xyz=0.006 0
  beta_xyz=1.65 0  xyz_low=-0.1 -0.1 xyz_upp=0.038 2.0 /
&EXP_PROD d_max=-5.7e19 xyz_max=0.4 0  a_xyz=0.071 0
  beta_xyz=1 0  xyz_low=0.0380001 -0.1 xyz_upp=2.0 2.0 /
&EXP_PROD d_max=-3.40e18 xyz_max=0.14 0  a_xyz=0.045 0
  beta_xyz=2.05 0  xyz_low=0.03 -0.1 xyz_upp=0.14 2.0 /
&EXP_PROD d_max=-3e19 xyz_max=0.4 0  a_xyz=0.12 0
  beta_xyz=1 0  xyz_low=0.14001 -0.1 xyz_upp=2.0 2.0 /
* Base
&EXP_PROD d_max=2.57e19 xyz_max=0.0285 0  a_xyz=0.0062 0
  beta_xyz=2.0 0  xyz_low=-0.1 -0.1 xyz_upp=2.0 2.0 /
&EXP_PROD d_max=6e16 xyz_max=0.0285 0  a_xyz=0.0082 0
```

```
  beta_xyz=1.95 0  xyz_low=-0.1 -0.1 xyz_upp=0.0285 2.0 /
&EXP_PROD d_max=1e14 xyz_max=0.4 0  a_xyz=0.29 0
  beta_xyz=1 0  xyz_low=0.0285001 -0.1 xyz_upp=2.0 2.0 /
* Germanium
&COMP_EXP mat_type='GE' c_max=0.25 xyz_max=0.038 0
      a_xyz=0.004  0.00 beta_xyz= 3.7103 0.00
      xyz_low=0.038001 -1 xyz_upp=2 2/
&COMP_EXP mat_type='GE' c_max=0.09 xyz_max=0.0195 0
      a_xyz=0.006  0.00 beta_xyz= 3.9 0.00
      xyz_low=0.0 -1 xyz_upp=0.0195 2/
&COMP_EXP mat_type='GE' c_max=0.16 xyz_max=0.031 0
      a_xyz=0.008  0.00 beta_xyz= 2.5 0.00
      xyz_low=0.0 -1 xyz_upp=0.031 2/
&COMP_LIN mat_type='GE' c_max=0.09 xyz_d=0 0
      xyz_low=0.019501 -1 xyz_upp=0.038 2/
&COMP_LIN mat_type='GE' c_max=-0.045 xyz_d=0 0
      xyz_low=-1 -1 xyz_upp=2 2/
&COMP_LIN mat_type='GE' c_max=0.16 xyz_d=0 0
      xyz_low=0.031001 -1 xyz_upp=0.038 2/
&COMP_LIN mat_type='GE' c_max=-0.08 xyz_d=0 0
      xyz_low=-1 -1 xyz_upp=2 2/
&COMP_TRAP mat_type='GE' c_max=0.02
      xyz_low=0.040 -1 xyz_upp=0.042 2
          xyz_dlow=0.002 0 xyz_dupp=0 0
          xyz_rlow=0 0 xyz_rupp=0 0/
&COMP_EXP mat_type='GE' c_max=0.02 xyz_max=0.042 0
      a_xyz=0.003  0.00 beta_xyz= 3.9 0.00
      xyz_low=0.0420001 -1 xyz_upp=2 2/
.PRO
```

```
ANALYZE
SOLUtion
  &SEMICON iter_max=1000 rtol_pot=1.0E-8 atol_pot=1.0E-8 rsum_cur
     =0.0 rho_pn=-100 /
  &SOLVE equa_name='poisson' solu_sequ=1 solu_meth='LUFA'
         iter_solv=300 dpot_lim=0.03 damp_newt=1
         iter_max=500/ orde_meth=''
  &SOLVE equa_name='cont_hole' solu_sequ=1 iter_max=500 dpot_lim
     =0.05
         damp_newt=1.0
         solu_meth='LUFA' orde_meth=0 iter_solv=300 relx_fac=1.0
            rtol_var=0.01 /
  &SOLVE equa_name='cont_elec' solu_sequ=1 iter_max=500 dpot_lim
     =0.05
         damp_newt=1.0
         solu_meth='LUFA' orde_meth=0 iter_solv=300 relx_fac=1.0
            rtol_var=0.01 /
.SOL

TCPU 10000000

DC
  &BIAS_INFO cont_name='E'  bias_fun='TAB' bias_val=0.0 /
  &BIAS_INFO cont_name='BL' bias_fun='TAB' bias_val=0.0 /
  &BIAS_INFO cont_name='BR' bias_fun='TAB' bias_val=0.0 /
  &BIAS_INFO cont_name='C'  bias_fun='TAB' bias_val=0.0 /
*  &QUAS_BIAS quas_cont='E C BL&BR' quas_val=1.0 /
  OUTPut
   &GEN_INFO algo_info=1 /
   &SPATIAL_DATA file_name='0V' num_start=0 data_format='WAVE'
          slice_dir='' var_list='*' /
  .OUT
RUN

END
```

M Sample Input File for Numerical Device Simulation of Resistance Structure (R_{Bx})

```
'NAME' 'RB_Simulation SYMMETRIC'
'*'
&SETTINGS lang_vers='1.80' num_dig=12 max_exp=300 tcpu_lim=1e6 /
'OPTIons' 1.0E-12 300.0 1E6
'*'
&SETTINGS lang_vers='1.90'/
STRUcture
 &GEN_INFO spat_dim=2 stru_file=' ' /
* TOTAL SIZE:   y(nm) = 0.234, 0.261, 0.301, 0.391, ...
