TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI INGEGNERIA BIOMEDICA, ELETTRONICA
E DELLE TELECOMUNICAZIONI

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

TRANSPORT MODELS AND ADVANCED NUMERICAL SIMULATION OF SILICON-GERMANIUM HETEROJUNCTION BIPOLAR TRANSISTORS

GRAZIA SASSO

Il coordinatore del Corso di Dottorato

Ch.mo Prof. Niccolò RINALDI

Il Tutore
Ch.mo Prof. Niccolò RINALDI

Anno Accademico 2009-2010

"I learned that there is no substitute for combining a good intuitive feel for devices with the predictions of your program. That is, if one has a good intuitive "feel" for devices, then the computer can be used to enhance that intuitive capability."

Don Scharfetter, speech at **SISPAD** 2000.

"Le occasioni della vita sono infinite e le loro armonie si schiudono ogni tanto a dar sollievo a questo nostro pauroso vagare per sentieri che non conosciamo."

Pier Vittorio Tondelli, ***Pao Pao*** (1982)

"No. Ma sta' attento: dato che non siamo calzini ma persone, non siamo qui con il fine principale di essere puliti. I desideri sono la cosa più importante che abbiamo e non si può prenderli in giro più di tanto. Così, alle volte, vale la pena di non dormire per star dietro ad un proprio desiderio. Si fa la schifezza e poi si paga. E' solo questo davvero importante: che quando arriva il momento di pagare uno non pensi a scappare e stia lì, dignitosamente, a pagare. Solo questo è importante."

Alessandro Baricco, ***Castelli di Rabbia*** (1991)

Acknowledgments

I would like to express my gratitude to a number of people who supported me during my work on this dissertation which would not have been possible without their help and challenge.

First, I would like to thank my supervisor and mentor Prof. Niccolò Rinaldi for giving me the opportunity to join his research group. He set high output standards offering learning, counsel and encouragement that have been indispensable to the completion of this work: the harder it is to keep up with high requirements, the sweeter the feeling to match them. His devotion to scientific research sets an excellent paradigm for me to follow.

I am very grateful to my current and former colleagues-friends at the Electronics group of *DIBET* for their support, cooperation and friendship. I was blessed with easygoing and genial colleagues, allowing all the days at work to be spent nicely.

I was lucky enough to be involved in a great and international research project called *DOTFIVE* and founded by European Commission through the Seventh Framework Program for Research and Technology Development. This research project allowed me to cooperate with academic and industrial top level partners and gave me the chance to work with Prof. Christoph Jungemann and Prof. Michael Schroeter during the time I spent at their departments, at *Bundeswehr University* and *TU Dresden* respectively. I would like to thank them for giving me the opportunity to join their group as a visiting researcher and letting me (partially) share their deep insight into semiconductor physics and modeling. Furthermore, I am grateful to Prof. Christoph Jungemann and his research group for kindly providing the spherical harmonics expansion simulator *SPRING* and for always finding the time to answer my questions. I also express my appreciation to Technology Modeling Team of *STMicroelectronics* in *Crolles* for supplying measurements wafer and TCAD results.

I want to thank my family. My parents made possible all of my studies by their continuous support. They have taught me how to work hard and bring out the best of myself.

The last - but not least - thank is for all of my friends. They always comprehended my changeable mood and cheered me up in difficulties I encountered.

Contents

Acknowledgments .... **I**

Contents .... **III**

Introduction **1**

Chapter 1 .... **5**

Silicon-germanium heterojunction bipolar transistors **5**

1.1 History, state-of-art and applications........ 6

1.2 Figures of Merit. 9

1.2.1 Cut-off frequency f_T........ 9

1.2.2 Maximum oscillation frequency f_{max}. 11

1.2.3 Impact ionization. 11

1.3 Carrier transport models and TCAD for SiGe HBTs development 14

1.3.1 Boltzmann transport equation 16

1.3.2 The drift-diffusion model 17

1.3.3 Energy transport and hydrodynamic models 18

1.3.4 The Monte Carlo model 20

1.3.5 The spherical harmonics expansion model 20

1.4 References 21

Chapter 2 .... **24**

Hydrodynamic model verification **24**

2.1 Hydrodynamic model in commercial TCAD. 25

2.2 Analysis of model parameters for a 100 GHz device........ 27

2.2.1 Parameter r 29
2.2.2 Parameter f^{td} 31
2.2.3 Parameter f^{hf} 32
2.2.4 Discussion 34
2.3 Optimization of standard models for a 100 GHz device 36
2.3.1 Four moments model 36
2.3.2 Stratton model 37
2.3.3 Blotekjær model 39
2.3.4 Hybrid optimization 41
2.3.5 Models' discussion and comparison 45
2.4 Analysis and optimization for a 450 GHz device 47
2.5 Analysis and optimization for a 700 GHz device 53
2.6 Unified optimization over different technological nodes 55
2.7 References 56

Chapter 3 58
Analytical models for transport parameters 58
3.1 Look-up table models 59
3.1.1 Bandgap narrowing model 61
3.2 Analytical models 63
3.2.1 Effective density of states 64
3.2.2 Low-field mobility 68
3.2.3 Energy relaxation time 80
3.2.4 Saturation velocity 81
3.2.5 High-field mobility 84
3.3 References 91

Chapter 4 96
Verification of transport models 96

4.1 Implementation of analytical models in *Sdevice* 96
4.2 Reference state-of-art models 97
4.3 Verification of models 98
4.3.1 One-dimensional 100 GHz device 98
4.3.2 One-dimensional 450 GHz device 100
4.3.3 Two-dimensional 230 GHz device............................ 102
4.4 *C++* code examples for *Sdevice* implementation 105
4.4.1 Low field mobility model... 105
4.4.2 Energy relaxation time model 113
4.5 References 116

Chapter 5.. 118
Avalanche multiplication measurements and modeling............ 118
5.1 Device simulation of the avalanche multiplication 119
5.1.1 Simulation approaches 119
5.1.2 Avalanche generation models for HD simulations 120
5.1.3 Okuto-Crowell model calibration............................. 122
5.2 Avalanche multiplication factor model 125
5.2.1 Main models review 125
5.2.2 A new multiplication factor model............................ 130
5.2.3 Measurements and model parameters' extraction...... 133
5.3 References 134

Conclusions and outlook.. 137

Introduction

I.1. Numerical simulation of silicon-germanium heterojunction bipolar transistors

Applications in the emerging high-frequency markets for millimeter wave applications more and more use SiGe components for cost reasons. The SiGe BiCMOS technology allows integration of analog and digital parts, providing high integration densities and saving costs: the combination of SiGe heterojunction bipolar transistors (HBTs) with advanced Si CMOS to form a SiGe BiCMOS technology represents a unique opportunity for Si-based RF system-on-a-chip solutions.

Current state-of-the-art research and development is taking place primarily in data communication and radar systems. Technologies with higher cut-off frequency f_T can directly lead to improved automotive radar systems with higher performance at lower power consumption, which increases road safety and energy budget. With an increased f_T completely new and highly integrated microwave sensor systems are feasible. However, until recently, this spectral region has resisted attempts to broadly harness its potential for everyday applications. This led to the expression *THz gap*, loosely describing the lack of adequate technologies to effectively bridge this transition region between microwaves and optics, both readily accessible via well developed electronic and laser-based approaches. THz technology is an emerging field which has demonstrated a wide-ranging potential. Extensive research in the last years has identified many attractive application areas and has paved the technological path towards broadly usable THz systems.

This thesis has been developed within the European project DOTFIVE, planning to establish the basis for fully integrated cost efficient electronic THz solutions. The aggressive technology development effort of the DOTFIVE project has been significantly advancing the performance of SiGe:C HBTs towards the Terahertz

range. To support such a technology effort, a reliable TCAD platform is required for developing predictive device and process simulation, exploring the physics and performance of extremely scaled devices, identifying operating limits and investigating new device concepts.

The main issue in the simulation of scaled devices is related to the limitations of the physical models used to describe charge carrier transport. A widely used approach is the so called drift-diffusion (DD) model, which treats carrier transport as diffusion and drift processes. However, as the device size approaches the nanometer range, charge transport becomes quasi-ballistic, and non-local effects such as velocity overshoot occur. In an attempt to capture these phenomena, more advanced transport models have been proposed, often termed hydrodynamic (HD) or energy-transport models. Both the DD and HD models can be viewed as approximations of the Boltzmann transport equation (BTE), which represents the most rigorous approach to model charge carrier transport in semiconductors. The BTE is an integro-differential equation in the six-dimensional phase space, which can be solved by a stochastic approach, the Monte Carlo (MC) method, or a deterministic approach, the spherical harmonics expansion (SHE) method. While MC/SHE simulation is the most accurate approach, it is also very time consuming. Long simulation times become a concern in the technology development cycle, where many simulation runs of complex device architectures are typically needed. For this reason calibrated HD simulation to support technology development are highly desired.

However, HD simulation of scaled devices is not a trivial task. This is primarily due to inherent approximations in the HD formalism. In addition, DD/HD simulation requires some transport parameters (mobility, energy relaxation times,…). This thesis work offers a reliable infrastructure for device simulation of SiGe HBTs, including an exhaustive survey of HD models capability and limitations and providing a complete set of analytical, calibrated and verified models for transport parameters.

I.2. Thesis contents

In Chapter 1, after a brief introduction to state-of-the-art SiGe heterostructure bipolar transistors and to their proper figures of merit (FoM), a neat overview of hierarchical TCAD tool architecture is traced.

In Chapter 2, inherent approximations in the HD formalism are discussed. Performing device simulation with standard nonlocal models and default parameters values, anomalous and unphysical effects appear, resulting in a negative slope for output characteristics. A detailed study focusing on the relation between terminal quantities (i.e. currents) and simulation parameters for bipolar transistor is reported, evaluating the link between the parameters and the above mentioned negative slope, providing assessment on rule played by each equation parameter in simulation results and tracing optimization procedure to obtain predictive simulation results. Analysis has been carried over different technology nodes, providing for the first time a complete survey of HD models capability and restrictions with scaling.

In Chapter 3, a complete set of models for transport parameters for HD device simulation is reported, including low-field mobility, energy relaxation time, saturation velocity, high-field mobility and effective density of state. Transport models for DD/HD simulation available in the literature refer mainly to silicon, and only partially include the dependence on all relevant parameters (Ge content, strain, doping, temperature,…). If not properly calibrated, HD simulation results can be strongly inaccurate, if not even unphysical. For this reason, analytical transport models for HD simulation of strained SiGe devices have been generated from MC data and, where possible, compared with experimental data.

In Chapter 4, implementation and verification of the novel transport models in a commercial device simulator is drawn. HD simulations including new analytical transport models are reported for several one-dimensional and two-dimensional structures with different f_T maximum. Findings are compared with simulation results obtained using a standard set of models and with trustworthy results (i.e. MC and SHE simulation results and experimental data), validating

proposed models and clarifying their reliability and accuracy over different technologies.

Finally, in Chapter 5, electrical breakdown phenomena in SiGe HBTs are analyzed. After an accurate calibration of avalanche generation models for non-local device simulation and a review of main models for multiplication factor (M), a novel complete model for M is reported, providing an exhaustive accuracy over a wide range of collector voltages, as required for a correct description of the pinch-in effect. The new model, combined with an analytical model for the base current-dependent base resistance, is suitable for being incorporated into HBTs compact models to properly describe device operation above BV_{CEO}.

Chapter 1

Silicon-germanium heterojunction bipolar transistors

The performance of semiconductor devices tends to improve as device dimensions shrink. This simple principle of scaling has been the key to the spectacular success of semiconductor industry over the past half-century. It has worked for virtually all types of transistors, including the Si-based bipolar transistor [1]. Historically, scaling has run into apparent hard limits multiple times in the course of bipolar technology evolution, which have been successfully overcome with help from material and structural innovations, such as the self-aligned base, poly emitter, epitaxial base, and, most recently, the SiGe base. SiGe heterojunction bipolar transistors (HBTs), which incorporate such a SiGe base and benefit from the availability of "bandgap engineering" owing to the smaller bandgap of SiGe than that of Si, are widely used for low-cost high frequency applications.

Excellent speed performance is favored for most practical semiconductor applications of today. The efficiency of information processing strongly depends on the speed of devices that compose the system. The speed of a device can be represented by various measures and corresponding speed parameters. Although some alternatives have been proposed [2], the most widely used speed parameters for bipolar transistors are the cutoff frequency (f_T) and the maximum oscillation frequency (f_{MAX}), which are defined as the frequency point where the current gain and the power gain become unity, respectively. Since the breakdown voltage introduces a trade-off with high frequency figures-of-merit (FoM) in SiGe HBTs, the definition of Safe Operating Area (SOA) limits is a pivotal issue for designers and technologists. SOA limits specification is not a trivial task, as the maximum usable output voltages and currents depend on the driving conditions at the input port [3]. An additional difficulty is related to the fact that impact ionization mechanisms concur and interact with other limiting mechanisms, i.e. self-heating and hot-carrier degradation.

The study of device performance and the exploration of new architecture to improve FoM demand predictive device simulation and technology computer-aided design (TCAD) tools.

1.1 History, state-of-art and applications

The bipolar junction transistor (BJT) was the first solid-state amplifier element and started the solid-state electronics revolution. Bardeen, Brattain and Shockley, while at Bell Laboratories, invented it in 1948 as part of a post-war effort to replace vacuum tubes with solid-state devices. Solid-state rectifiers were already in use at the time and were preferred over vacuum diodes because of their smaller size, lower weight and higher reliability. A solid-state replacement for a vacuum triode was expected to yield similar advantages. The work at Bell Laboratories was highly successful and culminated in Bardeen, Brattain and Shockley receiving the Nobel Prize in 1956. Their work led them first to the point-contact transistor and then to the bipolar junction transistor [4]. They used germanium as the semiconductor of choice because it was possible to obtain high purity material. The extraordinarily large diffusion length of minority carriers in germanium due to its high minority carrier mobility provided functional structures despite the large dimensions of the early devices.

Although the small bandgap of germanium results in a small voltage drop across germanium pn junctions under forward bias and accordingly in reduced power dissipation, the resulting intrinsic carrier density is much larger in germanium than in silicon. As a consequence, germanium junctions suffer from a substantial reverse current that increases rapidly with temperature. This, along which many others advantages of silicon, was quickly recognized and, since the application of the Czochralski method to the growth of single crystal silicon, the silicon technology has progressed rapidly. The development of a planar process yielded the first circuits on a chip and bipolar transistor operational amplifiers, like the 741, and digital TTL circuits were for a long time the workhorses of any circuit designer.

In the seventies, the spectacular rise of the MOSFET market share has completely removed the bipolar transistor from center stage.

Almost all logic circuits, microprocessor and memory chips contain exclusively MOSFETs. Nevertheless, bipolar transistors remain important devices for ultra-high-speed discrete logic circuits such as emitter coupled logic (ECL), for power-switching applications and in microwave power amplifiers. Meanwhile, the heterojunction bipolar transistor (HBT) has emerged as the device of choice for cell phone amplifiers and other demanding applications.

The principle of heterojunction bipolar transistors has been known since the 1950s [5]: wide-bandgap emitters were recognized as a means to reduce minority-carrier injection into the emitter region, resulting in increased current gain and cut-off frequency. The realization of hetero-junctions in silicon technology, however, failed because the formation of silicon alloys always results in a significant change of the lattice constant, preventing from the formation of single-crystal heterostructures; therefore, heterojunction bipolar transistors were the exclusive domain of compound semiconductor technology. Successively, novel epitaxial techniques made it possible to grow very thin layers of SiGe on silicon crystals, with the same lattice constant. This is not possible without strain, thence solely thin layers can be grown without the formation of dislocations, which would relax the strain due to the different lattice constant.

The first SiGe HBT has been reported in the late eighties [6]. The use of heterojunctions provides an additional degree of freedom, which can result in vastly improved devices compared to the homojunction counterparts. Germanium has a smaller bandgap compared to silicon; since the current gain depends exponentially on the difference in bandgap energy between base and emitter regions [4], it's possible to obtain a very large current gain in a heterojunction bipolar transistor though the base doping N_B is significantly larger that the emitter doping density N_E and the base can be much thinner even for the same punchthrough voltage. As a result, one can reduce the base transit time without increasing the emitter charging time, while maintaining the same emitter current density.

The advantages of SiGe HBTs are mostly in RF performances. Most of all, RF performance of SiGe HBT has been strongly improved by grading the composition of the germanium base layer such that it causes an electric field which reduces the transit time. Therefore, SiGe HBTs are widely used for high-frequency applications allowing to reach a good performance with the limited cost of BiCMOS

technology. SiGe technology is the driving force behind the explosion of low-cost, lightweight, personal communications devices like digital wireless handsets, as well as other entertainment and information technologies like digital set-top boxes, Direct Broadcast Satellite (DBS), automobile collision avoidance systems, and personal digital assistants. SiGe extends the life of wireless phone batteries, and allows smaller and more durable communication devices. Products combining the capabilities of cellular phones, global positioning, and Internet access in one package, are being designed using SiGe technology. These multifunction, low-cost, mobile client devices capable of communicating over voice and data networks represent a key element of the future of computing.

This thesis work has been developed within the frame of the european project DOTFIVE [7], whose aim is the development of SiGe HBTs with a maximum oscillation frequency of 500 GHz. At the present time, transit frequencies of 265 GHz and maximum oscillation frequencies of 400 GHz have been reached as record peak at room temperature for conventional double-polysilicon fully self-aligned selective epitaxial growth HBT [8], as highlighted in the high frequency performance trend depicted in Fig. 1.1.

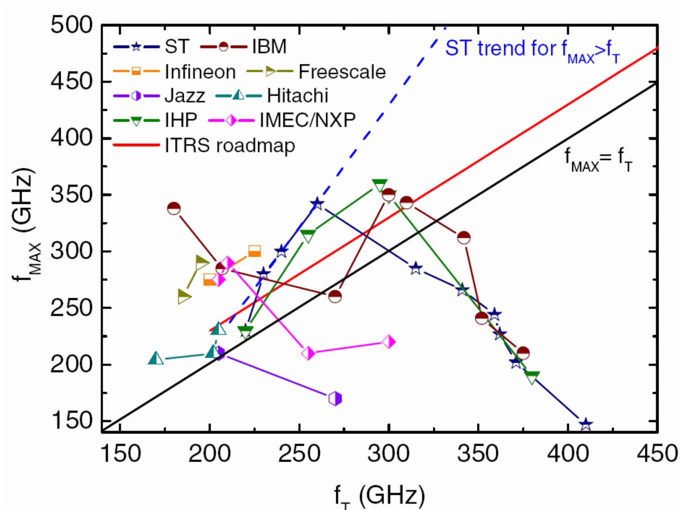


Fig. 1.1. f_{MAX} vs f_T trade-off: current state-of-the art vs. STMicroelectronics trend for HBTs featuring $f_{MAX} \geq f_T$.