* BASE CURRENT GENERATION BY RECOMBINATION
 &REGION_DEF reg_mat='CONT' mod_name='CON0' low_xyz= 0.134 0.346
    upp_xyz= 2.334 0.346 cont_name='B' / int #1
* silicided
 &REGION_DEF reg_mat='SEMI' mod_name='SILA' low_xyz= 0.000 0.000
    upp_xyz= 2.334 0.234 / left silicided #2
 &REGION_DEF reg_mat='SEMI' mod_name='SILA' low_xyz= 0.000 0.234
    upp_xyz= 0.067 0.346 / top silicided #3
* BASE PERIMETER LINK
 &REGION_DEF reg_mat='SEMI' mod_name='SILC' low_xyz= 0.067 0.234
    upp_xyz= 2.334 0.261 / left silicided #4
 &REGION_DEF reg_mat='SEMI' mod_name='SILC' low_xyz= 0.067 0.261
    upp_xyz= 0.094 0.346 / top silicided #5
* BASE PERIMETER SPACER
 &REGION_DEF reg_mat='SEMI' mod_name='SILD' low_xyz= 0.094 0.261
    upp_xyz= 2.334 0.301 / base perim  left spacer #6
 &REGION_DEF reg_mat='SEMI' mod_name='SILD' low_xyz= 0.094 0.301
    upp_xyz= 0.134 0.346 / base perim  top spacer #7
* EMITTER
 &REGION_DEF reg_mat='SEMI' mod_name='SILB' low_xyz= 0.134 0.301
    upp_xyz= 2.334 0.346 / base perim base #8
* FOXI & BASE CONTACT
 &REGION_DEF reg_mat='CONT' mod_name='CON0' low_xyz= 0.000 0.000
    upp_xyz= 2.334 0.000 cont_name='A' / int #9
```

```
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz= 0.000 0.000
    upp_xyz= 0.000 0.346 / top #10
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz= 2.334 0.000
    upp_xyz= 2.334 0.346 / bottom #11
 &REGION_DEF reg_mat='FOXI' mod_name='FOXI' low_xyz= 0.000 0.346
    upp_xyz= 0.134 0.346 / right #12
.STR
DISC
 &AUTO_GRID x_pnts= 40 y_pnts=10 gen_type='keep' int_pol='lin'
    pot_ratio=1.0 /
.DIS
&SETTINGS lang_vers='1.80'/
'MODELS'
 'SEMI' 'SILA' Silicide          15 Ohm/Square
  'PMOB' 41667 41667 3.2e17 0.63 -0.35 2.5 0.0 0.0 6.1e20 2.0 1.0
  'PSRH' 0.0 0.0 1.0e30 0.0
  'NSRH' 0.0 0.0 1.0e30 0.0
'*' BASE PERIM
 'SEMI' 'SILC' Link: 400 Ohm/Square
  'PMOB' 156.24 156.24 3.2e17 0.63 -0.35 2.5 0.0 0.0 6.1e20 2.0
     1.0
  'PSRH' 0.0 0.0 1.0e30 0.0
  'NSRH' 0.0 0.0 1.0e30 0.0
'*' BACKUP
 'SEMI' 'SILD' Poly-on-mono: 6000 Ohm/Square
  'PMOB' 10.417 10.417 3.2e17 0.63 -0.35 2.5 0.0 0.0 6.1e20 2.0
     1.0
  'PSRH' 0.0 0.0 1.0e30 0.0
  'NSRH' 0.0 0.0 1.0e30 0.0
'*' BASE CURRENT INJECTION
 'SEMI' 'SILB' Emitter:          1 Ohm/Square
  'PMOB' 625000 625000 3.2e17 0.63 -0.35 2.5 0.0 0.0 6.1e20 2.0
     1.0
  'PSRH' 0.0 0.0 1.0e30 0.0
  'NSRH' 0.0 0.0 1.0e30 0.0
 'CONT' 'CON0' 0.0 0.0
```

```
'.MOD'
'*'
'PROFILE'
  'CON' 1.0e17 0.0 15.0 0.0 15.0
'.PRO'
'*'
'ANALyze'
 'TEMP' 300
 'TCPU' 36000
'*'
'SOLUTION'
 'SEMI'
  'TOL' 1.0e-6 1.0e-11 0.0
  'ITER' 100000
  'SIMU' -10
  'POISSON' 1
   'ORDE' 2
   'SLOR' 10 -1.1 0.01
   'LIMI' 15 0.5 1.0
  'ELEC' 0
   'HOLE' 1
 '.SEM'
'.SOL'
'*'
'DC'
 'BIAS' 'A' 'TAB' 1 0.01
 'BIAS' 'B' 'TAB' 1 0.00
 'OUTPut' 'WAVE'
 'ALGO' 1
 'STOR' 'Dfile1' 1 400
 'PRIN' '*'
 'PLOF' 'PLO0'
 '.OUT'
'RUN '
'END'
```

N Sample Input File for Electrical Field Solver POICAPS

```
&set dim=2 num_cores=4 dbl_binary=T solve_poi=T len_unit='nm' /
&box btype='pass' x_s=0.000000 x_e=200.000000 y_s=0.000000
            y_e=10.0000 eps_r=1    pot=1.000000 keep_pot=T /
&box btype='pass' x_s=200.0000 x_e=550.000000 y_s=0.000000
            y_e=5.00000 eps_r=1    pot=1.000000 keep_pot=T /
&box btype='pass' x_s=550.0000 x_e=900.000000 y_s=0.000000
            y_e=32.0000 eps_r=1    pot=1.000000 keep_pot=T /
&box btype='pass' x_s=890.0000 x_e=900.000000 y_s=32.00000
            y_e=563.000 eps_r=1    pot=1.000000 keep_pot=T /
&box btype='pass' x_s=0.000000 x_e=200.000000 y_s=216.0000
            y_e=306.000 eps_r=1    pot=0.000000 keep_pot=T /
&box btype='pass' x_s=200.0000 x_e=500.000000 y_s=221.0000
            y_e=301.000 eps_r=1    pot=0.000000 keep_pot=T /
&box btype='pass' x_s=500.0000 x_e=550.000000 y_s=167.0000
            y_e=563.000 eps_r=1    pot=0.000000 keep_pot=T /
&box btype='pass' x_s=550.0000 x_e=890.000000 y_s=32.00000
            y_e=563.000 eps_r=3.9 keep_pot=F /
&box btype='pass' x_s=477.0000 x_e=550.000000 y_s=5.000000
            y_e=167.000 eps_r=3.9 keep_pot=F /
&box btype='pass' x_s=477.0000 x_e=500.000000 y_s=167.0000
            y_e=221.000 eps_r=7.8 keep_pot=F /
&box btype='pass' x_s=200.0000 x_e=477.000000 y_s=5.000000
            y_e=221.000 eps_r=3.9 keep_pot=F /
&box btype='pass' x_s=0.000000 x_e=200.000000 y_s=10.00000
            y_e=216.000 eps_r=3.9 keep_pot=F /
&box btype='pass' x_s=200.0000 x_e=477.000000 y_s=301.0000
            y_e=563.000 eps_r=3.9 keep_pot=F /
&box btype='pass' x_s=477.0000 x_e=500.000000 y_s=301.0000
            y_e=563.000 eps_r=7.8 keep_pot=F /
&box btype='pass' x_s=0.000000 x_e=200.000000 y_s=306.0000
            y_e=563.000 eps_r=3.9 keep_pot=F /
&inp fname='BC_B55_dBE' /
&out fname='BC_B55_dBE' cquant='all' /
end
```

References

[1] S. Iyer, G. Patton, S. S. Delage, S. Tiwari, and J. M. C. Stork, "Silicon-germanium base heterojunction bipolar transistors by molecular beam epitaxy," in *Electron Devices Meeting, 1987 International*, vol. 33, 1987, pp. 874–876.