1.2 Figures of Merit.

The cut-off frequency f_T and the maximum frequency of oscillation f_{max} are widely used figures of merit (FoM) to characterize high-frequency bipolar technologies. Since the parameters of a bipolar transistor are correlated and a modification that improves one parameter may cause a deterioration of another one, in order to obtain a figure of merit that takes account for how well a trade-off between different parameters is performed, the product of their values is frequently considered [4]. However, for devices towards the terahertz range just the product of the collector-emitter breakdown voltage and the cut-off frequency f_T could be of interest. The trade of between the breakdown voltage and the cut-off frequency is also known as the Johnson limit [9], [10].

1.2.1 Cut-off frequency f_T.

The cut-off frequency f_T represents the frequency at which the gain of a bipolar transistor drops to unity, as illustrated in Fig. 1.2. Beyond this frequency the gain of the transistor is less than unity, so it is no longer useful as either an amplifying or a switching device. In practice, it becomes increasingly difficult to design circuits as the required circuit operating frequency approaches the cut-off frequency of the transistor. More precisely, the cut-off frequency of a bipolar transistor is defined as the frequency at which the extrapolated common emitter, small-signal current gain drops to unity under conditions of a short-circuit load.

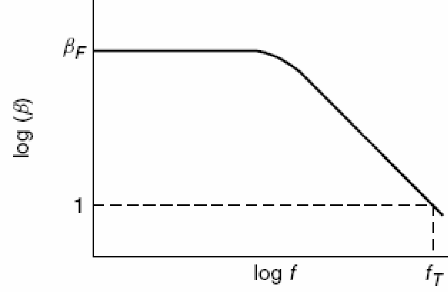


Fig. 1.2. Variation of small-signal current gain with frequency and definition of the cut-off frequency f_T.

Starting from the small-signal circuit model [11] and applying the f_T definition leads to:

$$f_T = \frac{1}{2\pi\left(\tau_F + \frac{kT}{qI_C}\left(C_{JE} + C_{JC}\right)\right)} \tag{1.1}$$

where C_{JE} and C_{JC} are the emitter/base and base/collector depletion capacitances and τ_F is the forward transit time given by:

$$\tau_F = \tau_E + \tau_{EBD} + \tau_B + \tau_{CBD} \tag{1.2}$$

Transit times τ_E, τ_{EBD}, τ_B and τ_{CBD} are associated with the excess minority carrier charges in the neutral emitter, the emitter/base depletion region, the base and the collector/base depletion region respectively.

In the case of an 1-D-device approximation, the electron total transit time is given as:

$$\tau(0,L) = \int_0^L q \left.\frac{dn}{dI_C}\right|_{V_{CE}=const} dx \tag{1.3}$$

where L is the device length, and the total transit time is inversely proportional to the cutoff frequency:

$$f_T = \frac{1}{2\pi\tau} \tag{1.4}$$

In this thesis the transit time and cutoff frequency are calculated in the quasi–stationary approximation of (1.3) and (1.4), in accordance with the definition given in [12].

1.2.2 Maximum oscillation frequency f_{max}.

Another important high-frequency parameter for a bipolar transistor is the maximum oscillation frequency f_{max}. This is defined as the frequency at which the power gain drops to unity. An approach similar to that followed for the evaluation f_T can be used to derive an expression for f_{max} [4]:

$$f_{\max} = \sqrt{\frac{f_T}{8\pi C_{JC} R_B}} \tag{1.5}$$

Since f_{max} is positively correlated to f_T, as indicated by (1.5), the benefits of vertical scaling on f_T also apply to f_{max}, although the impact is relatively lower. However, f_{max} depends on the base resistance R_B and on the collector/base capacitance C_{JC}, which are actually degraded by vertical scaling [1]. On the other hand, lateral scaling, which has only a limited influence on f_T, plays a major impact on f_{max} [1].

1.2.3 Impact ionization.

Generation-recombination processes are processes that exchange carriers between the conduction band and the valence band and are very important in the operation of bipolar devices. Avalanche multiplication or impact ionization is by far the most common breakdown mechanism in practical bipolar transistors [11]. In a reverse biased pn junction, electron-hole pairs are continually being generated by thermal agitation. At low reverse voltages this gives rise to a leakage generation current, but at high reverse voltages the generated carriers gain sufficient kinetic energy between collisions with the silicon lattice to be able to shatter the silicon-silicon bonds.

This mechanism is referred to as impact ionization and leads to the generation of an electron–hole pair. The original carrier and the electron and hole generated are then accelerated in opposite directions by the electric field, and in turn are able to produce further electron–hole pairs by impact ionization. This process, known as avalanche multiplication, rapidly leads to the generation of a large number of carriers and hence to a large current. For avalanche multiplication to occur, a critical electric field E_{crit} must be established across the reverse-biased junction. Since the depletion width depends upon the doping concentration, it is clear that the breakdown voltage BV will also depend on the doping concentration. For a one-sided step junction the breakdown voltage is given by [4]

$$BV = \frac{\varepsilon_0 \varepsilon_r E_{crit}^2}{2qN_L} \tag{1.6}$$

where N_L is the doping concentration of the lightly doped side of the junction.
In bipolar transistors, the breakdown voltage depends on the way the bipolar transistor is connected in the circuit. In the common base connection, the breakdown voltage of the collector junction is the same as that predicted by equation (1.6), whereas in the common emitter connection the breakdown voltage is considerably lower, [11]. In practice, the breakdown voltage in bipolar transistors is measured with the emitter open circuit, and hence in common base mode the breakdown voltage is referred to as BV_{CBO} (breakdown voltage in common base connection with the emitter open circuit). In the common emitter mode the breakdown voltage is referred to as BV_{CEO} (breakdown voltage in common emitter connection with the base open circuit).

The lower breakdown voltage in a common emitter connection can be understood by considering the currents flowing in the transistor when it is connected in a common emitter configuration. With reference to Fig. 1.3, if the current flowing across the emitter/base junction is I_F, a fraction of this current is collected at the collector/base junction, given by $\alpha \cdot I_F$, where α is the common base current gain. In addition, there will be a component of current at the

collector due to the leakage current of the collector/base junction I_{CBO}. In this case, we can write:

$$I_E = I_F$$
$$I_C = \alpha I_F + I_{CBO} \tag{1.7}$$

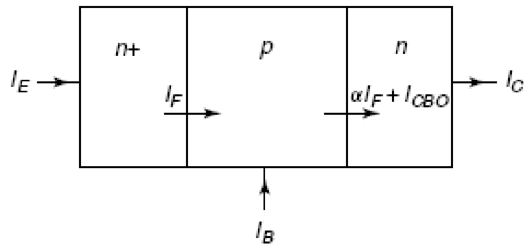


Fig. 1.3. Schematic view of current flow in forward active bias mode.

When the collector/base junction is breaking down, the current across the junction is multiplied by the electron–hole pairs created by avalanche breakdown. In this case, the current at the collector/base junction is multiplied by M which yields.

$$I_C = M\left(\alpha I_F + I_{CBO}\right) \tag{1.8}$$

being M is the collector current multiplication factor due to impact ionization [13].

If the base is open circuit, the emitter current must equal the collector current, so that equation (1.8) becomes:

$$I_C = -I_E = I_C = M\left(\alpha I_C + I_{CBO}\right) \tag{1.9}$$

From (1.9) we obtain

$$I_{CEO} = \frac{MI_{CBO}}{1-\alpha M} \tag{1.10}$$

where I_{CEO} is the current flowing between emitter and collector when the base is open circuit.

Equation (1.10) shows that the collector/emitter current begins to increase very rapidly when $\alpha \cdot M$ approaches unity. In contrast, in the common base mode the collector/base leakage current only begins to

increase when M approaches infinity. This explains why the breakdown voltage in the common emitter mode BV_{CEO} is lower than that in the common base mode BV_{CBO}.

1.3 Carrier transport models and TCAD for SiGe HBTs development

Device simulation is now an integral part of SiGe technology development, and is routinely used for understanding SiGe HBT operation and device optimization, [14]. All major commercial device simulators support SiGe simulation [15]; they are typically part of a technology computer-aided-design (TCAD) package, which includes process simulation, device simulation, and parameter extraction programs.

Due to the rapid evolution of semiconductor technologies, the use of technology computer-aided design (TCAD) has become a critical enablement path in the industrial development cycle. The capability to perform numerical process and device simulation that is able to accurately predict device performance under various process conditions can significantly reduce the time and cost of technology development. With experimental wafer processing costs increasing dramatically for advanced generation technologies, the economic necessity for accurate predictive TCAD is apparent. But it is not only the wafer cost and device complexity that is continuing to drive the need for accurate predictive SiGe TCAD. Unlike high-performance CMOS logic that can have years between generations, the analog and mixed-signal market that encompasses the majority of SiGe applications demands a significantly reduced product cycle time [15]. This aggressive development schedule is generating a new paradigm for the relationship between TCAD, process development, and circuit designers.

The traditional relationship of TCAD with technology development consisted of having SIMS profiles, TEMs, and various other physical and electrical measurements provided to the TCAD engineer to calibrate the process and device models in parallel to the process development effort, as shown in Fig. 1.4. The flow is then

iterated until target device characteristics are achieved and representative hardware can be furnished to the compact modeling team. The compact modeling team must then fully characterize the devices to generate a physically based model parameter set, which is then combined with the complete technology design kit. Even if the process of record is known, circuit designers cannot begin to design for a significant period of time before the processing, characterization, and design kit cycle is complete. With accurate process and device simulation early in the development cycle, the TCAD engineer can create the electrical device characteristics to derive an initial predictive compact model. The compact model can then be provided to circuit designers in a substantially earlier timeframe than traditionally available, based on calibrated physical simulations. As the technology development progresses, updates to the compact models can be provided to ensure the designers will have minimal design impact from the initial models to the final hardware-based models.

The main issue in the simulation of scaled devices is related to the limitations of the physical models used to describe charge carrier transport. A widely used approach is the so called drift-diffusion (DD) model, which treats carrier transport as diffusion and drift processes [16]. However, as the device size approaches the nanometer range, charge transport becomes quasi-ballistic, and non-local effects such as velocity overshoot occur. In the attempt to capture these phenomena, more advanced transport models have been proposed, often termed hydrodynamic (HD) or energy-transport models.

Both the DD and HD models can be viewed as approximations of the Boltzmann transport equation (BTE), which represents the most rigorous approach to model charge carrier transport in semiconductors [17]. The BTE is an integro-differential equation in the six-dimensional phase space, which can be solved by a stochastic approach, the Monte Carlo (MC) method [16], or by a deterministic approach, the spherical harmonics expansion (SHE) method [18].

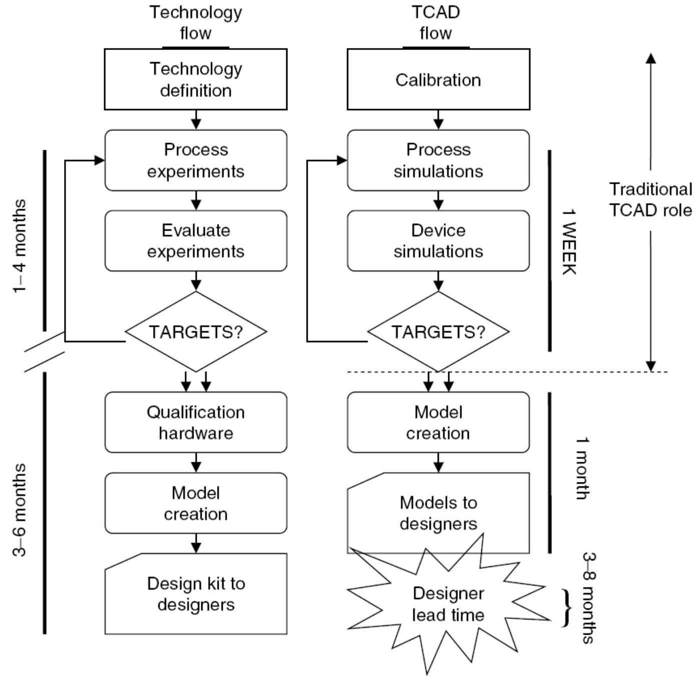


Fig. 1.4. Flow diagram showing the traditional TCAD relationship with technology development and the new extended paradigm [15].

1.3.1 Boltzmann transport equation

Transport equations used in semiconductor device simulation are normally derived from the BTE, a semiclassical kinetic equation, which reads [19]:

$$\partial_t f + \mathbf{u} \cdot \nabla_{\mathbf{r}} f + \frac{\mathbf{F}}{\hbar} \cdot \nabla_{\mathbf{k}} f = C[f] \tag{1.11}$$

where $f(\mathbf{k},\mathbf{r},t)$ represents the carrier distribution function in the six-dimensional phase space, and the term on the right side represents the rate of change due to collisions. The BTE is valid for general inhomogeneous materials with arbitrary band structure [20]. The group velocity $\mathbf{u}$ is defined as:

$$\mathbf{u}(\mathbf{k},\mathbf{r}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon\left(\mathbf{k},\mathbf{r}\right) \tag{1.12}$$

where ε represents the carrier kinetic energy.

The BTE is an integro-differential equation in the seven-dimensional space ($\mathbf{k}$,$\mathbf{r}$,t). To solve this equation numerically by discretization of the differential and integral operators is computationally very expensive.

1.3.2 The drift-diffusion model

At the very beginnings of semiconductor technology, electrical device characteristics could be estimated using simple analytical models relying on the drift-diffusion (DD) formalism. Various approximations have to be made to obtain DD equations from BTE in §1.3.1, but the resulting model captures the basic features of the devices.

The complete drift-diffusion model is based on the following set of equations [13]:

1. Current equations:

$$\begin{aligned} J_n &= qn\mu_n\mathbf{E} + qD_n\nabla n \\ J_p &= qp\mu_p\mathbf{E} - qD_p\nabla p \end{aligned} \tag{1.13}$$

2. Continuity equations:

$$\begin{aligned} \frac{\partial n}{\partial t} &= \frac{1}{q}\nabla\cdot\mathbf{J}_n + U_n \\ \frac{\partial p}{\partial t} &= -\frac{1}{q}\nabla\cdot\mathbf{J}_p + U_p \end{aligned} \tag{1.14}$$

3. Poisson's equation:

$$\nabla\cdot\varepsilon\nabla V = -\left(p - n + N_D^+ - N_A^-\right) \tag{1.15}$$

The continuity equations are the conservation laws for the charge carriers, which may be easily derived taking the zeroth moment of the

time dependent BTE. The mobilities μ_n, μ_p and the net generation-recombination rates U_n and U_p have to be modeled as functions of temperature, carrier concentration and electric field strength [4].

Numerical device simulation based on the carrier transport equations (1.13) ÷ (1.15) dates back to the famous work of Scharfetter and Gummel [21], who proposed a robust discretization of the DD equations which is still in use today. The DD model is the simplest current transport model which can be derived from Boltzmann's transport equation by either the method of moments [16] or from basic principles of irreversible thermodynamics [22]. For many decades, the DD model has been the backbone of semiconductor device simulation. In this model, the electron current density is phenomenologically expressed as consisting of two components. The drift component is driven by the electric field and the diffusion component by the electron density gradient.

However, as semiconductor devices were scaled into the submicrometer regime, the assumptions underlying the DD model lost their validity. Therefore, the transport models have been continuously refined and extended to more accurately capture transport phenomena occurring in submicrometer devices. As device dimensions shrink and doping levels increase to improve high frequency performance, the electric field inside the devices increases. A large electric field which rapidly changes over small length scales gives rise to nonlocal and hot-carrier effects, which begin to dominate device performance. An accurate description of these phenomena is required and is becoming a primary concern for industrial applications. To overcome some of the limitations of the DD model, extensions have been proposed. These extensions basically add an additional balance equation for the average carrier energy and an additional driving term in the current relation; the additional driving term is proportional to the gradient of the carrier temperature.

1.3.3 Energy transport and hydrodynamic models

Modeling of deep-submicrometer devices with the DD model is becoming more and more problematic. Although successful

reproduction of terminal characteristics of nanoscale devices has been reported with the DD model [23], the values used for the physical parameters significantly violate basic physical principles. In particular, the saturation velocity had to be set to more than twice the value observed in bulk measurements. This implies that the model is no longer capable of reproducing the results of bulk measurements and as such loses its consistency. Furthermore, the model can hardly be used for predictive simulations. These solutions may provide short-term fixes to available models, but obtaining “correct” results from the wrong physics is unsatisfactory in the long run [19].

To overcome some of the limitations of the DD model, energy balance and hydrodynamic models have been developed. However, a vast number of these models exists, and there is a considerable amount of confusion as to their relation to each other. Here the most important models, available in commercial device simulators [24], are summarized. The HD model equations read:

$$\begin{aligned} J_n &= \mu_n \left(n\nabla E_C + kT_n \nabla n + f_n^{td} kn\nabla T_n - 1.5knT_n \nabla \ln m_n \right) \\ J_p &= \mu_p \left(p\nabla E_V - kT_p \nabla p - f_p^{td} kp\nabla T_p - 1.5kpT_p \nabla \ln m_p \right) \end{aligned} \tag{1.16}$$

$$\begin{aligned} \frac{\partial W_n}{\partial t} + \nabla \cdot S_n &= J_n \cdot \nabla E_C + \left.\frac{dW_n}{dt}\right|_{coll} \\ \frac{\partial W_p}{\partial t} + \nabla \cdot S_p &= J_p \cdot \nabla E_V + \left.\frac{dW_p}{dt}\right|_{coll} \\ \frac{\partial W_L}{\partial t} + \nabla \cdot S_L &= \left.\frac{dW_L}{dt}\right|_{coll} \end{aligned} \tag{1.17}$$

$$\begin{aligned} S_n &= -\frac{5r_n}{2}\left(\frac{kT_n}{q} J_n + f_n^{hf} \kappa_n \nabla T_n \right) \qquad & \kappa_n &= \frac{k^2}{q} n\mu_n T_n \\ S_p &= -\frac{5r_p}{2}\left(\frac{kT_p}{q} J_p + f_p^{hf} \kappa_p \nabla T_p \right) \qquad & \kappa_p &= \frac{k^2}{q} p\mu_p T_p \\ S_L &= -\kappa_L \nabla T_L \end{aligned} \tag{1.18}$$

where (1.16) describe current densities, (1.17) represent energy balance equations and (1.18) are energy flux densities. Parameters r, f^{td} and f^{hf} are accessible and allow to choice between different ET/HD models, as it will be discussed in Chap. 2.