[2] D. Harame and B. Meyerson, "The early history of ibm's SiGe mixed signal technology," *Electron Devices, IEEE Transactions on*, vol. 48, no. 11, pp. 2555–2567, Nov 2001.

[3] Project profile on CATRENE call CT209, "From RF to mmw and thz silicon system-on-chip technologies [RF2THz SiSoC]," 2011.

[4] "Dotfive: Towards 0.5 terahertz silicon/germanium heterojunction bipolar technology," 2010, a European Integrated Project supported through the Seventh Framework Programme for Research and Technological Development.

[5] "Dotseven: Towards 0.7 thz silicon-germanium heterojunction bipolar technology (0.7 thz sige hbt)," 2012, the DOTSEVEN project addresses the area More than Moore, Beyond (and ahead) of CMOS, to develop SiGe heterojunction bipolar transistors operating at maximum oscillation frequencies of up to 700 GHz at room temperature.

[6] D. Knoll, "Bicmos integration of photonic components," in *Open Bipolar Workshop, Bordeaux*, 2013.

[7] P. Chevalier, "A 55-nm bicmos platform for optical and millimeter-wave systems-on-chip," in *Open Bipolar Workshop, Bordeaux*, 2013.

[8] G. Avenier, P. Chevalier, G. Troillard, B. Vandelle, F. Brossard, L. Depoyan, M. Buczko, S. Boret, S. Montusclat, A. Margain, S. Pruvost, S. Nicolson, K. Yau, D. Gloria, D. Dutartre, S. Voinigescu, and A. Chantre, "0.13 um SiGe bicmos technology for mm-wave applications," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2008. BCTM 2008. IEEE*, oct. 2008, pp. 89–92.

[9] D. Celi, "About modeling the reverse early effect in hicum level 0," in *6th European HICUM Workshop*, 2006.

[10] A. Pawlak, M. Schrter, and J. Krause, "A hicum extension for medium current densities," in *9th European HICUM Workshop, Wuerzburg*, October 2009.

[11] Z. Huszka, D. Celi, and E. Seebacher, "A novel low-bias charge concept for HBT/BJT models including heterobandgap and temperature effects-part i: Theory," *Electron Devices, IEEE Transactions on*, vol. 58, no. 2, pp. 348–356, feb. 2011.

[12] ——, "A novel low-bias charge concept for HBT/BJT models including heterobandgap and temperature effects-part ii: Implementation, parameter extraction and verification," *Electron Devices, IEEE Transactions on*, vol. 58, no. 2, pp. 357–363, feb. 2011.

[13] M. Schroter, A. Mukherjee, and A. Pawlak, *HICUM/L2 version 2.30: Release Notes*, 2011.

[14] J.-S. Rieh, D. Greenberg, A. Stricker, and G. Freeman, "Scaling of SiGe heterojunction bipolar transistors," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1522–1538, sept. 2005.

[15] J. Dunn, D. Ahlgren, D. Coolbaugh, N. B. Feilchenfeld, G. Freeman, D. Greenberg, R. Groves, F. Guarin, Y. Hammad, A. Joseph, L. Lanzerotti, S. St.Onge, B. Orner, J.-S. Rieh, K. Stein, S. Voldman, P.-C. Wang, M. Zierak, S. Subbanna, D. Harame, D. Herman, and B. Meyerson, "Foundation of rf cmos and SiGe bicmos technologies," *IBM Journal of Research and Development*, vol. 47, no. 2.3, pp. 101–138, 2003.

[16] S. P. Voinigescu, T. Chalvatzis, K. H. K. Yau, A. Hazneci, A. Garg, S. Shahramian, T. Yao, M. Gordon, T. Dickson, E. Laskin, S. T. Nicolson, A. Carusone, L. Tchoketch-Kebir, O. Yuryevich, G. Ng, B. Lai, and P. Liu, "SiGe bicmos for analog, high-speed digital and millimetre-wave applications beyond 50 GHz," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2006*, Oct 2006, pp. 1–8.

[17] A. H. Pawlikiewicz and D. Hess, "Choosing RF cmos or SiGe bicmos in mixed-signal design," *RF DESIGN*, vol. 29, no. 3, pp. 36–45, 2006.

[18] F. Schaffler, *Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe*. John Wiley & Sons, Inc., New York, 2001.

[19] J. Cressler and G. Niu, *Silicon-germanium Heterojunction Bipolar Transistors*, ser. Artech House microwave library. Artech House, 2003.

[20] J. Cressler, "SiGe HBT technology: a new contender for si-based RF and microwave circuit applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 46, no. 5, pp. 572–589, may 1998.

[21] E. Crabbe, B. Meyerson, D. Harame, J. Stork, A. Megdanis, J. Cotte, J. Chu, M. Gilbert, C. Stanis, J. Comfort, G. Patton, and S. Subbanna, "113-GHz ft graded-base SiGe HBTs," in *Device Research Conference, 1993. 51st Annual*, June 1993, pp. 22–23.

[22] M. S. Latham, J. D. Cressler, A. J. Joseph, and R. C. Jaeger, "The impact of ge grading on the bias and temperature characteristics of SiGe HBT precision voltage references," *J. Phys. IV France*, vol. 06, pp. C3–113–C3–118, 1996.

[23] S. Salmon, J. Cressler, R. Jaeger, and D. Harame, "The impact of ge profile shape on the operation of SiGe HBT precision voltage references," in *Bipolar/BiCMOS Circuits and Technology Meeting, 1997. Proceedings of the*, sep 1997, pp. 100–103.

[24] S. L. Salmon, J. Cressler, R. Jaeger, and D. Harame, "The influence of ge grading on the bias and temperature characteristics of SiGe HBTs for precision analog circuits," *Electron Devices, IEEE Transactions on*, vol. 47, no. 2, pp. 292–298, Feb 2000.

[25] "DotFive: Towards 0.5 terahertz: WP4, compact modeling and device characterization," 2011.

[26] S. Selberherr, H. Stippel, and E. Strasser, *Simulation of Semiconductor Devices and Processes: vol. 5*, ser. Computational Microelectronics Series. Pineridge, 1993.

[27] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell System Technical Journal*, vol. 28, no. 3, pp. 435–489, July 1949.

[28] J. Ebers and J. Moll, "Large-signal behavior of junction transistors," *Proceedings of the IRE*, vol. 42, no. 12, pp. 1761–1772, dec. 1954.

[29] H. K. Gummel, "A charge control relation for bipolar transistors," in *Bell Sys. Techn. Journal (BSTJ)*, vol. 49. Bell Systems, Jan 1970, pp. 115–120.

[30] H. Poon and H. Gummel, "Modeling of emitter capacitance," *Proceedings of the IEEE*, vol. 57, no. 12, pp. 2181–2182, dec. 1969.

[31] H.-M. Rein and M. Schroter, "A compact physical large-signal model for high-speed bipolar transistors at high current densities-part ii: Two-dimensional model and experimental results," *Electron Devices, IEEE Transactions on*, vol. 34, no. 8, pp. 1752–1761, aug 1987.

[32] W. Webster, "On the variation of junction-transistor current-amplification factor with emitter current," *Proceedings of the IRE*, vol. 42, no. 6, pp. 914–920, june 1954.

[33] C. T. Kirk, "A theory of transistor cutoff frequency (ft) falloff at high current densities," *IEEE Transactions on Electron Devices*, vol. 9, pp. 164–174, Mar 1962.

[34] G. Kull, L. Nagel, S.-W. Lee, P. Lloyd, E. Prendergast, and H. Dirks, "A unified circuit model for bipolar transistors including quasi-saturation effects," *Electron Devices, IEEE Transactions on*, vol. 32, no. 6, pp. 1103–1113, jun 1985.

[35] H. de Graaff and W. Kloosterman, "New formulation of the current and charge relations in bipolar transistor modeling for cacd purposes," *Electron Devices, IEEE Transactions on*, vol. 32, no. 11, pp. 2415–2419, 1985.

[36] M. Schroter, M. Friedrich, and H.-M. Rein, "A generalized integral charge-control relation and its application to compact models for silicon-based HBT's," *Electron Devices, IEEE Transactions on*, vol. 40, no. 11, pp. 2036–2046, nov 1993.

[37] M. Schroter, "A survey of present compact models for high-speed bipolar transistors," *Frequenz*, vol. 47, no. 7-8, pp. 178–190, July 1993.

[38] H. Stubing and H.-M. Rein, "A compact physical large-signal model for high-speed bipolar transistors at high current densities-part i: One-dimensional model," *Electron Devices, IEEE Transactions on*, vol. 34, no. 8, pp. 1741–1751, aug 1987.

[39] P. Chevalier, B. Barbalat, M. Laurens, B. Vandelle, L. Rubaldo, B. Geynet, S. Voinigescu, T. Dickson, N. Zerounian, S. Chouteau, D. Dutartre, A. Monroy, F. Aniel, G. Dambrine, and A. Chantre, "High-speed SiGe bicmos technologies: 120-nm status and end-of-roadmap challenges," in *Silicon Monolithic Integrated Circuits in RF Systems, 2007 Topical Meeting on*, Jan 2007, pp. 18–23.

[40] M. Schroter, A. Pawlak, J. Krause, and P. Sakalas, "Latest developments of hicum/l2 for mm-wave applications," in *OBip: BCTM Open Bipolar Workshop, Bordeaux*, 2013.

[41] B. Ardouin, "Tools and environment for sub-thz circuit design," in *OBip: BCTM Open Bipolar Workshop, Bordeaux*, 2013.

[42] J. Early, "Effects of space-charge layer widening in junction transistors," *Proceedings of the IRE*, vol. 40, no. 11, pp. 1401–1406, nov. 1952.

[43] M. Schroter, *DEVICE - A Mixed-Mode Simulator for Three-Dimensional Heterostructure Semiconductor Devices and Circuits, User's Guide to Version 2.03*, 2008.

[44] M. Schroter and A. Chakravorty, *Compact Hierarchical Bipolar Transistor Modeling With Hicum*, ser. International Series on Advances in Solid State Electronics and Technology. World Scientific, 2010.

[45] M. Schroter and H. Tran, "Charge-strorage related parameter calculations for si and sige bipolar transistors from device simulation," *Proc. WCM*, pp. 735–740, 2006.

[46] B. Ardouin, C. Raya, M. Schroter, A. Pawlak, D. Celi, F. Pourchon, K. Aufinger, T. Meister, and T. Zimmer, "Modeling and parameter extraction of SiGe: C HBT's with hicum for the emerging terahertz era," in *Microwave Integrated Circuits Conference (EuMIC), 2010 European*, sept. 2010, pp. 25–28.

[47] M. Schroter and A. Chakravorty, *HICUM - A Geometry Scalable Physics-Based Compact Bipolar Transistor model - Documentation of model level 2 version 2.2*, 2005.

[48] M. Schroter, A. Pawlak, and A. Mukherjee, *HICUM/L2 - A geometry scalable physics-based compact bipolar transistor model*, 2013, documentation of model version 2.32.

[49] A. Pawlak, M. Schroter, and J. Krause, "A HICUM extension for medium current densities," *9th HICUM Workshop, Wuerzburg*, October 2009.

[50] A. Pawlak, M. Schroter, A. Mukherjee, and T. Kessler, "HICUM/l2 v2.30 overview," *10th HICUM Workshop*, 2010.

[51] J. Paasschens, W. Kloosterman, and R. v.d. Toorn, *Model derivation of Mextram 504, The physics behind the model*, Philips Nat.Lab., 2005.

[52] R. van der Toorn, J. Paasschens, and W. Kloosterman, *The Mextram bipolar transistor model, level 504*, Philips Nat.Lab., 2008.

[53] J. Paasschens, W. Kloosterman, and R. Havens, "Modelling two SiGe HBT specific features for circuit simulation," in *Bipolar/BiCMOS Circuits and Technology Meeting, Proceedings of the 2001*, 2001, pp. 38–41.

[54] C. Kittel and H. Kroemer, *Thermal Physics*. W. H. Freeman, 1980.

[55] H. Poon, H. Gummel, and D. Scharfetter, "High injection in epitaxial transistors," *Electron Devices, IEEE Transactions on*, vol. 16, no. 5, pp. 455–457, may 1969.

[56] E. Johnson, "Physical limitations on frequency and power parameters of transistors," in *IRE International Convention Record*, vol. 13, mar 1965, pp. 27–34.

[57] S. Tiwari, "A new effect at high currents in heterostructure bipolar transistors," *Electron Device Letters, IEEE*, vol. 9, no. 3, pp. 142–144, march 1988.

[58] M. S. Peter, J. Slotboom, and D. Terpstra, "Impact ionization and neutral base recombination in sige hbts," in *Bipolar/BiCMOS Circuits and Technology Meeting, 1999. Proceedings of the 1999*, 1999, pp. 58–61.