1.3.4 The Monte Carlo model

A widely used numerical method for solving the BTE is the Monte Carlo (MC) method. The full-band MC (FB-MC) model is currently the most popular simulation method within the framework of semi-classical device physics [28]. Another important feature of the FB-MC model is that it captures the full anisotropy of the band structure, which plays a significant role in the quasi-ballistic transport in deep submicron devices.

Despite its considerable success, the FB-MC method is applied in this thesis only to perform exemplary simulations. Due to its computational expense, it is used to generate reference results, to be used for calibration of lower level accurate semi-classical device models. Additionally, a proper analysis of the stochastic error must be performed to assess the accuracy and efficiency of Monte Carlo device simulations [25].

MC simulation results reported in this thesis work have been generated by Bundeswehr University using FB-MC developed by Prof. Jungemann and described in [17].

1.3.5 The spherical harmonics expansion model

For extremely scaled devices, carrier transport cannot be described accurately by momentum-based models (drift-diffusion or hydrodynamic models) [18]. In such a case, a full solution of the BTE is required. Although the Monte Carlo approach is the standard method to solve the Boltzmann equation, it has many disadvantages due to its stochastic nature [25]. A deterministic Boltzmann equation solver based on the spherical harmonics expansion (SHE) of the

distribution function is a viable alternative to the Monte Carlo approach [26], [27]. Although its application to device simulation was hampered for a long time mainly due to its huge memory consumption, the exponential growth of computer memory in the last decades makes this method more and more attractive.

A general higher order SHE solver which includes anisotropic band structure for the conduction band, captures FB effects and can be used with a higher order SHE solver has been developed by *Bundeswehr University* and described in [27]. This simulator enables accurate modeling of the quasi-ballistic transport in nanoscale devices and of high-energy effects, like impact ionization. It has been verified that this SHE solver, called *SPRING*, provides simulation results comparable to results of a FB Monte Carlo simulator.

The *SPRING* simulator is available on our cluster by courtesy of Prof. Jungemann and has been used to generate all SHE simulations reported in this thesis work.

1.4 References

[1] 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, September 2005.

[2] R. A. Gosser, O. Foroudi, and S. Flanyak, “New bipolar figure of merit ‘fo’,” *BIPOLAR/BiCMOS Circuits and Technology Meeting Proceedings*, pp. 128–135, 2002.

[3] J. D. Cressler, “Emerging SiGe HBT Reliability Issues for Mixed-Signal Circuit Applications,” *IEEE Transactions on Device and Materials Reliability*, vol. 4, no. 2, pp. 222-236, June 2004.

[4] M. Reisch, High-frequency Bipolar Transistors: Physics, Modelling, Applications, Springer, 2003.

[5] H. Kroemer, “Theory of wide-gap emitter for transistors,” *Proceedings of the IRE*, vol. 45, no. 11, pp. 1535-1537, November 1957.

[6] S. S. Iyer, G. L. Patton, S. S. Delage, S. Tiwari, and J. M. C. Stork, “Silicon-Germanium Base Heterojunction Bipolar

Transistors by Molecular Beam Epitaxy," *Proceeding of IEEE International Electron Device Meeting*, pp. 874-876, 1987.
[7] http: //www.dotfive.eu
[8] P. Chevalier , F. Pourchon, T. Lacave, G. Avenier, Y. Campidelli, L. Depoyan, G. Troillard, M. Buczko, D. Gloria, D. Céli, C. Gaquière, and A. Chantre, "A conventional Double-Polysilicon FSA-SEG Si/SiGe:C HBT Reaching 400 GHz fmax," *BIPOLAR/BiCMOS Circuits and Technology Meeting Proceedings*, pp. 1-4, October 2009.
[9] E. Johnson, "Physical limitations on frequency and power parameters of transistors," *Proceeding of IRE International Convention Record*, pp. 27-34, March 1965.
[10] K. K. Ng, M. Frei, C. A. King, "Reevaluation of the fT-BV_{CEO} limit on Si bipolar transistors", *IEEE Transactions on Device and Materials Reliability*, vol. 45, no.8, pp. 1854-1855, August 1998.
[11] P. Ashburn, SiGe Heterojunction Bipolar Transistors, John Wiley & Sons, 2003.
[12] H. K. Gummel, "On the definition of the cutoff frequency f_T," *Proceedings IEEE*, vol. 57, no. 12, p. 2159, December 1969.
[13] S. M. Sze, Physics of Semiconductors Devices. New York, Wiley, 1981.
[14] M. Al-Sa'di, V. d'Alessandro, S. Fregonese, S.-M. Hong, C. Jungemann, C. Maneux, I. Marano, A. Pakfar, N. Rinaldi, G. Sasso, M. Schröter, A. Sibaja-Hernandez, C. Tavernier, and G. Wedel, "TCAD simulation and development within the European DOTFIVE project on 500GHz SiGe:C HBT's", *European Microwave Integrated Circuits Conference*, pp. 29-32, September 2010.
[15] J. D. Cressler, The Silicon Heterostructure Handbook: Materials, Fabrication, Devices, Circuits, and Applications of SiGe and Si Strained-Layer Epitaxy. CRC Press, New York.
[16] S. Selberherr, Analysis and Simulation of Semiconductor Devices. Wien, New York, Springer, 1984.
[17] C. Jungemann and B. Meinerzhagen, Hierarchical Device Simulation -The Monte-Carlo Perspective. Berlin, Springer, 2003.
[18] S.-M. Hong, and C. Jungemann, "Electron Transport in Extremely Scaled SiGe HBTs," *BIPOLAR/BiCMOS Circuits and Technology Meeting Proceedings*, pp. 67-74, 2009.

[19] T. Grasser, T. W. Tang, H. Kosina, and S. Selberherr, "A Review of Hydrodynamic and Energy-Transport Models for Semiconductor Device Simulation," *Proceedings IEEE*, vol. 91, no. 2, pp. 251-274, Feb. 2003

[20] E. M. Azoff, "Generalized energy-momentum conservation equation in the relaxation time approximation," *IEEE Journal of Solid-State Electronics* vol. 30, no. 9, pp. 913-917, September 1987.

[21] D. L. Scharfetter, and H. K. Gummel, "Large-signal analysis of a silicon read diode oscillator," *IEEE Transactions on Electron Devices*, vol. 16, no. 1, pp. 64–77, January 1969.

[22] G. K. Wachutka, "Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling," *IEEE Transaction on Computer-Aided Design*, vol. 9, no. 11, pp. 1141-1149, November 1990.

[23] J. D. Bude, "MOSFET modeling into the ballistic regime," *Proceeding of Simulation of Semiconductor Processes and Devices*, pp. 23-26, 2000.

[24] Synopsys TCAD Software, Release 2007.03.

[25] C. Jungemann, and B. Meinerzhagen, "Analysis of the stochastic error of stationary Monte Carlo device simulations," *IEEE Transactions on Electron Devices*, vol. 48, no. 5, pp. 985–992, May 2001.

[26] D. Ventura, A. Gnudi, G. Baccarani, and F. Odeh, "Multidimensional spherical harmonics expansion of Boltzmann equation for transport in semiconductors," *Applied Mathematics Letters*, vol. 5, no. 3, pp. 85-90, May 1992.

[27] S.-M. Hong, G. Matz, and C. Jungemann, "A Deterministic Boltzmann Equation Solver Based on a Higher Order Spherical Harmonics Expansion With Full-Band Effects," *IEEE Transactions on Electron Devices*, vol. 57, n. 10, pp. 2390-2397, October 2010.

[28] K. Hess, Monte Carlo Device Simulation: Full Band and Beyond. Kluwer, Boston, 1991.

Chapter 2

Hydrodynamic model verification

A large number of energy transport (ET) and hydrodynamic (HD) models have been developed during the last decades. Since these models base on simplifying assumptions, it' s necessary to verify their accuracy with more rigorous approaches, such as Monte Carlo (MC) or spherical harmonic expansion simulation (SHE). Several papers related to ET and HD models are available in the Literature, addressing the accuracy of HD/ET models. However, these studies always refer to simplified unipolar devices (n^+ - n^- - n^+ structure) or field effect transistors, and do not consider simulation issues in bipolar devices. In this thesis it will be shown that performing device simulation with standard nonlocal models and default parameters values, can yield anomalous and unphysical effects, such as a negative slope in the output characteristics. Therefore, although ET and HD models' limitations have been widely investigated and many references are available for parameters values, a detailed analysis of relation between terminal quantities (i.e. currents) and simulation parameters for bipolar transistors is available.

ET and HD models available in commercial device simulators are based on several and widely discussed assumptions, [1] and [2]. Although different and more sophisticated models have been proposed, e.g. [3], they are not implemented in commercial tools.

In this thesis a detailed analysis has been performed to assess the role played by each equation parameter in simulation results, thus providing a user-addressed tool to avoid anomalous and unphysical results, [4]. Although this work has been developed using *TCAD Sentaurus* by *Synopsys*, [5], the hydrodynamic models formulation and parameters definition are similar in all major commercial simulation codes, e.g. [6]. Therefore, the results presented here can be also applied to other device simulator. More specifically, the present analysis clarifies the role of each model's parameter and suggest a procedure to optimize the parameters to obtain predictive simulation results. All results presented in the first part of this work refer to a 100 GHz reference structure; later, the analysis is repeated for a 450 GHz

device and a 700 GHz structure. Therefore, a complete study on HD simulation of SiGe HBTs has been carried for different technological nodes, thus providing a complete survey of HD models accuracy and limitations in connection with device scaling.

2.1 Hydrodynamic model in commercial TCAD.

The HD equations implemented in *Sdevice* [5] read as the system of equations expressed by (2.1) ÷ (2.3), where current densities are given by equations (2.1), energy balance equations are given by relations in (2.2) and energy flux densities are equations (2.3).

$$\begin{aligned} J_n &= \mu_n \left(n\nabla E_C + kT_n \nabla n + f_n^{td} kn\nabla T_n - 1.5knT_n \nabla \ln m_n \right) \\ J_p &= \mu_p \left(p\nabla E_V - kT_p \nabla p - f_p^{td} kp\nabla T_p - 1.5kpT_p \nabla \ln m_p \right) \end{aligned} \tag{2.1}$$

$$\begin{aligned} \frac{\partial W_n}{\partial t} + \nabla \cdot S_n &= J_n \cdot \nabla E_C + \left.\frac{dW_n}{dt}\right|_{coll} \\ \frac{\partial W_p}{\partial t} + \nabla \cdot S_p &= J_p \cdot \nabla E_V + \left.\frac{dW_p}{dt}\right|_{coll} \\ \frac{\partial W_L}{\partial t} + \nabla \cdot S_L &= \left.\frac{dW_L}{dt}\right|_{coll} \end{aligned} \tag{2.2}$$

$$\begin{aligned} S_n &= -\frac{5r_n}{2}\left(\frac{kT_n}{q} J_n + f_n^{hf} \kappa_n \nabla T_n \right) \qquad & \kappa_n &= \frac{k^2}{q} n\mu_n T_n \\ S_p &= -\frac{5r_p}{2}\left(\frac{kT_p}{q} J_p + f_p^{hf} \kappa_p \nabla T_p \right) \qquad & \kappa_p &= \frac{k^2}{q} p\mu_p T_p \\ S_L &= -\kappa_L \nabla T_L \end{aligned} \tag{2.3}$$

The last terms on the right hand side of (2.1) account for the additional driving force due to the change in the effective masses in hetero-structure devices, so that the force related to the change in the

band edge energies is included in the valence and conduction band energy gradients.

Parameters r, f^{td} and f^{hf} are accessible by user in *TCAD Sentaurus* [5], and can be modified for both electrons and holes. Table 2.1 shows the parameter values according to the Stratton model [7], the Blotekjær model [7] and the four-moments model [9].

Table 2.1. HD parameters for the Stratton, the Blotekjær and the four moments model.

	$\mathbf{r_n = r_p}$	$\mathbf{f_n^{td} = f_p^{td}}$	$\mathbf{f_n^{hf} = f_p^{hf}}$
Stratton	variable range	0 ÷0.5	1
Blotekjær	1	1	variable range
Four-moments	variable range	1	1

The Stratton model [7], generally referred as an energy transport model, is derived from the Boltzmann Transport Equation (BTE) introducing certain assumptions on the form of the distribution function, while the Blotekjær [8] model equations are derived by considering the first three moments of the BTE. The most important difference between Stratton and Blotekjær approaches is the mobility definition. In the Stratton model the parameter f^{td} appearing in (2.1) is defined according to (2.4) after the rearrangement of (2.5),where the mobility is inside the spatial gradient:

$$f_n^{td} = 1 + \nu_n, \qquad \nu_n = \frac{T_n}{\mu}\frac{\partial \mu}{\partial T_n} = \frac{\partial \ln \mu}{\partial \ln T_n} \tag{2.4}$$

$$J_n = q\mu_n \left(n\nabla E_C + \frac{k}{q}\nabla\left(n\mu T_n\right) - 1.5\frac{k}{q} n T_n \nabla \ln m_n \right) \tag{2.5}$$

On the contrary, in the Blotekjær equations the mobility is outside the gradient, (2.6), so that f^{td} must be set to one:

$$J_n = q\mu_n \left(n\nabla E_C + \frac{k}{q}\mu\nabla\left(nT_n\right) - 1.5\frac{k}{q} n T_n \nabla \ln m_n \right) \tag{2.6}$$

This difference in the definition yields different mobility values for inhomogeneous devices, where the electric field varies rapidly. In the Blotekjær model the mobility in non-uniform devices can be approximated by its bulk value, while in the Stratton model is always different. Therefore, the Blotekjær formulation is more suitable for commercial device simulators, where the mobility is modeled as a function of the carrier energy only, without any dependence upon the electric field.

The method of moments transforms the BTE into an equivalent, infinite set of equations. To solve this equation set, a severe approximation is required, namely the truncation to a finite number of equations (normally three or four). The highest order equation contains the moment of the next order, which has to be suitably approximated using the available information. Therefore, the main problem of moment-based models, usually referred to as hydrodynamic models, is that they deliver more unknown than equations; this issue has to be solved by separate closure relations. The four-moments model is an extension of the Blotekjær model where the fourth moment of the BTE is included to overcome problems related to the energy flux density closure.

2.2 Analysis of model parameters for a 100 GHz device

In order to find a proper and possibly unique parameters set for SiGe HBT device simulation, the influence of each parameter on simulation accuracy has been first investigated by a simulation study of the 100 GHz reference structure, depicted in Fig. 2.1.

To clearly elucidate the influence of each parameter, the parameters have been modified individually (starting from the Blotekjær default set, where all parameters in Table 2.1 are set to 1). Simulation results have been compared to MC data. The MC data used for this analysis were provided by Bundeswehr University in Munich (Germany).

Comparison involves terminal currents for transfer and output characteristics, as well as internal quantities. For internal quantities we

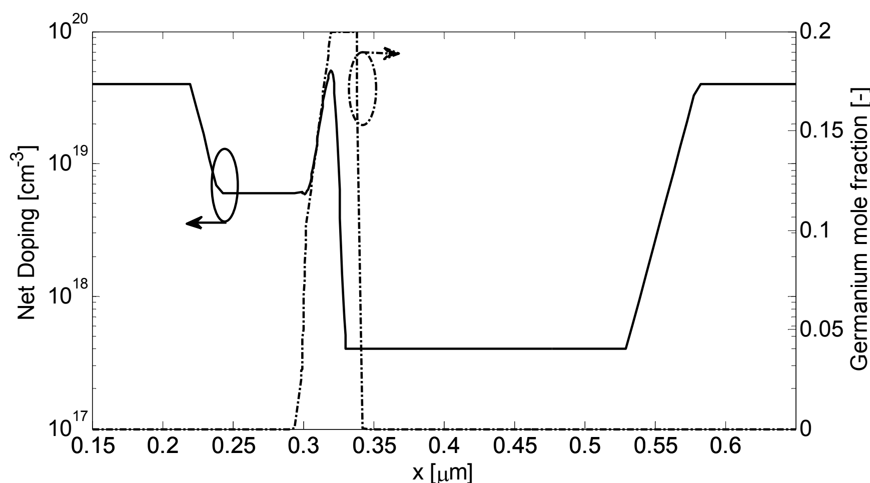


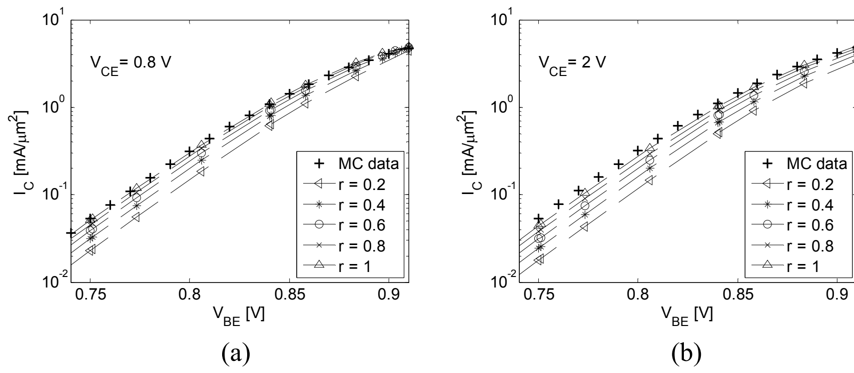
Fig. 2.1. Doping profile of the 100-GHz SiGe HBT.

only show the results pertaining the electron temperature and the electron velocity, as they play a key role in determining the frequency performance. Moreover, the electron velocity is closely related to the electron carrier density in the neutral base, as the product between carrier velocity and density is constant when generation-recombination process can be neglected. As a consequence, carrier velocity is linked to the collector current as well. In this comparison we also include the cut-off frequency. Additionally, carrier velocity plots are useful to clarify how velocity overshoot effects are modeled.

Velocity overshoot occurs at the base-collector junction, where the electric field increases rapidly. On the other hand, in HD simulation a spurious velocity overshoot located in the epitaxial collector region can be observed. This overshoot is referred to as "spurious" because it has no physical background and it is related to limitations of models. In [10] it was pointed out for the simplified unipolar structure (n^+-n^--n^+), that spurious velocity overshoot is strictly related to the closure relation in the moments definition, since the overshoot appears also for a six moments model but it's removed when the closure is taken from MC simulation.

2.2.1 Parameter *r*

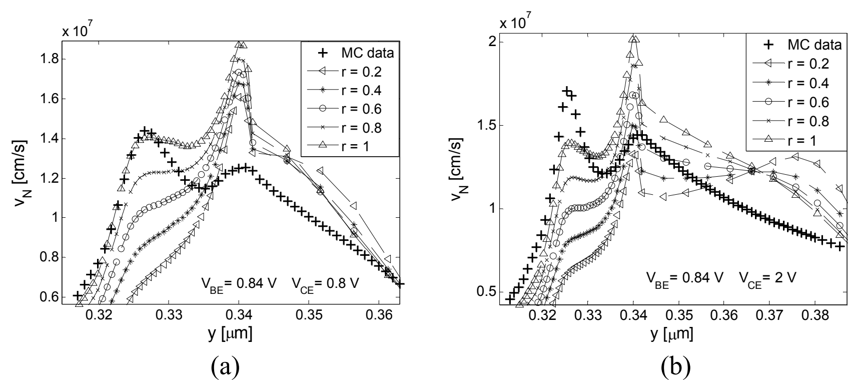
The parameter *r* affects the energy flux density (2.3). It appears in the Stratton formulation after defining the microscopic relaxation time using a power law and its impact on simulation results is reported in Fig. 2.2 for the 100 GHz reference structure.