[59] K. Roenker, S. Alterovitz, and C. Mueller, "Device physics analysis of parasitic conduction band barrier formation in SiGe HBTs," in *Silicon Monolithic Integrated Circuits in RF Systems, 2000. Digest of Papers. 2000 Topical Meeting on*, 2000, pp. 182–186.

[60] S. Fregonese, T. Zimmer, C. Maneux, and P. Sulima, "Barrier effects in SiGe HBT: modeling of high-injection base current increase," in *Bipolar/BiCMOS Circuits and Technology, 2004. Proceedings of the 2004 Meeting*, sept. 2004, pp. 104–107.

[61] M. Schroter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener, and A. Koldehoff, "Physics- and process-based bipolar transistor modeling for integrated circuit design," *Solid-State Circuits, IEEE Journal of*, vol. 34, no. 8, pp. 1136–1149, Aug 1999.

[62] G. Avenier, M. Diop, P. Chevalier, G. Troillard, N. Loubet, J. Bouvier, L. Depoyan, N. Derrier, M. Buczko, C. Leyris, S. Boret, S. Montusclat, A. Margain, S. Pruvost, S. Nicolson, K. Yau, N. Revil, D. Gloria, D. Dutartre, S. Voinigescu, and A. Chantre, "0.13 um SiGe bicmos technology fully dedicated to mm-wave applications," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 9, pp. 2312–2321, sept. 2009.

[63] B. Tyrrell, P. Rhyins, P. Martin, M. Fritze, D. Astolfi, R. Mallen, and B. Wheeler, "Investigation of the physical and practical limits of dense-only phase shift lithography for circuit feature definition*," *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 1, no. 3, pp. 243–252, 2002.

[64] R. F. Pease and S. Y. Chou, "Lithography and other patterning techniques for future electronics," *Proceedings of the IEEE*, vol. 96, no. 2, pp. 248–270, 2008.

[65] E. Canderle, P. Chevalier, A. Montagne, L. Moynet, G. Avenier, P. Boulenc, M. Buczko, Y. Carminati, J. Rosa, C. Gaquiere, and A. Chantre, "Extrinsic base resistance optimization in dpsa-seg sige:c hbts," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2012 IEEE*, Sept 2012, pp. 1–4.

[66] M. Schroter, S. Lehmann, and D. Celi, "Non-standard geometry scaling effects in high-frequency sige bipolar transistors," *Proc. WCM, International NanoTech Meeting*, pp. 603–608, 2007.

[67] A. Pawlak, M. Schroter, and A. Fox, "Geometry scalable model parameter extraction for mm-wave sige-heterojunction transistors," *Proc. IEEE BCTM*, 2013.

[68] H.-M. Rein, "A simple method for separation of the internal and external (peripheral) currents of bipolar transistors," *Solid-State Electronics*, vol. 27, no. 7, pp. 625 – 631, 1984.

[69] S. Sze and K. Ng, *Physics of semiconductor devices*, ser. Wiley-Interscience publication. Wiley-Interscience, 2007.

[70] S. Marschmeyer and B. Heinemann, "Dry chemical opening of emitter windows of high speed pnp SiGe:C HBTs in a complementary bicmos technology," *Proc. 19th International Symposium on Plasma Chemistry*, vol. 19, no. 19, pp. 514–517, 2009, iSPC19, Bochum, Germany.

[71] M. Schroter and D. Walkey, "Physical modeling of lateral scaling in bipolar transistors," *Solid-State Circuits, IEEE Journal of*, vol. 31, no. 10, pp. 1484–1492, oct 1996.

[72] K. M. Walter, B. Ebersman, D. A. Sunderland, G. D. Berg, G. Freeman, R. A. Groves, D. Jadus, and D. Harame, "A scalable, statistical spice gummel-poon model for sige hbts," in *Bipolar/BiCMOS Circuits and Technology Meeting, 1997. Proceedings of the*, Sep 1997, pp. 32–35.

[73] M. Schroter, "Modeling of the low-frequency base resistance of single base contact bipolar transistors," *Electron Devices, IEEE Transactions on*, vol. 39, no. 8, pp. 1966–1968, aug 1992.

[74] S. Verhaeren, "Mise en oeuvre d'un modele de transistor bipolaire RF parametre," STMicroelectronics, Tech. Rep., 1999.

[75] M. Schroter, J. Krause, S. Lehmann, and D. Celi, "Compact layout and bias-dependent base-resistance modeling for advanced SiGe HBTs," *Electron Devices, IEEE Transactions on*, vol. 55, no. 7, pp. 1693–1701, july 2008.

[76] H. Murrmann and D. Widmann, "Current crowding on metal contacts to planar devices," *Electron Devices, IEEE Transactions on*, vol. 16, no. 12, pp. 1022–1024, Dec 1969.

[77] N. Kauffmann, C. Raya, F. Pourchon, S. Ortolland, and D. Celi, "Determination of the collector resistance rcx of bipolar transistor," *5th European HICUM Workshop*, 2005.

[78] C. Raya, N. Kauffmann, F. Pourchon, D. Celi, and T. Zimmer, "Scalable approach for external collector resistance calculation," in *Microelectronic Test Structures, 2007. ICMTS '07. IEEE International Conference on*, March 2007, pp. 101–106.

[79] M. J. W. Rodwell, M. Urteaga, T. Mathew, D. Scott, D. Mensa, Q. Lee, J. Guthrie, Y. Betser, S. Martin, R. Smith, S. Jaganathan, S. Krishnan, S. Long, R. Pullela, B. Agarwal, U. Bhattacharya, L. Samoska, and M. Dahlstrom, "Submicron scaling of hbts," *Electron Devices, IEEE Transactions on*, vol. 48, no. 11, pp. 2606–2624, Nov 2001.

[80] H. BECKRICH-ROS, "Contribution a la caracterisation et a la modelisation de transistors bipolaires de puissance integres dans une filiere bicmos submicronique," Ph.D. dissertation, UNIVERSIT BORDEAUX I, COLE DOCTORALE DE SCIENCES PHYSIQUES ET DE LINGNIEUR, 2006.

[81] R. Hull, *Properties of Crystalline Silicon*, ser. EMIS datareviews series. INSPEC, the Institution of Electrical Engineers, 1999.

[82] P. Mars, "Temperature dependence of avalanche breakdown voltage in pn junctions," in *Int. Journal Electronics*, vol. 32, no. 1, 1971, pp. 23–37.

[83] C. Raya, "Modlisation et optimisation de transistors bipolaires htrojonction si/sigec ultra rapides pour applications millimtriques," Ph.D. dissertation, University of Bordeaux 1, 2008.

[84] H.-M. Rein and M. Schroter, "Experimental determination of the internal base sheet resistance of bipolar transistors under forward-bias conditions," *Solid-State Electronics*, vol. 34, no. 3, pp. 301 – 308, 1991.

[85] F. Stein, N. Derrier, C. Maneux, and D. Celi, "Advanced extraction procedure for parasitic collector series resistance contributions in high-speed bicmos technologies," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2013 IEEE*, Sept 2013, pp. 33–36.

[86] R. Johnson, J. Evans, P. Jacobsen, J. Thompson, and M. Christopher, "The changing automotive environment: high-temperature electronics," *Electronics Packaging Manufacturing, IEEE Transactions on*, vol. 27, no. 3, pp. 164–176, July 2004.

[87] T.-Y. Lee, M. Schroter, and M. Racanelli, "A scaleable model generation methodology of bipolar transistors for rfic design," in *Bipolar/BiCMOS Circuits and Technology Meeting, Proceedings of the 2001*, 2001, pp. 171–174.

[88] J. Berkner, *Kompaktmodelle fur Bipolartransistoren: Praxis der Modellierung, Messung und Parameterbestimmung-SGP, VBIC, HICUM und MEXTRAM*, ser. Reihe Technik. Expert-Verlag GmbH, 2002.

[89] D. Berger, D. Celi, M. Schroter, M. Malorny, T. Zimmer, and B. Ardouin, "Hicum parameter extraction methodology for a single transistor geometry," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2002. Proceedings of the 2002*, 2002, pp. 116–119.

[90] D. Celi, "Step by step extraction of hicum/l2 high-current parameters," in *8th HICUM Workshop*, 2008.

[91] A. Pawlak, M. Schroter, J. Krause, D. Celi, and N. Derrier, "HICUM/2 v2.3 parameter extraction for advanced SiGe-heterojunction bipolar transistors," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2011 IEEE*, oct. 2011, pp. 195–198.

[92] R. Tinti, F. Paolini, T. Eguchi, and F. Sischka, "Ic-cap waferpro: A new software environment for automated dc/cv and rf measurements in ic-cap," Agilent Technologies, Agilen Whitepaper 5990-6494EN, 2010.

[93] R. Tinti, "Automating on-wafer measurements with the new agilent ic-cap waferpro," Agilent Technologies, Jan 2011, waferPro Webcast.

[94] D. Celi, "An attempt to determine the emitter size of bipolar transistors from electrical measurements," in *9th HICUM Workshop*, 2009.

[95] S. S. Cohen, "Contact resistance and methods for its determination," *Thin Solid Films*, vol. 104, no. 3-4, pp. 361–379, 1983.

[96] W. Loh, S. Swirhun, T. A. Schreyer, R. Swanson, and K. Saraswat, "Modeling and measurement of contact resistances," *Electron Devices, IEEE Transactions on*, vol. 34, no. 3, pp. 512–524, Mar 1987.

[97] G. Gildenblat, *Compact Modeling: Principles, Techniques and Applications*. Springer, 2010.

[98] T. Nakadai and K. Hashimoto, "Measuring the base resistance of bipolar transistors," in *Bipolar Circuits and Technology Meeting, 1991., Proceedings of the 1991*, 1991, pp. 200–203.

[99] M. Schroter, "Methods for extracting parameters of geometry scalable compact bipolar transistor models," University of Technology Dresden, CEDIC internal document, 2000.

[100] W. Sansen and R. Meyer, "Characterization and measurement of the base and emitter resistances of bipolar transistors," *Solid-State Circuits, IEEE Journal of*, vol. 7, no. 6, pp. 492–498, 1972.

[101] Y. Gobert, P. Tasker, and K.-H. Bachem, "A physical, yet simple, small-signal equivalent circuit for the heterojunction bipolar transistor," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 45, no. 1, pp. 149–153, 1997.

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

[103] C. Mcandrew, "BJT base and emitter resistance extraction from dc data," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2006*, Oct 2006, pp. 1–4.

[104] F. Stein, Z. Huszka, N. Derrier, C. Maneux, and D. Celi, "Extraction procedure for emitter series resistance contributions in sigec bicmos technologies," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, 25-27th Mar 2014, pp. 20–25.

[105] Z. Huszka, E. Seebacher, and W. Pflanzl, "An extended two-port method for the determination of the base and emitter resistance," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2005. Proceedings of the*, 2005, pp. 188–191.

[106] R. Martin and D. Thomson, "Robust-resistant spectrum estimation," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1097–1115, 1982.

[107] R. D. Martin, "Robust-resistant spectral analysis," in *Time Series in the Frequency Domain*, ser. Handbook of Statistics, D. Brillinger and P. Krishnaiah, Eds. Elsevier, 1983, vol. 3, pp. 185–219.

[108] B. Ardouin, T. Zimmer, D. Berger, D. Celi, H. Mnif, T. Burdeau, and P. Fouillat, "Transit time parameter extraction for the hicum bipolar compact model," in *Bipolar/BiCMOS Circuits and Technology Meeting, Proceedings of the 2001*, 2001, pp. 106–109.

[109] G. Wedel, *POICAPS - A Multidimentional Numerical Capacitance Simulator*, CEDIC, Dresden University of Technology, 2012.

[110] G. Fischer, "Substrate coupling for 500ghz hbts," in *AK Bipolar, Campeon Munich*, 2012.

[111] M. Schroter, Z. Yan, T. Lee, and W. Shi, "A compact tunneling current and collector breakdown model," in *Bipolar/BiCMOS Circuits and Technology Meeting, 1998. Proceedings of the 1998*, Sep 1998, pp. 203–206.

[112] F. Stein, Z. Huszka, N. Derrier, C. Maneux, and D. Celi, "Extraction of the emitter related space charge weighting factor parameters of hicum l2.30 using the lambert w function," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2012 IEEE*, Sept 2012, pp. 1–4.

[113] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the lambert w function," in *Advances in Computational Mathematics*, vol. 5, 1996, pp. 329–359.

[114] B. Ardouin and C. Raya, "Improved HBT parameter extraction methodology [dotfive-wp4-d4.1.5]," Tech. Rep., 04 2011.

[115] M. Reisch, *High-Frequency Bipolar Transistors: Physics, Modelling, Applications*, ser. Advanced microelectronics. Springer Berlin Heidelberg, 2003.