(a) (b)

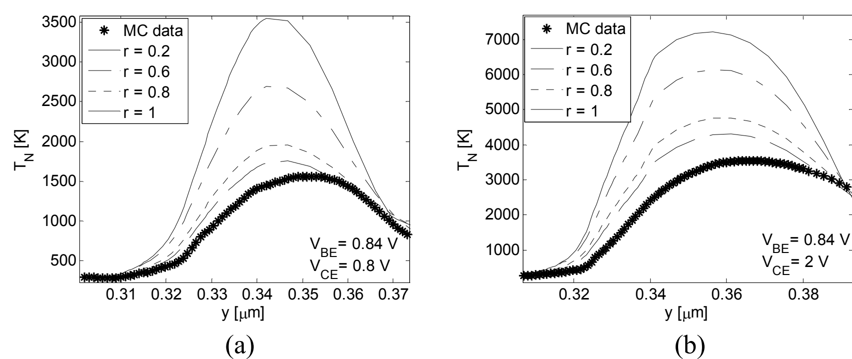
Fig. 2.2. Transfer characteristics at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b) for several *r* values.

The results depicted in Fig. 2.2 clarify that the reduction in energy flux densities yields lower collector current values. This result is expected, since the energy density gradient appears in the current density equations (2.5) and (2.6). It can be noted that the dependence of the current on the parameter *r* is more marked for V_{CE} = 2 V. A more detailed view (see Fig. 2.3 and Fig. 2.4) shows that decreasing *r* increases the electron spurious velocity overshoot, removes the real one at base-collector junction, and strongly overestimates electron temperature. This behaviour occurs for both values of the collector-emitter voltage, but velocity profiles become more complicated and unrealistic at V_{CE} = 2 V. Finally, as depicted in Fig. 2.5, the value of *r* has no impact on the output characteristic slope. This indicates that the absolute value of the energy density has no influence on the sign of the output characteristic slope.



(a) (b)

Fig. 2.3. Electron velocity for several values of *r* at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b); $V_{BE} = 0.84$ V.



(a) (b)

Fig. 2.4. Electron temperature at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b) for several values of *r*; $V_{BE} = 0.84$ V.

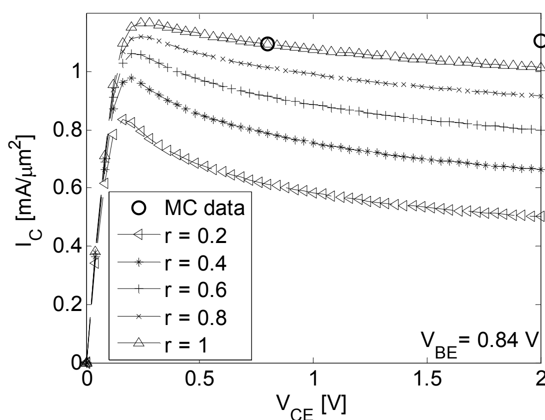


Fig. 2.5. Output characteristic for several values of *r* at $V_{BE} = 0.84$ V.

2.2.2 Parameter f^{td}

The parameter f^{td} modifies the value of the carrier temperature diffusive component in the current density with respect to the other components present in the drift-diffusion formulation, (2.1). In the moment-based models it should be set to 1.

As f^{td} is reduced, the collector current increases, yielding values higher than those given by MC simulations, as shown in Fig. 2.6. On the contrary, decreasing f^{td} yields a better fit of carrier velocity and temperature (see Fig. 2.7 and Fig. 2.8) and the negative output resistance disappears, Fig. 2.9.

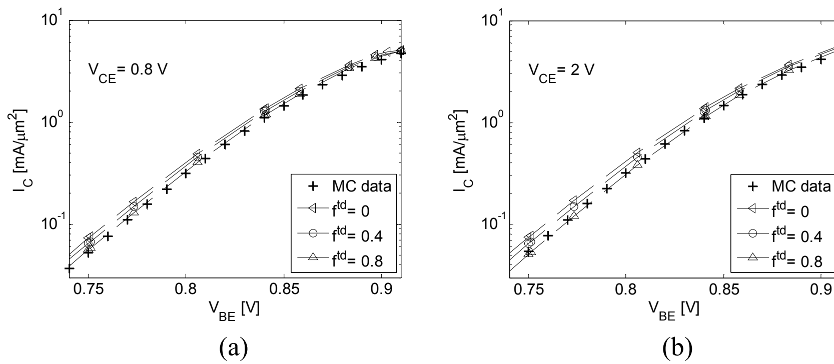


Fig. 2.6. Transfer characteristics for several values of f^{td} at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b).

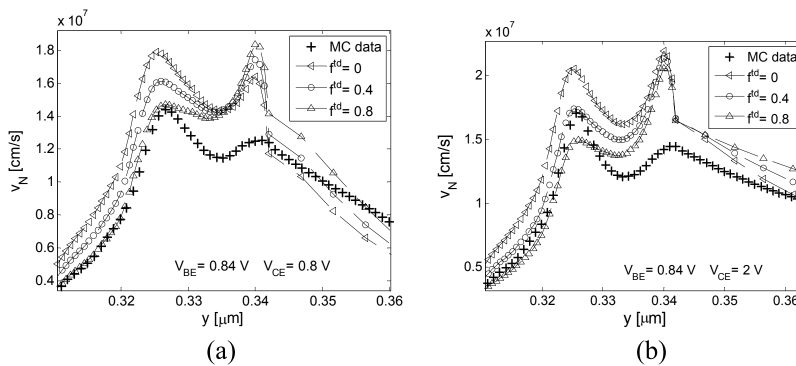


Fig. 2.7. Electron velocity at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b) for several f^{td} values; V_{BE} = 0.84 V.

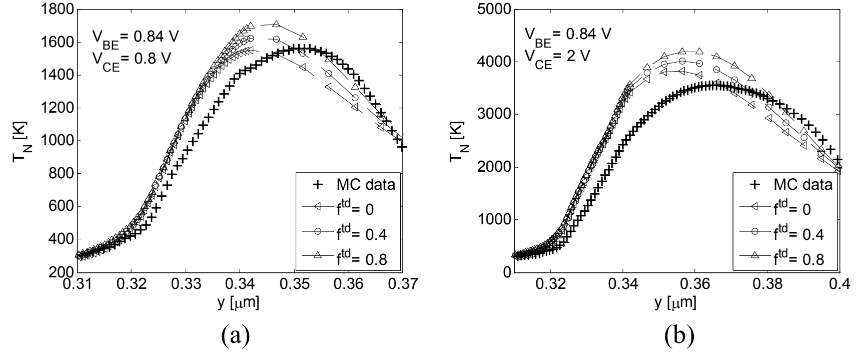


Fig. 2.8. Electron temperature at $V_{CE} = 0.8$ V and $V_{CE} = 2$ V for several values of f^{td}; $V_{BE} = 0.84$ V.

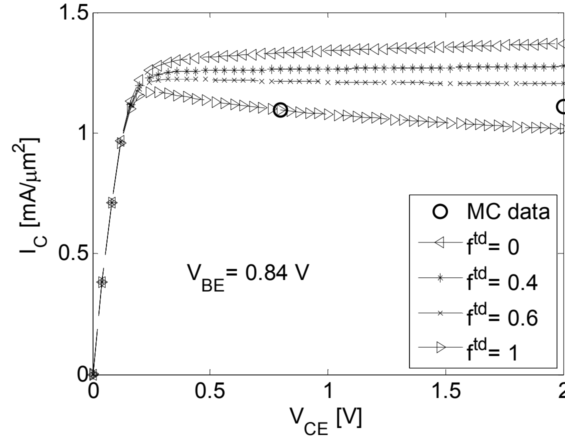
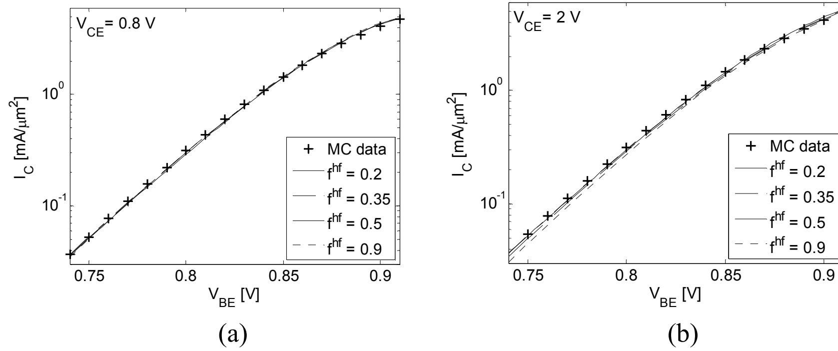


Fig. 2.9. Output characteristic for several values of f^{td} at $V_{BE} = 0.84$ V.

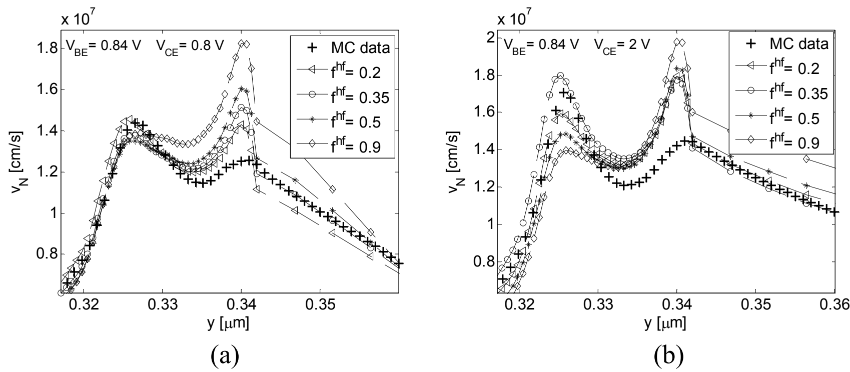
2.2.3 Parameter f^{hf}

The parameter f^{hf} affects the carrier temperature diffusive component in the energy flux density (2.3). Its value modifies the transfer characteristics at high V_{CE} voltages (see Fig. 2.10). On one hand, the reduction of f^{hf} improves the agreement between HD simulation results and MC data for the transfer characteristics depicted in Fig. 2.10 and electron velocity (see Fig. 2.11), and avoids negative output resistance (see Fig. 2.13). On the other hand, however the carrier temperature overestimation is more remarked, as highlighted in Fig. 2.12.



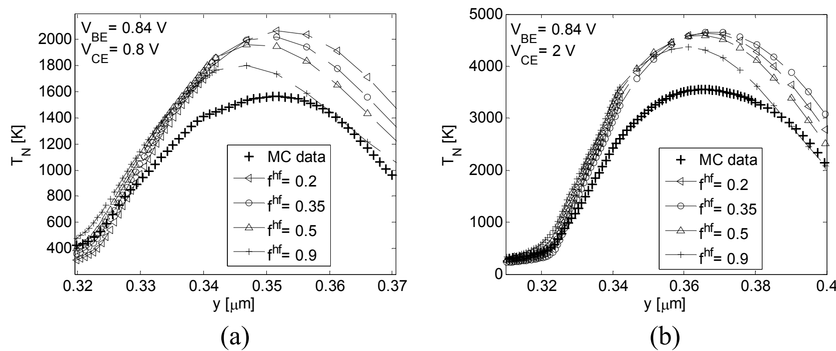
(a) (b)

Fig. 2.10. Transfer characteristics for several f^{hf} values at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b).



(a) (b)

Fig. 2.11. Electron velocity at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b) for several f^{hf} values; V_{BE} = 0.84 V.



(a) (b)

Fig. 2.12. Electron temperature at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b) for several f^{hf} values; V_{BE} = 0.84 V.

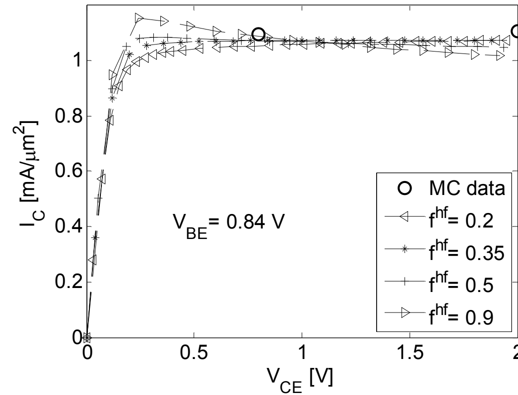


Fig. 2.13. Output characteristic for several f^{hf} values at $V_{BE} = 0.84$ V.

2.2.4 Discussion

The parameter r impacts the collector current, but it has no relevant influence on the trans-conductance and on the output conductance. The parameter f^{td} influences the collector current and output characteristics slope. f^{td} must be lower than 0.4 for the 100 GHz profile at V_{BE} = 0.84 V, in order to avoid a negative output conductance. The parameter f^{hf} influences the output characteristic slope. It has been found that the slope is positive for f^{hf} values lower than 0.36 for the 100 GHz profile at V_{BE} = 0.84 V when f^{td} and r are set to one. Since f^{td} and f^{hf} determine the value of the carrier temperature diffusive component in current density and energy flux density, respectively, the negative output conductance is due to an overestimation of carrier diffusive components in models equation.

In order to elucidate the reason for the negative output resistance, simulation results for f^{hf} 0.2 and 0.5 are reported for three different values of the collector voltage, whereas the base voltage is set to 0.84 V. Fig. 2.14 shows the temperature distribution. This figure clarifies that for high f^{hf} values the electron temperature increases with the collector voltage, while for low f^{hf} values it slightly decreases with V_{CE}. This different temperature behaviour determines different trends for mobility, which is energy dependent and decreases with increasing energy (see Fig. 2.15). Since the collector current is closely related to

the minority carrier mobility in the neutral base, for high f^{hf} values the current decreases when the collector voltage increases.

The parameter f^{hf} determines the balance between the convective and diffusive components in the energy flux density equation (2.3). Since an unphysical negative slope appears in the output characteristic when its value is high, it can be concluded that HD models overestimate the heat flux (diffusive component), which has to be reduced in order to reproduce real device behaviour. A similar discussion applies for f^{td}.

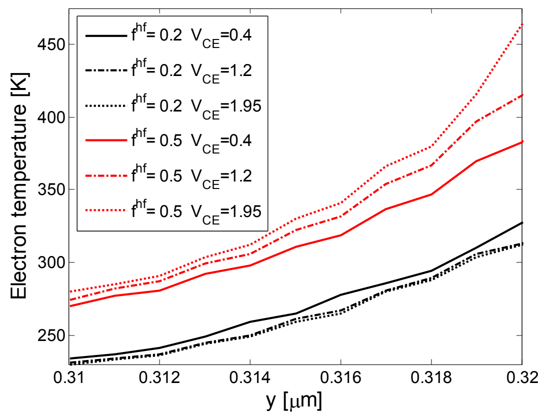


Fig. 2.14. Electron temperature distribution at V_{BE} = 0.84 V, zoom in the neutral base region of the device in Fig. 2.1.

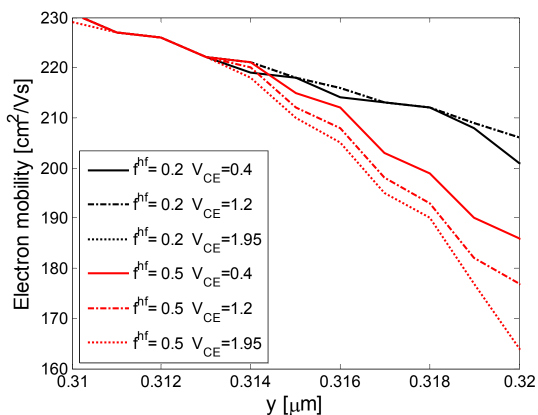


Fig. 2.15. Electron mobility distribution at V_{BE} = 0.84 V, zoom in the neutral base region of the device in Fig. 2.1.

The heat flux formulation in ET and HD models is based on the approximation of the distribution function with an heated Maxwellian [2], providing the closure relation for the highest order moment. By relaxing the heated Maxwellian assumption and reducing empirically the energy flux, a reasonable accuracy can be achieved: the heat flux reduction improves the accuracy of the electron velocity profile, reduces the spurious velocity overshoot and modifies the carrier temperature distribution, yielding a positive slope in the output characteristic. The overestimation of the diffusive flux has already been addressed as a limitation of HD model in [11]. The analysis presented in [11] is performed for a standard MOSFET and shows that the values of f^{hf} can be adjusted to achieve a better agreement of the current with MC data. The simulation results presented in [11] indicate that the spurious negative slope in the output characteristics does not occur in MOSFETs.

2.3 Optimization of standard models for a 100 GHz device

Bearing in mind the physical models' formulations and the corresponding parameters' sets, displayed in Table 2.1, an optimization can be performed for each model, in order to fit HD simulation results and MC data.

2.3.1 Four moments model

The four moments formulation [10] fixes f^{td} and f^{hf} to one, while r is a variable parameter. The parameter r changes the collector current value, but has no relevant influence on the transconductance and on the output conductance. The output resistance is negative for the whole range of r values (see Fig. 2.5). Therefore, the four-moments model appears not to be suitable for HD simulations of SiGe HBT, since it produces a negative slope which can never be removed by changing the parameter r alone.

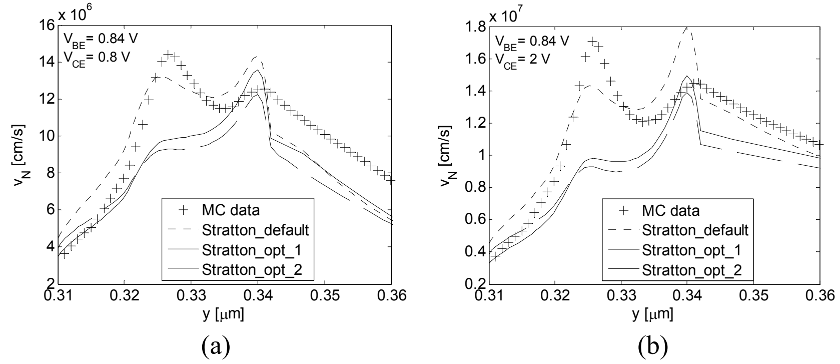
2.3.2 Stratton model

Recalling that when parameters f^{td} and r are changed the Stratton ET model is selected, we first assumed $f^{hf} = 1$. Results in Par. 2.2.2 clarify that f^{td} influences both the current values and the output characteristic slopes. The value of f^{td} must be reduced in order to avoid negative output conductance and the current overshoot in the output characteristic. Since the Stratton model involves two parameters, f^{td} and r, we can try to optimize both using MC data. This optimization has been performed and two different fitted parameter sets were obtained (see Table 2.2).

Table 2.2. HD parameters for optimized Stratton model.

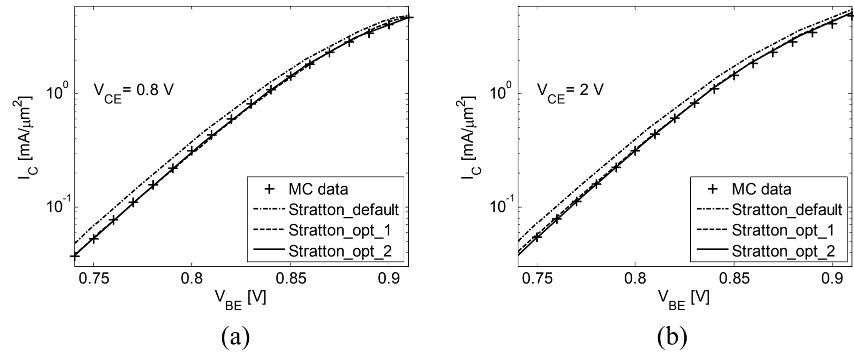
	$\mathbf{r_n = r_p}$	$\mathbf{f_n^{td} = f_p^{td}}$	$\mathbf{f_n^{hf} = f_p^{hf}}$
Stratton_default	variable range	0 ÷0.5	1
Stratton_opt_1	0.2	0	1
Stratton_opt_2	0.3	0.2	1

Using the optimized values in Table 2.2, a reasonable approximation is achieved for the collector current (see Fig. 2.17 and Fig. 2.19), but simulation results are unsatisfactory for internal quantities, such as electron velocity (Fig. 2.16), since the velocity overshoot at the base-collector junction disappears. Moreover, when f^{td} is set to zero, the thermal diffusion current vanishes, which is questionable. However, this mismatch in internal quantities does not significantly affect the high frequency behavior, since the cut-off frequency is a global quantity related to the overall velocity profile.



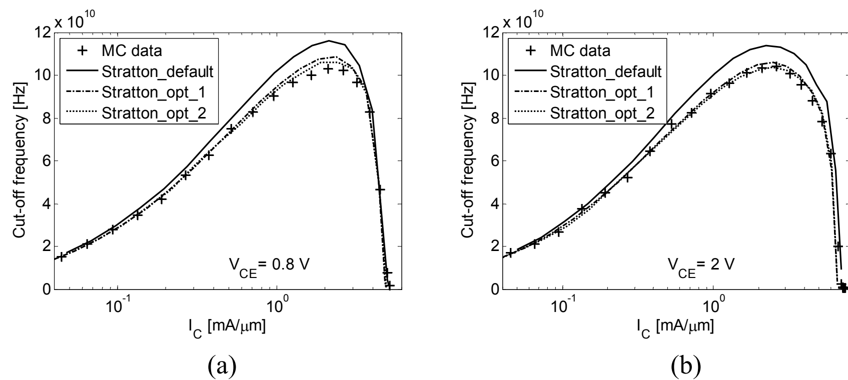
(a) (b)

Fig. 2.16. Electron velocity distribution for default and optimized Stratton model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b); $V_{BE} = 0.84$ V.



(a) (b)

Fig. 2.17. Transfer characteristics for default and optimized Stratton model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b).



(a) (b)

Fig. 2.18. Cut-off frequency for default and optimized Stratton model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b).

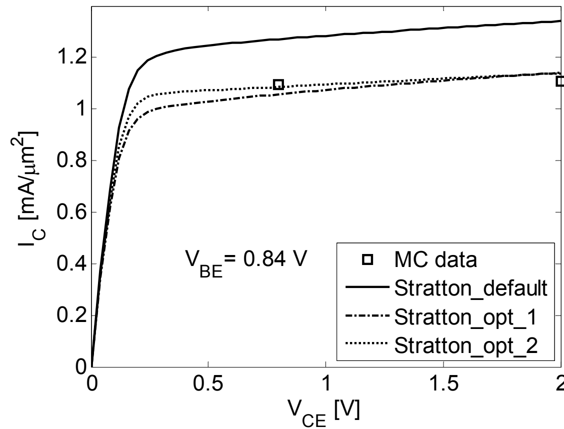


Fig. 2.19. Output characteristic for default and optimized Stratton model.

2.3.3 Blotekjær model

The Blotekjær model fixes f^{td} and r to one, while f^{hf} is a variable parameter. The parameter f^{hf} determines the output characteristic slope. It has been found that the slope is positive when f^{hf} values are sufficiently low (see Fig. 2.13). For the 100 GHz device it was found that the best f^{hf} value for fitting MC data is 0.2, see Fig. 2.20 ÷ Fig. 2.23.

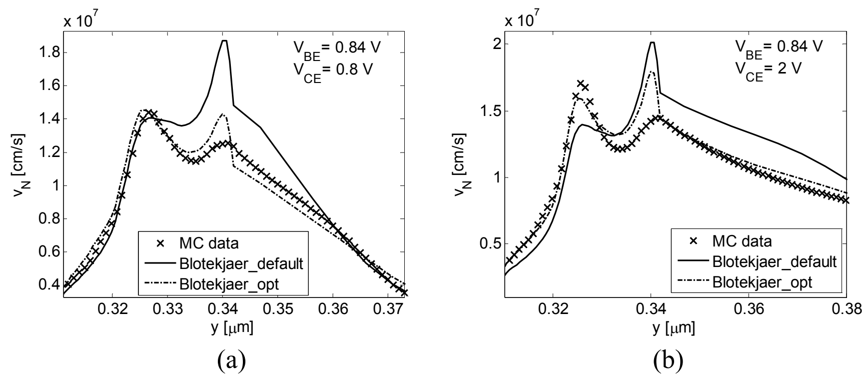
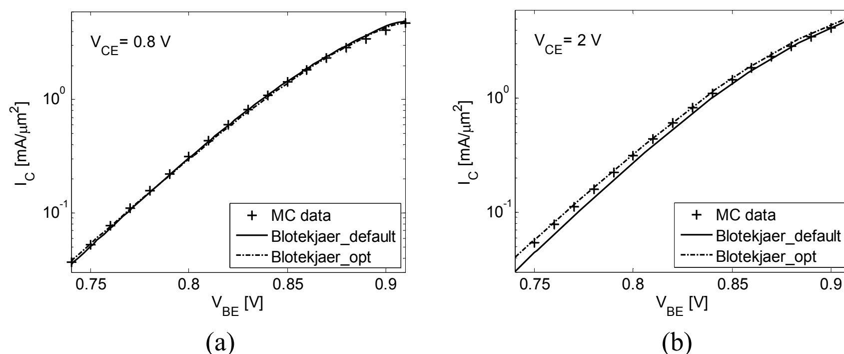
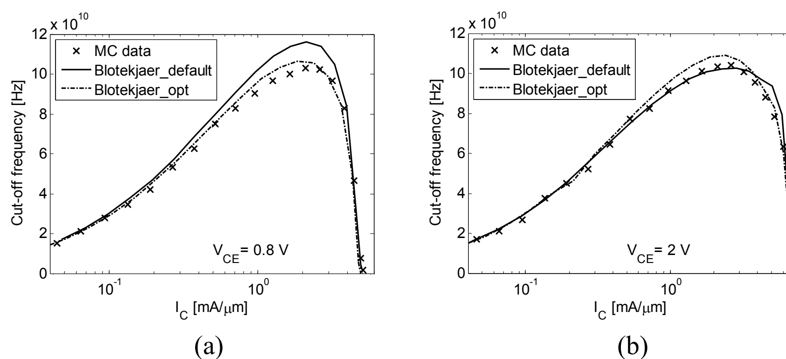


Fig. 2.20. Electron velocity distribution for default and optimized Blotekjær model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b); $V_{BE} = 0.84$ V.



(a) (b)

Fig. 2.21. Transfer characteristics for default and optimized Blotekjær model at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b).



(a) (b)

Fig. 2.22. Cut-off frequency for default and optimized Blotekjær model at V_{CE} = 0.8 V (a) and V_{CE} = 2 V (b).

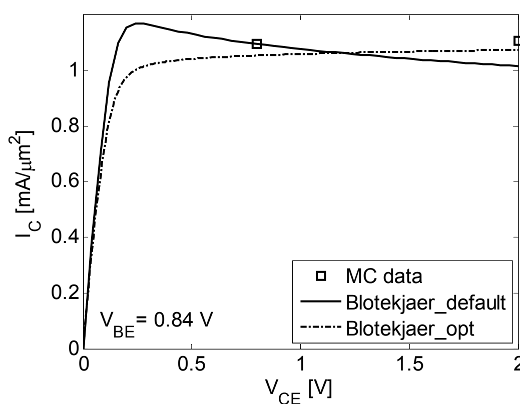


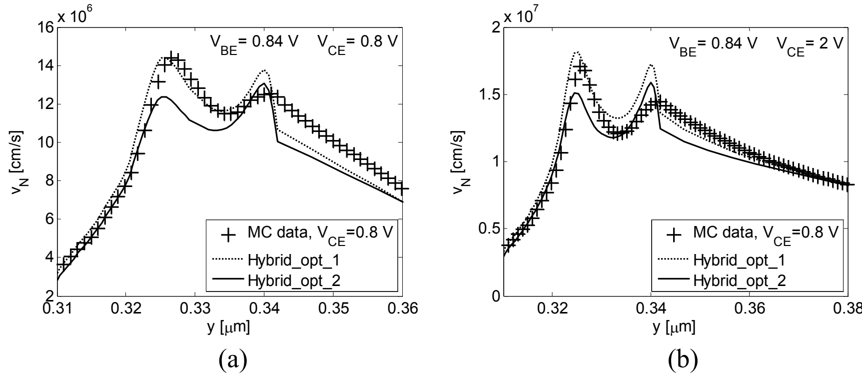
Fig. 2.23. Output characteristic for default and optimized Blotekjær model.

2.3.4 Hybrid optimization

The optimization of the HD models presented in sections 2.3.1 ÷ 2.3.3, with the exception of the model of Blotekjær, yielded unsatisfactory results. For this reason a full optimization of all available parameters was performed. Since the parameter sets found by means of a fitting procedure do not maintain a firm connection to the physical formulation, the optimum parameter sets are labeled as "hybrid". The optimization of r and f^{hf} was carried out first, providing the values reported in Table 2.2. Unfortunately, the results obtained were still unsatisfactory (see Fig. 2.24 ÷ Fig. 2.27).

Table 2.2. HD parameters for optimized partial hybrid model

	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$
Hybrid_opt_1	0.95	1	0.125
Hybrid_opt_2	0.8	1	0.07



(a) (b)

Fig. 2.24. Electron velocity distribution for optimized partial hybrid model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b); $V_{BE} = 0.84$ V.

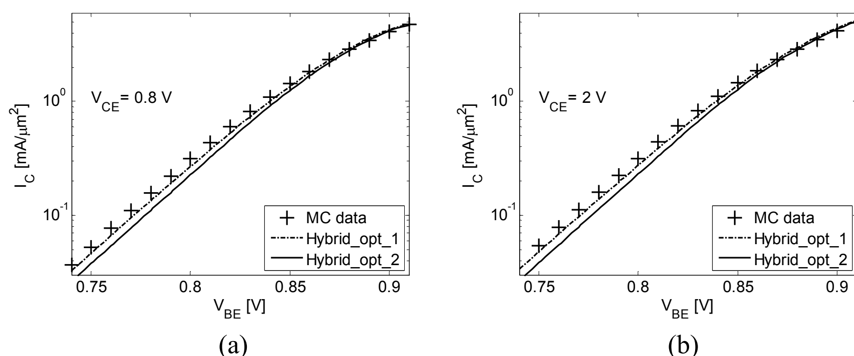


Fig. 2.25. Transfer characteristics for optimized partial hybrid model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b).

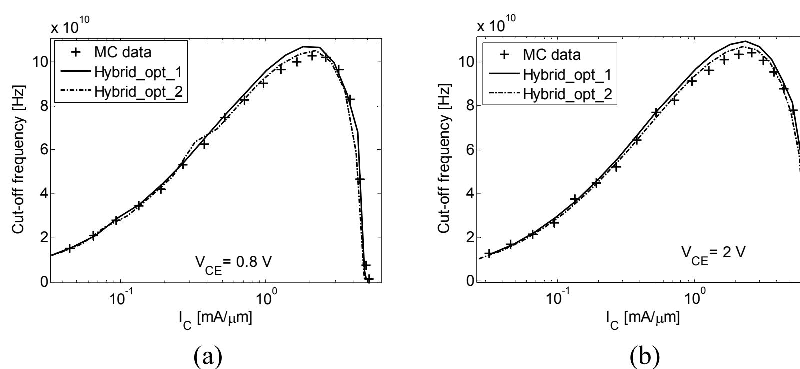


Fig. 2.26. Cut-off frequency for optimized partial hybrid model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b).

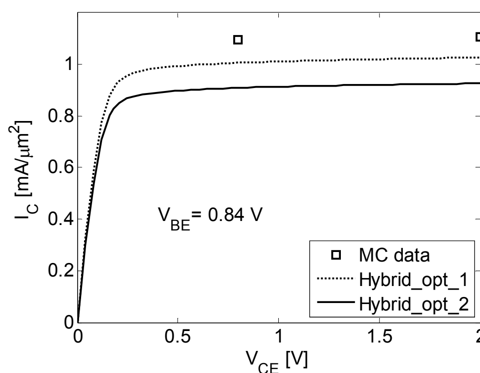


Fig. 2.27. Output characteristic for optimized partial hybrid model.

Next, the simultaneous optimization of all available parameters was carried out, which did provide reasonable results. Since the three parameters have an influence on both external simulation data (I_C, f_T) and internal variables (electron velocity, electron temperature,...), the optimization is not straightforward. The best results are displayed in Table 2.3, and corresponding plots are depicted in Fig. 2.28 ÷ Fig. 2.32, where a comparison is made with the Blotekjær optimum set already discussed.

Table 2.3. HD parameters for optimized full hybrid model.

	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$
Hybrid_opt_3	0.915	0.6	0.4
Hybrid_opt_4	0.85	0.2	0.4

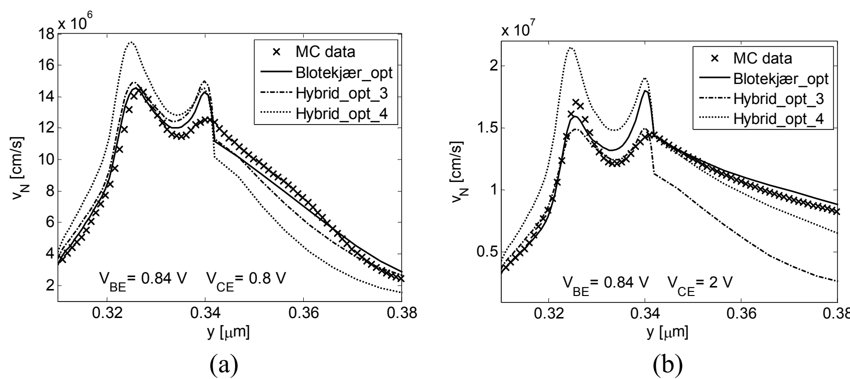


Fig. 2.28. Electron velocity distribution for optimized full hybrid model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b); $V_{BE} = 0.84$ V.

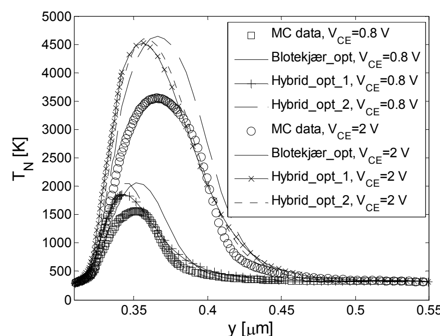


Fig. 2.29. Electron temperature distribution for optimized full hybrid model.

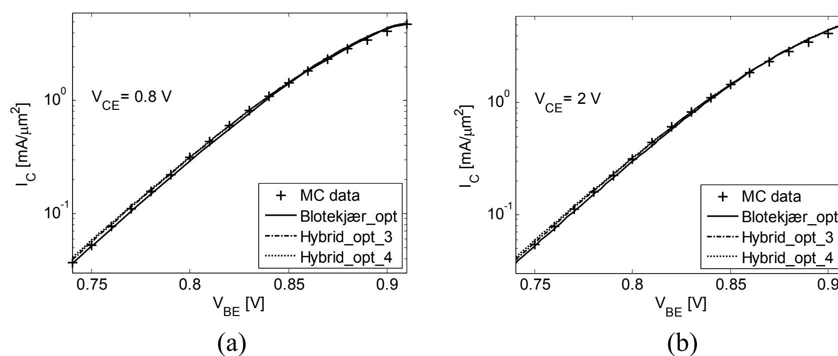


Fig. 2.30. Transfer characteristics for optimized full hybrid model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b).

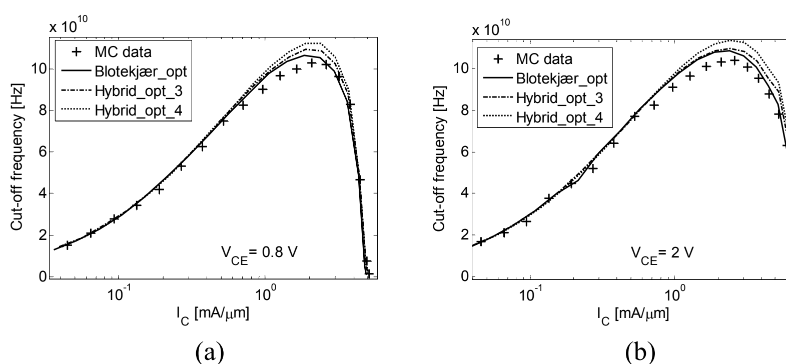


Fig. 2.31. Cut-off frequency for optimized full hybrid model at $V_{CE} = 0.8$ V (a) and $V_{CE} = 2$ V (b).

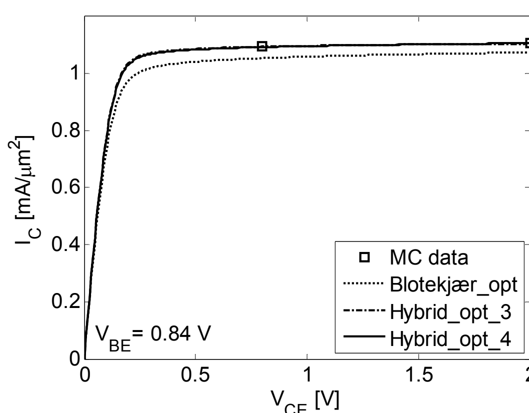


Fig. 2.32. Output characteristic for optimized full hybrid model.