[116] M. Schroter, A. Mukherjee, and A. Pawlak, *HICUM Level2 version 2.30: Release Notes*, Chair for Electron Devices & Integrated Circuits (CEDIC) University of Technology Dresden, Germany, February 2011.

[117] F. Chapeau-Blondeau and A. Monir, "Numerical evaluation of the lambert w function and application to generation of generalized gaussian noise with exponent 1/2," *Signal Processing, IEEE Transactions on*, vol. 50, no. 9, pp. 2160–2165, sep 2002.

[118] R. Fox, S.-G. Lee, and D. Zweidinger, "The effects of BJT self-heating on circuit behavior," *Solid-State Circuits, IEEE Journal of*, vol. 28, no. 6, pp. 678–685, Jun 1993.

[119] M. Schroter and H.-M. Rein, "Investigation of very fast and high-current transients in digital bipolar ic's using both a new compact model and a device simulator," *Solid-State Circuits, IEEE Journal of*, vol. 30, no. 5, pp. 551–562, May 1995.

[120] N. Rinaldi, "Small-signal operation of semiconductor devices including self-heating, with application to thermal characterization and instability analysis," *Electron Devices, IEEE Transactions on*, vol. 48, no. 2, pp. 323–331, Feb 2001.

[121] J.-S. Rieh, D. Greenberg, B. Jagannathan, G. Freeman, and S. Subbanna, "Measurement and modeling of thermal resistance of high speed SiGe heterojunction bipolar transistors," in *Silicon Monolithic Integrated Circuits in RF Systems, 2001. Digest of Papers. 2001 Topical Meeting on*, Sept 2001, pp. 110–113.

[122] J. C. J. Paasschens, S. Harmsma, and R. van der Toorn, "Dependence of thermal resistance on ambient and actual temperature," in *Bipolar/BiCMOS Circuits and Technology, 2004. Proceedings of the 2004 Meeting*, Sept 2004, pp. 96–99.

[123] J. Berkner, "Extraction of thermal resistance and its temperature dependence using dc methods," in *7th European HICUM Workshop, Dresden*, vol. 7, 2007.

[124] M. Pfost, V. Kubrak, and P. Brenner, "A practical method to extract the thermal resistance for heterojunction bipolar transistors," in *European Solid-State Device Research, 2003. ESSDERC '03. 33rd Conference on*, Sept 2003, pp. 335–338.

[125] M. Reisch, "Self-heating in {BJT} circuit parameter extraction," *Solid-State Electronics*, vol. 35, no. 5, pp. 677 – 679, 1992.

[126] M. Weis, S. Fregonese, M. Santorelli, A. Sahoo, C. Maneux, and T. Zimmer, "Pulsed i(v), pulsed rf measurement system for microwave device characterization with 80ns/45ghz," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, Sept 2012, pp. 189–192.

[127] A. K. Sahoo, S. Fregonese, M. Weiß, B. Grandchamp, N. Malbert, and T. Zimmer, "Characterization of self-heating in si–ge HBTs with pulse, dc and ac measurements," *Solid-State Electronics*, vol. 76, pp. 13–18, 2012.

[128] J. Lindmayer, "Power gain of transistors at high frequencies," *Solid State Electronics*, vol. 5, pp. 171–175, Jun. 1962.

[129] C.-H. Choi, P. R. Chidambaram, R. Khamankar, C. Machala, Z. Yu, and R. Dutton, "Dopant profile and gate geometric effects on polysilicon gate depletion in scaled mos," *Electron Devices, IEEE Transactions on*, vol. 49, no. 7, pp. 1227–1231, Jul 2002.

[130] M. Sarkar, A. C. Hoe, H. Jiayi, and T. Chen, "Impact of nonuniform graded dopant profile in polysilicon gate on gate leakage current," *Electron Devices, IEEE Transactions on*, vol. 52, no. 6, pp. 1200–1204, June 2005.

[131] A. Khakifirooz and D. Antoniadis, "Mosfet performance scaling part i: Historical trends," *Electron Devices, IEEE Transactions on*, vol. 55, no. 6, pp. 1391–1400, June 2008.

[132] M. Schroter, S. Lehmann, and D. Celi, "Non-standard geometry scaling effects in sige hbts," *7th HICUM Workshop*, 2007.

[133] M. Schroter, "Hbt geometry scaling: issues and solutions," *13th HICUM Workshop*, 2013.

[134] W. Chen, *VLSI Technology*, ser. Principles and applications in engineering. Taylor & Francis, 2003.

[135] J. D. Cressler, *Silicon Heterostructure Handbook: Materials, Fabrication, Devices, Circuits and Applications of SiGe and Si Strained-Layer Epitaxy*. Hoboken, NJ: CRC Press, 2005.

[136] T. Lacave, P. Chevalier, Y. Campidelli, M. Buczko, L. Depoyan, L. Berthier, G. Avenier, C. Gaquieandre, and A. Chantre, "Vertical profile optimization for +400 GHz fmax si/SiGe:C HBTs," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2010 IEEE*, oct. 2010, pp. 49–52.

[137] J. John, F. Chai, D. Morgan, T. Keller, J. Kirchgessner, R. Reuter, H. Rueda, J. Teplik, J. White, S. Wipf, and D. Zupac, "Optimization of a SiGe:C HBT in a bicmos technology for low power wireless applications," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2002. Proceedings of the 2002*, 2002, pp. 193–196.

[138] P. Chevalier, F. Pourchon, T. Lacave, G. Avenier, Y. Campidelli, L. Depoyan, G. Troillard, M. Buczko, D. Gloria, D. Celi, C. Gaquiere, and A. Chantre, "A conventional double-polysilicon fsa-seg si/SiGe:C HBT reaching 400 GHz fmax," in *Bipolar/BiCMOS Circuits and Technology Meeting, 2009. BCTM 2009. IEEE*, Oct 2009, pp. 1–4.

[139] B. Kramer, *Advances in Solid State Physics*, ser. Advances in Solid State Physics. Springer, 2002.

[140] J. Cressler, *Fabrication of SiGe HBT BiCMOS Technology*. Taylor & Francis, 2007.

[141] S. Franssila, *Introduction to Microfabrication*. Wiley, 2010.

[142] D. Harame, J. Boquet, G. Masini, T. Krishnamohan, B. Tillack, S. Bedell, S. Miyazaki, A. Reznicek, and S. Koester, *SiGe, Ge, and Related Compounds 4: Materials, Processing, and Devices*, ser. ECS transactions. Electrochemical Society, 2010, no. Nr. 6.