A comparison between hybrid optimum values and the optimized Blotekjær model indicates that a trade-off appears in the fitting procedure. The optimized Blotekjær model gives the best fit for carrier velocity and cut-off frequency, but a slightly higher error for the collector current. On the other hand, the collector current is well described by the hybrid models which, however, lack accuracy in the carrier velocity distribution and cut-off frequency. Since the differences between are limited, both optimized parameter sets provide reasonable results and can be used to perform HD simulation matched to MC data.

2.3.5 Models' discussion and comparison

In the previous sections different simulation approaches were investigated, which refer to different non-local models and hybrid methods. A comparison is discussed here, in order to remark the differences between the models and provide a simple guideline to set the equations parameters properly, in order to avoid unphysical results and to achieve agreement with MC data.

The parameter set labelled "Stratton_1" in Table 2.2 provides the most straightforward and simplest way to avoid negative output resistance, but the carrier velocity profile is significantly different from MC data, since velocity overshoot at the B-C junction disappears. Moreover, it removes the thermal diffusion current inside the device, as f^{td} is set to zero, which is questionable. The "Stratton_2" model accounts for a small thermal diffusion current, but the poor accuracy for internal quantities is still present. The optimization of Stratton model's parameters for terminal currents requires to set r to small values, so as to reduce or delete the thermal diffusion current term in the current equation to avoid unphysical effects. For the optimized parameters this modification implies a strong underestimation of the velocity overshoot and a strong overestimation of the spurious one. Therefore, the electron velocity profile is never in accordance with MC data. This determines an overestimation of the cut-off frequency, mainly in the high-injection regime. Moreover, the r factor was introduced in the Stratton model in the definition of the

microscopic relaxation time and it has to be approximated by an average value to cover the relevant scattering processes. Therefore, this value actually depends on the doping profile and the applied field and no unique value can be given for all devices and operating conditions.

The Blotekjær approach is more suitable for optimization. It requires the optimization of one only parameter and provides a good approximation for both terminal and internal quantities, ensuring a proper description of the velocity overshoot at the base-collector junction. Since it is based on a macroscopic definition of the energy relaxation time, the same parameter value should be able to fit MC data for different doping profiles.

The full parameter optimization provide good results as well, as compared to the optimized Blotekjær model, but three parameters are involved.

From the analysis above it can be concluded that the best way to get reasonable and physical results using non-local transport models is to use the Blotekjær model, which require a single parameter optimization. The parameter must be reduced, to decrease the heat flux term in the energy flux equation. Since the energy flux in the Blotekjær model has been derived using an heated Maxwellian shape for the distribution function, the empirical reduction of the heat flux parameter helps in mitigating this assumption.

Several analysis reported for a simplified unipolar structure, showed that the spurious velocity overshoot peak can be reduced by decreasing the components of electron current density and energy flux density, since they are related to the carrier temperature gradient, [12]. Here, we have shown that these results are still valid for bipolar devices. In addition, it has been clarified that for bipolar transistors an overestimation of the thermal diffusion current and of the heat flux has a strong influence on carrier transport properties, such as electron temperature and velocity. In turns, this leads to inaccurate results and introduces an unphysical effect, i.e. a negative output resistance in the output characteristics, which is related to carrier temperature variations with collector voltage.

We also noted that the well-known cut-off frequency overestimation imputed to HD models can be drastically reduced by setting properly equation parameters; Fig. 2.22 confirms that the

Blotekjær model optimization reduces the maximum cut-off frequency error by about one order of magnitude, from 13% to 3%, when compared with the default one.

2.4 Analysis and optimization for a 450 GHz device

The optimization analysis of model parameters described in §2.2 and §2.3 has been repeated for a 1D structure with maximum cut-off frequency of about 450 GHz (Fig. 2.33). The accuracy of HD simulation for this device has been investigated by means of comparison with *SPRING* [13].

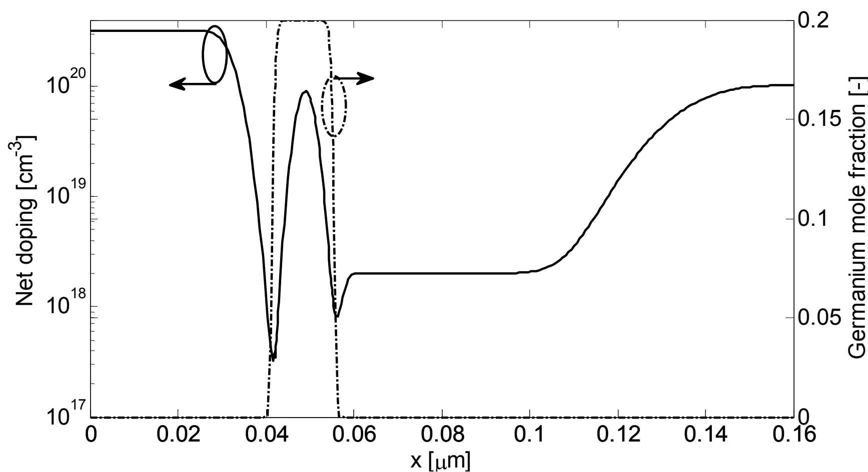
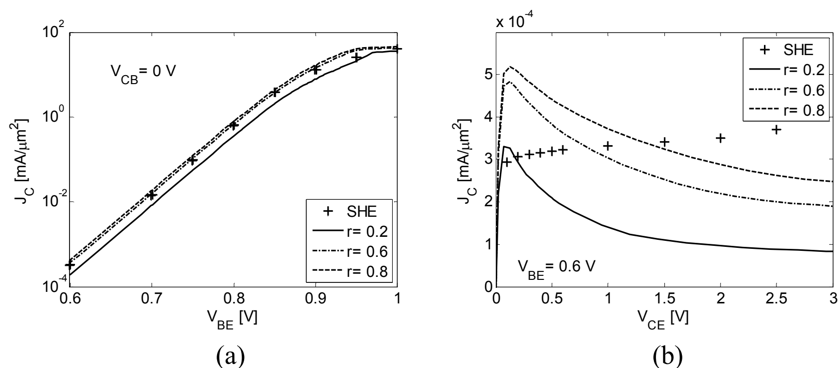


Fig. 2.33. Doping profile of the 450 GHz SiGe HBT.

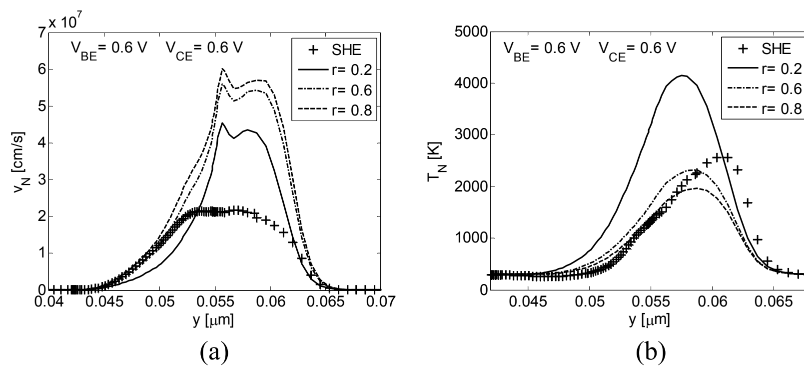
Again, to clearly show their influence, all parameters have been modified individually and simulation results have been compared to those obtained with *SPRING*.

Simulation results for parameter r are reported in Fig. 2.34 ÷ Fig. 2.36. It can be seen that a reduction of this parameter leads to lower collector current values, but the influence on the output characteristic slope, as already highlighted in 2.2.1. Withal, although electron drift velocity values change with r, a parameter reduction doesn't affects the velocity overshoots.



(a) (b)

Fig. 2.34. Transfer characteristic at V_{CE} = 0 V (a) and output characteristic at V_{BE} = 0.6 V (b) for several *r* values.



(a) (b)

Fig. 2.35. Electron velocity (a) and electron temperature (b) at V_{BE} = 0.6 V and V_{CE} = 0.6 V for several *r* values.

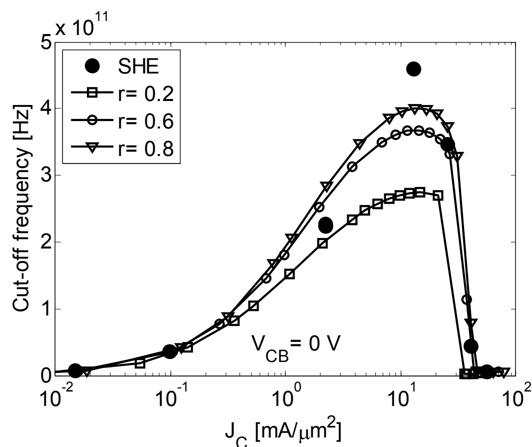


Fig. 2.36. Cut-off frequency at V_{CB} = 0 V for several *r* values.

The simulation results for parameter f^{td} are reported in Fig. 2.37 ÷ Fig. 2.39. Again, as f^{td} is reduced, the collector current slightly increases and negative output resistance disappears (see Fig. 2.37). However, a reasonable fit with *SPRING* simulations cannot be reached for carrier velocity and temperature (see Fig. 2.38 and Fig. 2.39).

The f^{hf} parameter's reduction improves the agreement between HD simulation results and reference data for transfer characteristics and removes the negative output resistance (see Fig. 2.40). However, the carrier temperature overestimation is exacerbated and the electron velocity is poorly described (see Fig. 2.41).

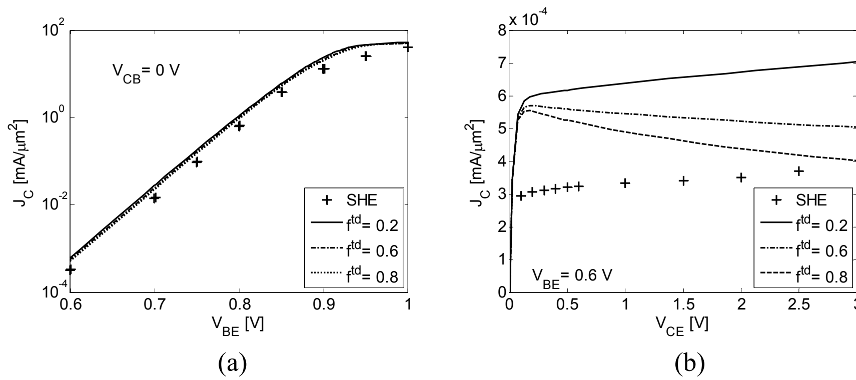


Fig. 2.37. Transfer characteristic at V_{CE} = 0 V (a) and output characteristic at V_{BE} = 0.6 V (b) for several f^{td} values.

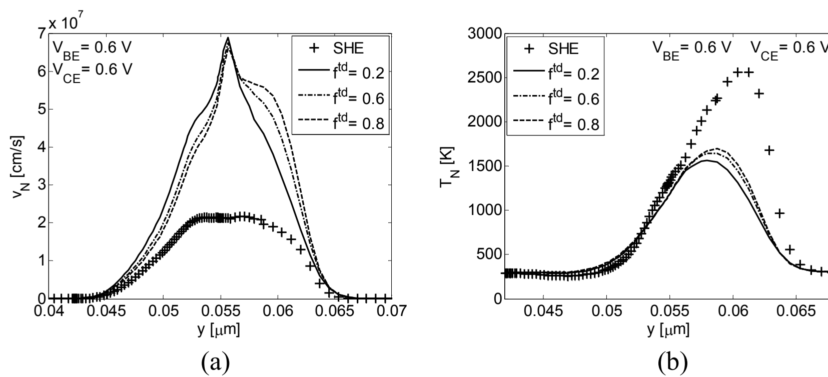


Fig. 2.38. Electron velocity (a) and electron temperature (b) at V_{BE} = 0.6 V and V_{CE} = 0.6 V for several f^{td} values.

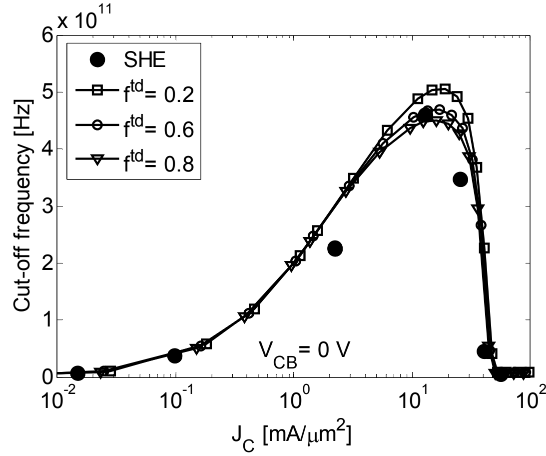


Fig. 2.39. Cut-off frequency at $V_{CB} = 0$ V for several f^{td} values.

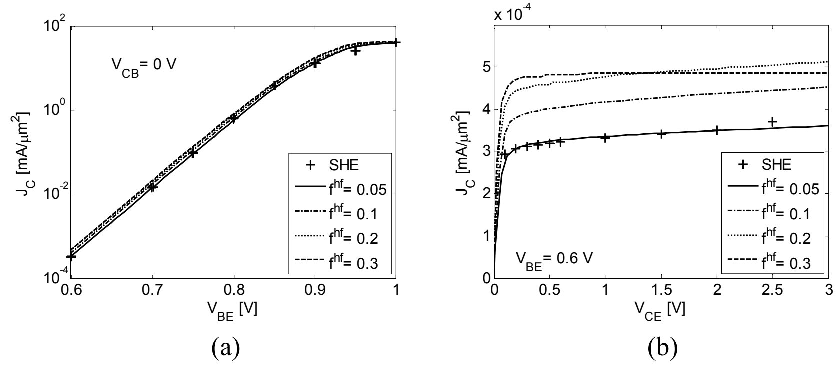


Fig. 2.40. Transfer characteristic at $V_{CE} = 0$ V (a) and output characteristic at $V_{BE} = 0.6$ V (b) for several f^{hf} values.

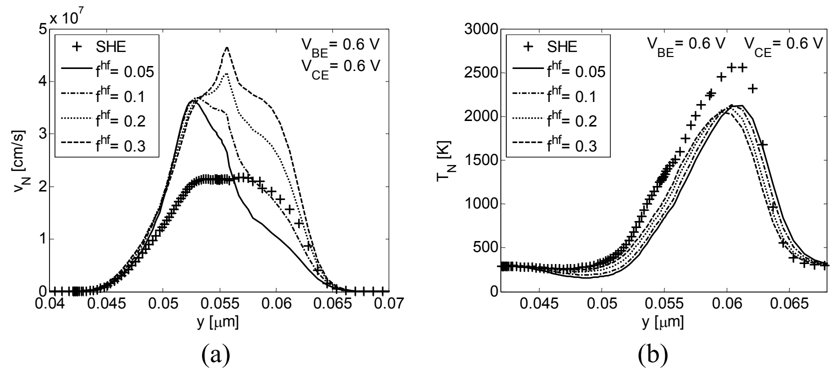


Fig. 2.41. Electron velocity (a) and electron temperature (b) at $V_{BE} = 0.6$ V and $V_{CE} = 0.6$ V for several f^{hf} values.

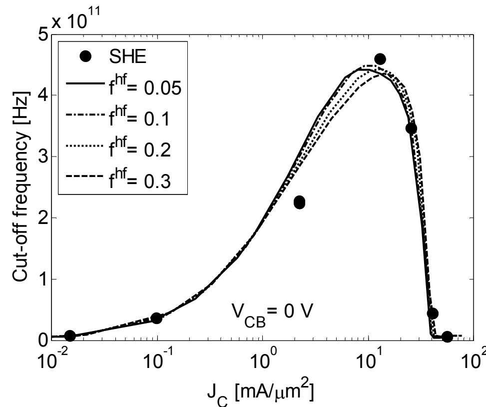


Fig. 2.42. Cut-off frequency at $V_{CB} = 0$ V for several f^{hf} values.

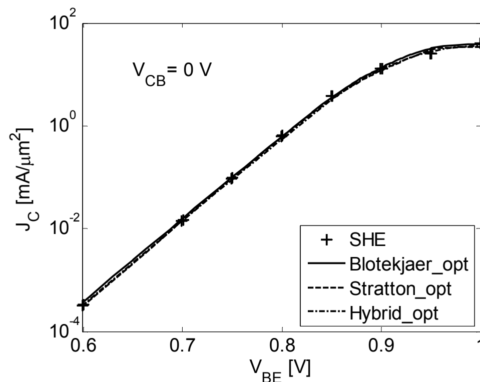


Fig. 2.43. Transfer characteristics for optimized HD models.

After the analysis of each parameter, an optimization study has been performed for Stratton, Blotekjær and hybrid approaches as described in §2.3. The results of this optimization are shown in Fig. 2.43 ÷ Fig. 2.46, and the corresponding parameters are listed in Table 2.4.

Table 2.4. HD parameters for optimized model.

	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$
Blotekjær_opt	1	1	0.07
Stratton_opt_2	0.1	0.4	1
Hybrid_opt_1	0.75	1	0.1

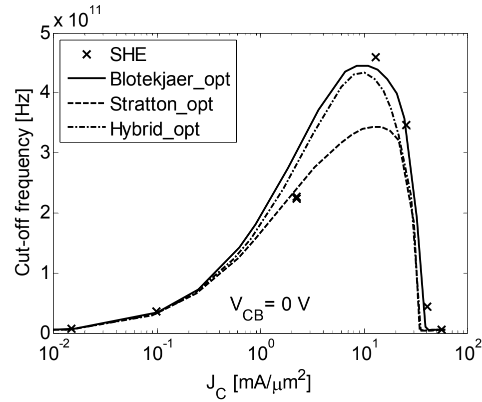


Fig. 2.44. Cut-off frequency for optimized HD models.

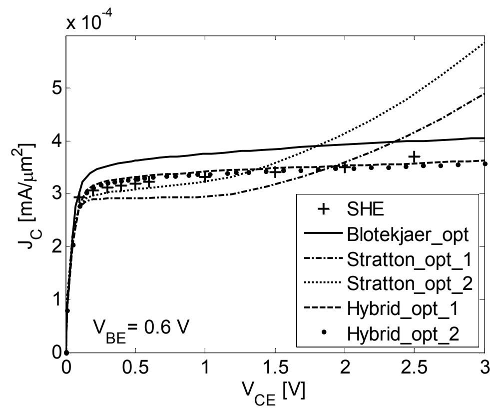


Fig. 2.45. Transfer characteristics for optimized HD models.