[143] J.-S. Rieh, B. Jagannathan, D. Greenberg, M. Meghelli, A. Rylyakov, F. Guarin, Z. Yang, D. Ahlgren, G. Freeman, P. Cottrell, and D. Harame, "SiGe heterojunction bipolar transistors and circuits toward terahertz communication applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, no. 10, pp. 2390–2408, Oct 2004.

[144] D. Celi, "Method for accurate determinaton of the intrinsic cut-off frequency of ic bipolar transistors," in *Microelectronic Test Structures, 1988. ICMTS. Proceedings of the 1988 IEEE International Conference on*, feb. 1988, pp. 200–203.

[145] S. Mason, "Power gain in feedback amplifier," *Circuit Theory, Transactions of the IRE Professional Group on*, vol. CT-1, no. 2, pp. 20–25, June 1954.

[146] M. Gupta, "Power gain in feedback amplifiers, a classic revisited," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 40, no. 5, pp. 864–879, 1992.

[147] M. Vaidyanathan and D. Pulfrey, "Extrapolated fmax of heterojunction bipolar transistors," *Electron Devices, IEEE Transactions on*, vol. 46, no. 2, pp. 301–309, Feb 1999.

[148] R. L. Pritchard, "Electric network representation of transistors-a survey," *Circuit Theory, IRE Transactions on*, vol. 3, no. 1, pp. 5–21, Mar 1956.

[149] J. M. Early, "Structure-determined gain-band product of junction triode transistors," *Proceedings of the IRE*, vol. 46, no. 12, pp. 1924–1927, Dec 1958.

[150] E. Kasper, *Properties of strained and relaxed silicon germanium*, ser. EMIS datareviews series. Kasper, E. and Institution of Electrical Engineers and INSPEC (Information service): INSPEC, 1995.

[151] J. Weber and M. I. Alonso, "Near-band-gap photoluminescence of si-ge alloys," *Phys. Rev. B*, vol. 40, pp. 5683–5693, Sep 1989.

[152] Agilent Technologies, "De-embedding and embedding s-parameter networks using a vector network analyzer, application note 1364-1," Agilent Technologies, Tech. Rep., 2004.

[153] D. Williams, R. B. Marks, and A. Davidson, "Comparison of on-wafer calibrations," in *ARFTG Conference Digest-Winter, 38th*, vol. 20, Dec 1991, pp. 68–81.

[154] A. J. Lord, "Comparing the accuracy and repeatability of on-wafer calibration techniques to 110GHz," in *Microwave Conference, 1999. 29th European*, vol. 3, Oct 1999, pp. 28–31.

[155] H. Heuermann, A. Rumiantsev, and S. Schott, "Advanced on-wafer multiport calibration methods for mono-and mixed-mode device characterization," in *ARFTG Conference Digest Spring, 2004. 63rd*, June 2004, pp. 91–96.

[156] A. Rumiantsev, P. Sakalas, F. Pourchon, P. Chevalier, N. Derrier, and M. Schroter, "Application of on-wafer calibration techniques for advanced high-speed bicmos technology," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2010 IEEE*, Oct 2010, pp. 98–101.

[157] Agilent Technologies, "Signal integrity analysis series part 3: The abcs of de-embedding, application note 5989-5765en," Agilent Technologies, Tech. Rep., 2007.

[158] Hewlett-Packard, "Understanding the fundamental principles of vector network analysis, application note 1287-1," Hewlett-Packard, Tech. Rep., 1997.

[159] A. Rumiantsev, P. Sakalas, N. Derrier, D. Celi, and M. Schroter, "Influence of probe tip calibration on measurement accuracy of small-signal parameters of advanced bicmos HBTs," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2011 IEEE*, Oct 2011, pp. 203–206.

[160] P. Van Wijnen, "On the characterization and optimization of high-speed silicon bipolar transistors," Ph.D. dissertation, Electrical Engineering, Mathematics and Computer Science, TU Delft, 1992.

[161] Cascade Microtech, "On-wafer vector network analyzer calibration and measurements, application note, pyroprobe 0597," Cascade Microtech, Tech. Rep., 1997.

[162] M. C. A. M. Koolen, J. A. M. Geelen, and M. P. J. G. Versleijen, "An improved deembedding technique for on-wafer high-frequency characterization," in *Bipolar Circuits and Technology Meeting, 1991., Proceedings of the 1991*, Sep 1991, pp. 188–191.

[163] L. Giacoletto, "Diode and transistor equivalent circuits for transient operation," *Solid-State Circuits, IEEE Journal of*, vol. 4, no. 2, pp. 80–83, Apr 1969.

[164] K. C. Gupta, R. Garg, and R. Chadha, "Computer aided design of microwave circuits," *NASA STI/Recon Technical Report A*, vol. 82, p. 39449, 1981.

[165] J. Choma and W. Chen, *Feedback Networks: Theory and Circuit Applications*, ser. Advanced series in circuits and systems. World Scientific, 2007.

[166] K. Kurokawa, "Power waves and the scattering matrix," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 13, no. 2, pp. 194–202, Mar 1965.

[167] Agilent Technologies, "An 154, s-parameter design, application note," Agilent Technologies, Tech. Rep., 2000.

[168] D. Frickey, "Conversions between s, z, y, h, abcd, and t parameters which are valid for complex source and load impedances," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 42, no. 2, pp. 205–211, Feb 1994.

[169] D. Roulston, *Bipolar semiconductor devices*, ser. McGraw-Hill series in electrical and computer engineering. McGraw-Hill Higher Education, 1990.

[170] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, "A review of hydrodynamic and energy-transport models for semiconductor device simulation," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 251–274, feb 2003.

[171] W. Van Roosbroeck, "Theory of the flow of electrons and holes in germanium and other semiconductors," *Bell System Technical Journal*, vol. 29, no. 4, pp. 560–607, October 1950.

[172] *HSPICE Reference Manual - Elements and Device Models*, version c-2009.03 ed., Synopsys, 2009.

[173] R. Miles, *Terahertz Frequency Detection and Identification of Materials and Objects*, ser. NATO Security through Science Series. Springer, 2007.

[174] A. Y. Pawar, D. D. Sonawane, K. B. Erande, and D. V. Derle, "Terahertz technology and its applications," *Drug Invention Today*, vol. 5, no. 2, pp. 157 – 163, 2013.

[175] I. Hosako, N. Sekine, M. Patrashin, S. Saito, K. Fukunaga, Y. Kasai, P. Baron, T. Seta, J. Mendrok, S. Ochiai, and H. Yasuda, "At the dawn of a new era in terahertz technology," *Proceedings of the IEEE*, vol. 95, no. 8, pp. 1611–1623, Aug 2007.

List of Figures

List of Tables