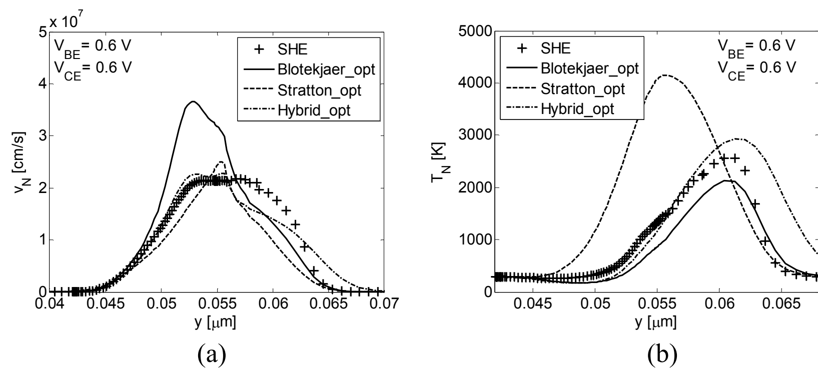


Fig. 2.46. Electron velocity (a) and electron temperature (b) for optimized HD models

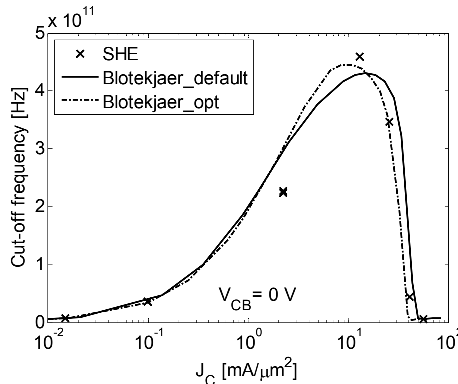


Fig. 2.47. Transfer characteristics for optimized HD models.

The results indicate that the optimization of the Blotekjær model provides the most reliable and accurate results for I_C and f_T, but a good fitting for the electron velocity and temperature is never earned. Again, we can state that the optimization of the Blotekjær model reduces the maximum cut-off frequency overestimation, from 7% to 3%, as depicted in Fig. 2.47.

2.5 Analysis and optimization for a 700 GHz device

Hydrodynamic model verification and optimization have been repeated for an extremely scaled profile with a maximum cut-off frequency of about 700 GHz (see Fig. 2.48).

The impact of each parameter has been found to be the same as described in 2.2 and 2.4. The best fitting results are obtained by the optimized Blotekjær model (see Table 2.5) providing the results shown in Fig. 2.49 and Fig. 2.50. However, for this extremely scaled structure internal distributions obtained by HD simulation are always far from MC data and SHE results even with optimized parameters. For this device the Blotekjær model optimization reduces the maximum cut-off frequency overestimation from 64% to 37% only.

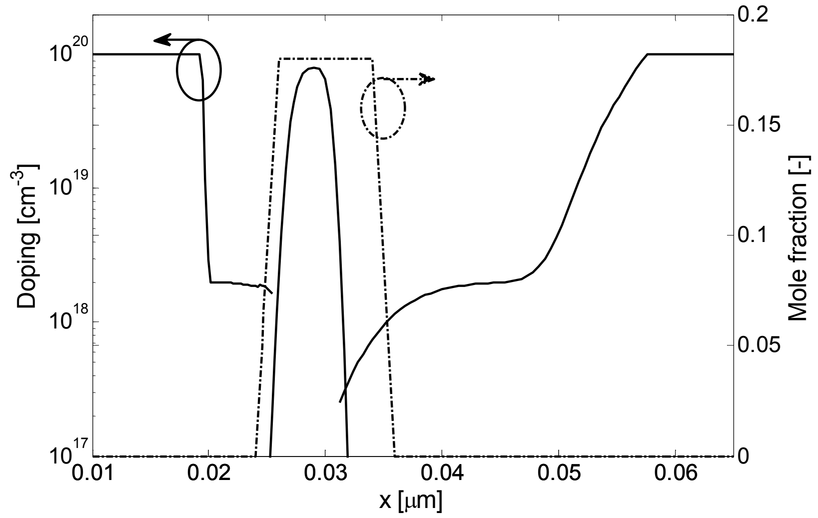


Fig. 2.48. Doping profile of the 700 GHz SiGe HBT

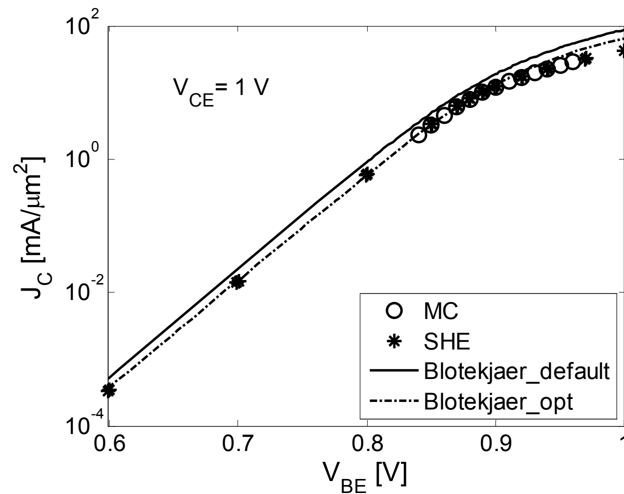


Fig. 2.49. Transfer characteristics for optimized HD models.

Table 2.5. HD parameters for optimized model.

	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$
Blotekjær_opt	1	1	0.05

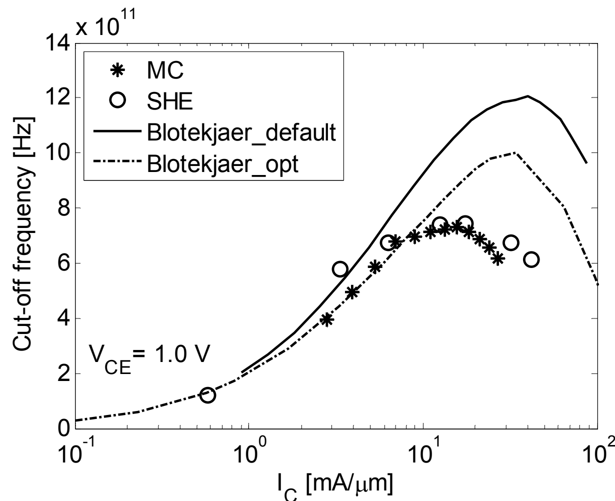


Fig. 2.50. Cut-off frequency for optimized HD models.

2.6 Unified optimization over different technological nodes

SiGe HBT HD simulation issues have been studied over different technological nodes. In particular, a detailed study of each parameter influence on simulation results and the optimization of parameters have been carried out. The analysis highlights that the influence of each parameter is similar for different technologies. For different scaled devices the unphysical effects in the output characteristics can be removed by using optimizing parameters, and the best results are achieved with the Blotekjær model. Therefore, the maximum cut-off frequency overestimation of HD models can be strongly mitigated for frequencies under the terahertz range. However, the optimization must be repeated for each different technological node. Since the overestimation of the diffusive component of the energy density increases with scaling, HD models are too inaccurate for modeling carrier transport in devices within the terahertz range, as summarized in

Table 2.6. For HBTs with a maximum cut-off frequency above 500 GHz, HD simulation becomes too inaccurate and more reliable approaches, such as Monte Carlo and Spherical Harmonic Expansion are needed.

Table 2.6. Optimum parameters and f_{TMAX} relative error of Blotekjær model with scaling.

f_{TMAX}	$r_n = r_p$	$f_n^{td} = f_p^{td}$	$f_n^{hf} = f_p^{hf}$	f_{TMAX} % error
100 GHz	1	1	0.2	3%(13%)
450 GHz	1	1	0.07	3% (7%)
700 GHz	1	1	0.05	37% (64%)

2.7 References

[1] M. A. Stettler, A. A. Muhammad, and M. S. Lundstrom, “A Critical Examination of The Assumptions Underlying Macroscopic Transport Equations for Silicon Devices,” *IEEE Transactions on Electron Devices*, vol. ED-40, pp.733-740, Apr. 1993.

[2] T. Grasser, T. W. Tang, H. Kosina, and S. Selberherr, “A Review of Hydrodynamic and Energy-Transport Models for Semiconductor Device Simulation,” *Proceedings IEEE*, vol. 91, no. 2, pp. 251-274, Feb. 2003.

[3] T. Grasser, H. Kosina, M. Gritsch, and S. Selberherr, “Using Six Moments of the Boltzmann’s Transport Equation for Device Simulation,” *Journal of Applied Physics*, vol. 90, no. 5, pp. 2389-2396, 2001.

[4] G. Sasso, G. Matz, C. Jungemann, and N. Rinaldi, “Accurate Mobility and Energy Relaxation Time Models for SiGe HBTs Numerical Simulation”, *International Conference on Simulation of Semiconductor Processes and Devices*, pp. 241-244, September 2009.

[5] Synopsys TCAD Software, Release 2007.03.
[6] Atlas Users Manual, SILVACO (2000).
[7] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers," *Phys. Rev.*, vol. 126, no. 6, pp. 2002-2014, 1962.
[8] K. Blotekjær, "Transport equations for electrons in two-valley semiconductors," *IEEE Transactions on Electron Devices*, vol. ED-17, pp.38-47, Jan. 1970.
[9] S. C. Lee and T. W. Tang, "Transport coefficients for a silicon hydrodynamic model extracted from inhomogeneous monte-carlo calculation," *Solid-State Electronics*, vol. 35, no. 4, pp.561-569, 1992.
[10] T. Grasser, H. Kosina, and S. Selberherr, "Investigation of Spurious Velocity Overshoot Using Monte Carlo Data," *Applied Physics Letters*, vol. 79, no. 12, pp. 1900-1903, 2001.
[11] I. Bork, C. Jungemann, B. Meinerzhagen, and W. L. Engl, "Influence of Heat Flux on The Accuracy of Hydrodynamic Models for Ultrashort Si Mosfets," *Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits*, Honolulu, HI, vol. 5, 1994.
[12] B. Meinerzhagen, R. Thoma, H. J. Peifer, and W. L. Egl, "On The Consistency of The Hydrodynamic and The Monte Carlo Models," *International Workshop on Computational Electronics*, Urbana-Champaign, IL, pp. 7-12, 1992.
[13] S.-M. Hong, G. Matz, and C. Jungemann, "A Deterministic Boltzmann Equation Solver Based on a Higher Order Spherical Harmonics Expansion With Full-Band Effects," *IEEE Transactions on Electron Devices*, vol. 57, n. 10, pp. 2390-2397, October 2010.

Chapter 3

Analytical models for transport parameters

During the past decades the complexity of IC fabrication processes constantly increased, whereupon the use of simulation tools for the development of new technologies has become more and more important to cope with explosive development costs and competition of today's semiconductor industry. As a result, technology-CAD (TCAD) tools that accurately predict the process and device characteristics of anticipated wafer fabrication technology is indispensable for IC fabrication technology and device development [1] and [2].

Challenges in TCAD for industry applications remain the improvement of accuracy and predictive capabilities of process and device simulators. Semiconductor device simulation concerns the description of charged carriers' distribution and transport inside the device. Carriers' distribution is described by the Poisson equation and the Boltzmann Transport Equation (BTE). The BTE can be solved numerically only at very high computational expense using Monte Carlo (MC) simulators. Therefore, the BTE is approximated, leading to the drift-diffusion equations and the hydro-dynamic equations, which account for non-local and hot-carrier effects.

Efficient device simulation demands well-calibrated models for transport parameters in order to accomplish realistic and predictive results. Several approaches can be followed to develop calibrated models: calibration can be performed using measured data or MC simulation results.

All device simulators include the most popular published models and the user can change their related parameters to match available data. Tuning the available model parameters allows getting calibrated model quite simply. However, the reliability of tuned models should be investigated carefully. Since models are empirically fitted to a

particular set of data, the model accuracy could be improved in the calibrated parameters range but it might have even worsened for different operating regions ignored in the calibration data. Besides, if model parameters have non-physical values, simulation predictiveness is missed [3]. Furthermore, tuning model parameters to fit available data could be unsuccessful even using an optimized fitting, since the data used for fitting could be inconsistent with available analytical model functions.

Most industrial device simulators allow introducing user customized models, both analytical and numerical [4]. Development of brand new analytical models (with a physical background) is the best solution to above-cited calibration related problems, although often not feasible because time-consuming. A relatively simple alternative calibration approach is to resort to look-up table based models, which rely on tables including all models dependences.

3.1 Look-up table models

A complete look-up table models set was developed in cooperation with Bundeswehr University in Munich (Germany). MC data have been generated by Bundeswehr University for holes and electrons from bulk simulations for n and p type doping [5], and for variables' range detailed in Table 3.1. Involved models with are reported in Table 3.2. Such models are the essential set to get consistent simulation results; dependences are reported too.

Table 3.1. Independent variables and ranges in look-up table models.

VARIABLE	RANGE
Doping (cm^{-3})	$10^{15} \div 10^{21}$
Lattice temperature (K)	$200 \div 500$
Germanium mole fraction	$0.0 \div 0.3$
Electric field (kV/cm)	$0 \div 350$

Table 3.2. Look-up table models and dependences for transport parameters .

PARAMETER	DEPENDENCES
Low-field mobility	N_D/N_A, T_L, x
High-field mobility (DD)	N, T_L, x, E
High-field mobility (HD)	N, T_L, x, E, T_c
Bandgap narrowing	N
Effective density of states	T_L, x
Energy relaxation time	N, T_L, x, T_c

Since a well-known trade-off exists between memory requirements and table density, a significant difficulty in look-up table models, as opposed to analytical ones, is the need to introduce and develop an interpolation technique, in order to construct new data points within the known ones. The choice of the interpolation method depends on the requirements to be met: a linear interpolation can be used and easily implemented but it doesn't assure derivatives continuity. Irregular derivatives can be allowed when they are not directly involved in the equations, e.g. mobility dependence on mole fraction. On the other hand, their continuity must be guaranteed in order to ensure simulation convergence, e.g. energy relaxation time dependence on carrier temperature. Additionally, in electro-thermal simulations all temperature derivatives should be continuous. When regular derivatives are needed, a more-sophisticated interpolation technique is essential, i.e. spline interpolation. The look-up table models where used, verified and compared to MC data, [5].

Look-up table models accuracy relies on the density of tables. Therefore memory requirements are pivotal. Besides, look-up table have no predictive capability outside the calibrated variables range and they don't provide a physical insight into device behavior. Hence, a suitable set of analytical transport parameters have been developed, with the exception of bandgap narrowing model. A look-up table formulation has been used for the bandgap narrowing, therefore it is described in details on end.

3.1.1 Bandgap narrowing model

The bandgap narrowing (BGN) model includes both heavy doping induced bandgap narrowing and germanium-induced bandgap narrowing [6]. Apparent BGN, describing the effect of heavy doping in silicon, can be determined using experimental results reported in [7] concerning the product between electron mobility and the square of the effective intrinsic density [6], as reported in (3.1):

$$\mu \cdot n_{i,e}^2 = \mu \cdot n_{i,0}^2 \cdot e^{\frac{\Delta E_g}{KT}} \tag{3.1}$$

Using experimental values in [7], the apparent bandgap narrowing can be easily determined using the mobility values. From equation (3.1) the apparent BGN can be computed as given in (3.2) if intrinsic density values are the same in device simulator [4], [7] and MC simulator [5].

$$\mu \cdot n_{i,e}^2 = \mu_{[2]} \cdot n_{i,e[2]}^2 \tag{3.2}$$

$$\Delta E_g = \Delta E_{g[2]} + KT \cdot \ln\left(\frac{\mu_{[2]}}{\mu}\right) \tag{3.3}$$

$\Delta E_{g[2]}$, $\mu_{[2]}$, $n_{i0[2]}$ denote the bandgap narrowing, mobility and intrinsic density reported in [7], while ΔE_g, μ, n_{i0} refer to the values used in this analysis. However, we should note that the intrinsic density has the same value in the Klaassen model and *Sdevice* simulator, but it has a different value in MC simulator, as shown in Table 3.3.

Table 3.3. Carrier intrinsic density value for Silicon

	n_{i0} [cm^{-3}]
Klaassen [7]	$1.194 \cdot 10^{10}$
Sdevice [4]	$1.18 \cdot 10^{10}$
MC [5]	$0.926 \cdot 10^{10}$

Hence, when the developed bandgap narrowing model has been included in MC simulator, in order to correct the difference between intrinsic densities and assuring consistency the offset

$$\Delta = -2kT \log\left(\frac{n_{i0[KLAASSEN]}}{n_{i0[MC-SIM]}}\right) \tag{3.4}$$

has been added to (3.3).

For each doping value, two p-doped bulk silicon *Sdevice* simulations (a sample doped resistor) have been performed using the *Philips unified mobility model* (which matches the experimental data in [6]) and our low-field mobility look-up table model respectively. Then, simulations results have been substituted in (3.2) to compute the apparent BGN in the base region. This calibration procedure has been repeated for n-doped bulk silicon for emitter apparent BGN computation. The resulting values of apparent bandgap narrowing are reported in Fig. 3.1. for the base and emitter regions, respectively. The doping range considered for this calibration is $10^{15} \div 10^{21}$ cm^{-3} for acceptor type doping and $10^{15} \div 5{\cdot}10^{21}$ cm^{-3} for donor type doping. The extracted BGN values, which form the BGN look-up table model, have been included directly into the simulator using the *TableBGN* model available in *Sdevice*.

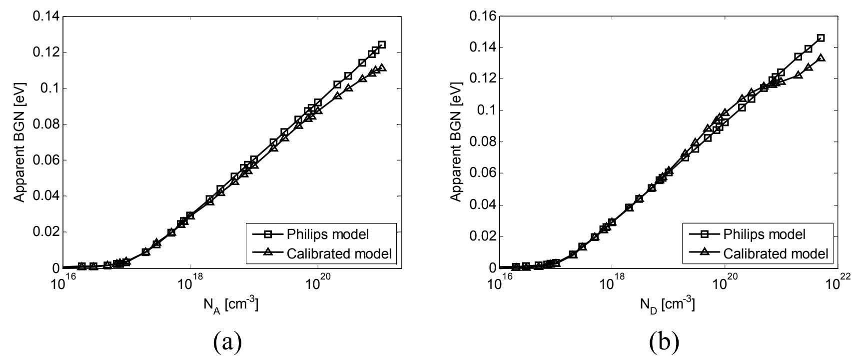


Fig. 3.1. Apparent bandgap narrowing for acceptor (a) and donor doping (b): default and calibrated values.

Additionally, since the lattice constant of silicon-germanium is larger than the lattice constant of silicon, the base region is a strained layer whose bandgap and effective densities of states (DOS) are smaller compared to silicon. Although not wholly correct, the apparent (heavy doping induced) BGN in silicon-germanium is assumed to be the same as in silicon and the effect of strain on BGN in silicon-germanium layers is computed separately and assumed to be [8]:

$$\Delta E_g \approx 0.74 \cdot x \qquad for\ x \leq 0.25 \tag{3.5}$$

Analyzing the silicon-germanium parameter file in *Sdevice*, it can be inferred that the default mole fraction dependent bandgap model is given by:

$$\Delta E_g = 0.73127 \cdot x \qquad for\ x \leq 0.245 \tag{3.6}$$

Therefore, the correct mole fraction dependent bandgap model is already implemented in *Sdevice*.

3.2 Analytical models

Commercial device simulation tools commonly used in industrial applications allow the simulation with DD and HD models of silicon-based hetero-junction devices, such as silicon-germanium HBTs. However, analytical transport models for DD/HD simulation available in the literature refer to silicon, and do not include the dependence on all relevant parameters, in particular germanium content. Therefore, based on the MC data described in §3.1, a full set of transport parameters have been generated and new analytical models have been developed which include the dependence upon all relevant quantities. In addition, new models have been compared with state-of-art models and experimental results whenever possible.

The full set of developed models is listed in Table 3.4. All models are analytical, with the exception of Bandgap narrowing parameters.

Table 3.4. Analytical models for transport parameters

PARAMETER
Effective density of states
Low-field mobility
Energy relaxation time
Saturation velocity
High-field mobility

3.2.1 Effective density of states

Silicon-germanium device simulation requires a suitable model for the effective Density of States (DOS), since it has a relevant impact on collector current computation through the effective intrinsic density, (3.7).

$$n_{ie} = \sqrt{N_C N_V} \exp(-E_G / 2k_B T_L) \tag{3.7}$$

A simplified approximation was proposed in [8], based on a constant ratio of the product of the effective DOS for silicon-germanium to the product of the effective DOS for silicon according to:

$$\frac{\left(\mathrm{N_C N_V}\right)_{\mathrm{SiGe}}}{\left(\mathrm{N_C N_V}\right)_{\mathrm{Si}}} \approx 0.4 \tag{3.8}$$

The constant ratio approximation disregards some critical dependences. In SiGe the bands degeneracies split due to strain, and the effective DOS are consequently modified. Moreover, the lattice temperature dependence should be introduced, accounting for the relative population of the upper bands which are degenerate in absence of strain. For a low germanium content the split is of the order of $k_B \cdot T_L$ [11], so that a strong dependence on lattice temperature is expected. Therefore, the dependence on the mole fraction and lattice temperature must be also taken into account in effective DOS models, as demonstrated in Fig. 3.2.

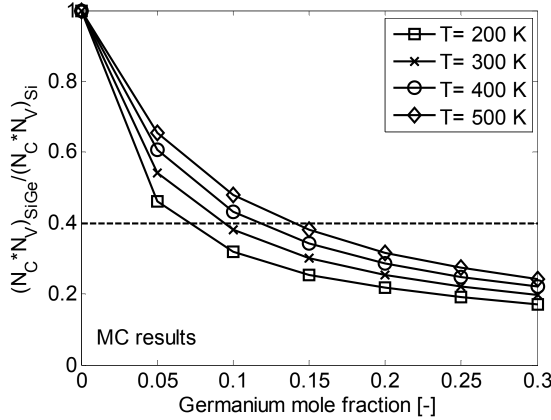


Fig. 3.2. MC data evaluation and constant ratio approximation in (3.8) as a function of germanium mole fraction for different lattice temperature values.

In device simulators effective densities are typically assumed to be a function of the effective masses. However, this analytical formulation including the dependence on the effective mass is based upon the assumption of a scalar effective mass. This simplification is not introduced in our analytical model, which is derived by fitting the data extracted from MC simulation, where the effective masses are tensors, in accordance with their definition from quantum mechanics. In order to overcome this limitation and develop a general effective DOS model, that can be easily implemented in different device simulators using their characteristic relation tying the effective DOS to the effective mass, we express the SiGe effective densities of state as functions of corresponding value in silicon. In detail, in *Sdevice* [4] the link between effective mass and effective DOS is expressed by:

$$N_{C,V} = 2.54 \cdot 10^{19} \left[\left(\frac{m_{e,h}^*}{m_0} \right) \left(\frac{T_L}{300} \right) \right]^{\frac{3}{2}} \tag{3.9}$$

or by the inverse relation:

$$\frac{m_{e,h}^*}{m_0} = \left[\frac{N_{C,V}}{2.54 \cdot 10^{19}} \right] \tag{3.10}$$

Firstly, we include the dependence on lattice temperature of effective Dos in silicon using the well-know power law dependence [12]:

$$N_C(T_L)=N_{C_300}\cdot\left(\frac{T_L}{T_0}\right)^{\alpha_C},\ T_0=300\,K \tag{3.11}$$

$$N_V(T_L)=N_{V_300}\cdot\left(\frac{T_L}{T_0}\right)^{\alpha_V},\ T_0=300\,K \tag{3.12}$$

The exponents in (3.11) and (3.12) have been optimized so as to fit MC data. The results of this optimization are depicted in Fig. 3.3 and model parameters are reported in Table 3.5 at the end of section.

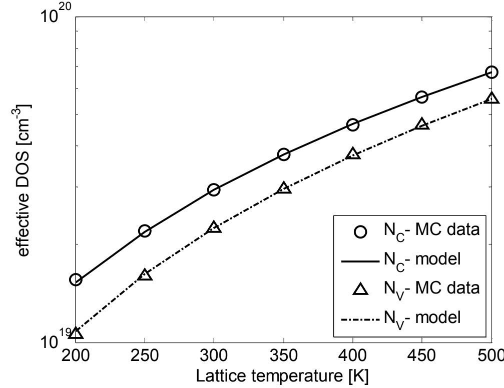


Fig. 3.3. Effective DOS for valence and conductive band in silicon as a function of lattice temperature: MC data and model results.

The model for SiGe conduction band proposed in [13] includes all main dependences. In this model, the effective DOS for electrons is calculated using the relation:

$$N_{C_Si_{1-x}Ge_x}(T_L,x)=N_{C_Si}(T_L)\cdot\frac{M_{C1}+M_{C2}\cdot\exp\left(\dfrac{\Delta E_C\cdot x}{k_B\cdot T_L}\right)}{M_{C1}+M_{C2}} \tag{3.13}$$

where all the parameters can be extracted from the schematic view of band degeneracy splitting proposed by Prinz [11]. The model in [11] provides an accurate agreement with our MC data (see Fig. 3.4).

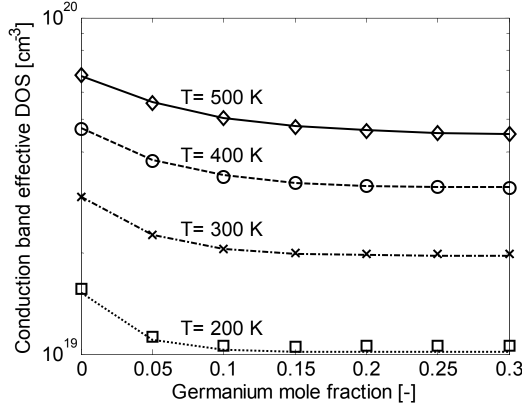


Fig. 3.4. Effective DOS for conductive band in $Si_{1-x}Ge_x$ as a function of germanium mole fraction for different lattice temperature values: MC data (symbols) and model results (lines).

On the other hand, valence band effective DOS for silicon-germanium is usually assumed to be identical to bulk silicon [13] or assumed to be linearly dependent on the Ge mole fraction [14]. We propose an analytical approximation for the hole effective DOS which includes the germanium mole fraction and lattice temperature dependences. A new model has been developed starting from the splitting schematic view in [11], which gives a value of 0.125 eV for the sum between ΔE_{V1} and ΔE_{V2} for a 20% germanium mole fraction. The new model equation reads:

$$N_{V_Si_{1-x}Ge_x}(T_L,x)=N_{V_Si}(T_L)\cdot\frac{M_{V1}+M_{V2}\cdot\exp\left(\frac{\Delta E_{V1}\cdot x}{k_B\cdot T_L}\right)+M_{V3}\cdot\exp\left(\frac{\Delta E_{V2}\cdot x}{k_B\cdot T_L}\right)}{M_{V1}+M_{V2}+M_{V3}\cdot\exp\left(\frac{\Delta E_{V3}}{k_B\cdot T_L}\right)} \quad (3.14)$$

where single terms have been arranged to fit (3.14) with MC data.

The results of the effective DOS model for holes are depicted in Fig. 3. **5**, while model parameters are summarized in Table 3.5. The new model provides an accurate approximation for valence band effective DOS in the mole fraction range of practical interest for SiGe HBTs (0 ÷ 0.2). The analytical models for the effective conduction and valence band effective density of states were successfully implemented in *Sdevice* using the physical model interface.

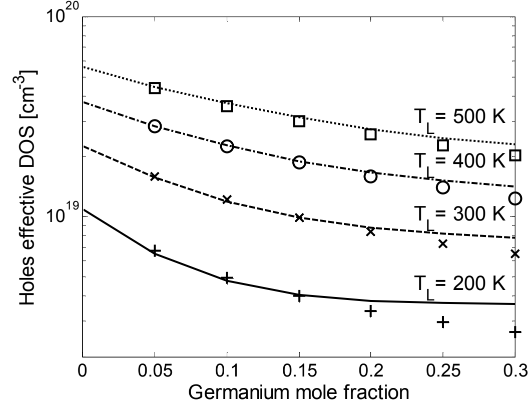


Fig. 3. 5. Effective DOS for valence band in $Si_{1-x}Ge_x$ as a function of germanium mole fraction for different lattice temperature values: MC data (symbols) and model results (lines)

Table 3.5. Parameters for SiGe alloy effective DOS.

Parameter	Value	Parameter	Value
N_{C_300} [cm^{-3}]	$2.94 \cdot 10^{19}$	N_{V_300} [cm^{-3}]	$2.24 \cdot 10^{19}$
α_C	1.62	α_V	1.79
M_{C1}	4	$M_{V1} = M_{V2} = M_{V3}$	1
M_{C2}	2	ΔE_{V1} [eV]	-0.31
ΔE_C [eV]	-0.6	ΔE_{V2} [eV]	-0.315
		ΔE_{V3} [eV]	-0.044

3.2.2 Low-field mobility

During the past decades several analytical models were introduced to model the mobility dependence on doping, lattice temperature. However, none of these models provides an adequate description of the dependence on Ge mole fraction. After a critical review of available analytical models, a novel semi-empirical analytical model for silicon bulk mobility has been developed. This model is more accurate than existing models at low temperatures and high-doping levels. Then, the dependence on Ge mole fraction has been included

for silicon-germanium alloys. A comparison with MC data and experimental results is reported below.

The first simple and analytical formulation was proposed by Chaughy and Thomas [14]

$$\mu\left(N_{D,A},T\right)=\frac{\mu_{\max}\left(T\right)-\mu_{\min}\left(N_{D,A}\right)}{1+\left(N_{D,A}/C_r\left(N_{D,A}\right)\right)^{\alpha\left(N_{D,A}\right)}}+\mu_{\min}\left(N_{D,A}\right) \quad (3.15)$$

This model predicts monotonic decreasing behavior with doping concentration, which turns out to be a poor approximation at high doping levels. Moreover no temperature dependence was included. the Chaughy-Thomas model was later modified by Arora in [15], by including the dependence upon temperature. However, in this formulation the role and meaning of the model parameters are not straightforward since the maximum mobility term disappeared. The “max-min” behaviour is not fairly accurate to model carrier scattering due to doping. Therefore several models were later developed to imporve the accuracy in the high doping range. The model [16] proposed by Masetti is based on the equation:

$$\mu\left(N_{D,A},T\right)=\frac{\mu_{\max}\left(T\right)-\mu_0\left(N_{D,A}\right)}{1+\left(N_{D,A}/C_r\left(N_{D,A}\right)\right)^{\alpha\left(N_{D,A}\right)}}+\mu_0\left(N_{D,A}\right)-\frac{\mu_1\left(N_{D,A}\right)}{1+\left(C_S\left(N_{D,A}\right)/N_{D,A}\right)^{\beta\left(N_{D,A}\right)}} \quad (3.16)$$

where the third term accounts for the decrease of mobility at doping values higher than 10^{20} cm^{-3}.

The Masetti model was later extended by researchers of Bologna University [17]. This model is given in the form:

$$\mu\left(N_D,N_A,T\right)=\frac{\mu_{\max}\left(T\right)-\mu_0\left(N_D,N_A,T\right)}{1+\left(N_D/C_{r1}\left(T\right)\right)^{\alpha_1}+\left(N_A/C_{r2}\left(T\right)\right)^{\alpha_2}}+\mu_0\left(N_D,N_A,T\right)-\frac{\mu_1\left(N_D,N_A,T\right)}{1+\left(N_D/C_{s1}\left(T\right)+N_A/C_{s2}\left(T\right)\right)^{-2}} \quad (3.17)$$

$$\mu_0\left(N_D,N_A,T\right)=\frac{\mu_{0d}N_D+\mu_{0a}N_A}{N_D+N_A} \quad (3.18)$$

$$\mu_1\left(N_D,N_A,T\right)=\frac{\mu_{1d}N_D+\mu_{1a}N_A}{N_D+N_A} \quad (3.19)$$

and includes the effects of impurity scattering, the dependence on lattice temperature and doping type (donor or acceptor). Equations (3.17) ÷ (3.19) allow to distinguish between majority and minority mobility, yet ensuring the continuity of mobility values. The doping type dependence is important as it can change mobility values up to a factor of three. It’s well know that for device simulation mobility should be expressed using a single function of donor and acceptor doping. Therefore, the doping type dependence is reported in a suitable way in equations (3.17), (3.18) and (3.19), guaranteeing with a single equation to distinguish between majority and minority mobility and with a continuous equation. We will refer to (3.17) ÷ (3.19) as the "Bologna University model".

An important and well-known model was proposed by Klaassen and workers [18], also known as Philips model. This model includes the difference between majority and minority mobility. A comparison between MC data and the Klaassen model is reported in Fig. 3.6 for both majority and minority electrons. For this analysis the Klaassen model was implemented using a numerical computing platform. The Klaassen model shows a reasonable agreement with MC data for doping levels up to 10^{19} cm^{-3} and for a lattice temperature higher than room temperature.

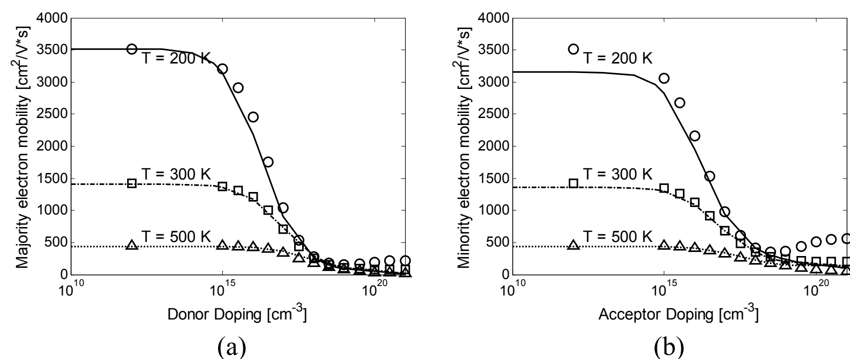
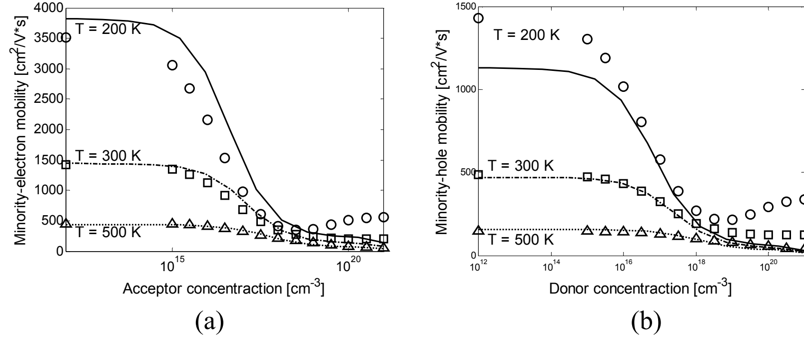


Fig. 3.6. Majority (a) and minority (b) electron mobility in pure silicon as a function of doping for different lattice temperatures; Klaassen model [18] (lines) and MC data (symbols).

In the state-of-the-art, as well as in next generation devices, doping profiles for high frequency applications exceed the value of 10^{20} cm^{-3} for the emitter region [19] and there is a growing interest in cryogenic applications. Thus this model cannot be regarded as future reference for silicon device mobility modeling. This limitation is common to all mobility models available in the Literature. In [18] it is stated that the model is suitable to describe ultra-high doping effects ($N > 10^{20}$ cm^{-3}), where the mobility decay is modelled using an effective "cluster" concentration function. Fig. 3.6 shows that this behaviour appears only from medium to high temperatures, where it is only roughly modeled. Since the mobility is seen to increase with doping at low temperatures and very high doping, the Klaassen model is unsatisfactory. The low temperature and high-doping mobility behaviour is widely documented in literature [15],[20], [21], [22], [23], but it is not properly accounted for in the existing analytical models. Additionally all available models have been calibrated at room temperatures above and for maximum doping values around 10^{20} cm^{-3}. The increase in mobility at very high-doping and low temperature appears for both minority [20], [21], [22], and majority carrier [15], although it's more marked for minority carriers. In [20] a simple explanation is suggested for these unexpected results. According to [20] very high free carrier densities screen the minority electrons from the charged impurities. However, no further physical studies have been proposed until now.

Considering the large number of parameters involved in Philips model, the optimization of parameters would be very complicated and partly incorrect, since there is no cluster concentration at low temperatures. On the contrary, the Bologna University model relies on a more simple formulation and can be extended to include high doping and low temperatures effects in a relative simple way. For this reason, it has been chosen as the starting point to develop a novel analytical calibrated model. Firstly, the original Bologna model formulation was implemented and comparison with MC data is depicted in Fig. 3.7. This model lacks accuracy for low temperatures and ultra-high doping levels; however. The poor accuracy at low temperatures was also noted by the authors.



(a) (b)

Fig. 3.7. Minority-electron (a) and minority-hole (b) mobility in pure silicon as a function of doping for different lattice temperatures; Bologna model [16] (lines) and MC data (symbols).

Based on MC simulations, an analytical low field mobility model has been developed for both minority and majority carriers. The model formulation is based on an extension to the model of Reggiani [17] for silicon, which is already available in commercial tools. The main model equation is:

$$\mu(N_D, N_A, T_L) = \frac{\mu_{\max}(T_L) - \mu_0(N_D, N_A, T_L)}{1 + \left(N_D / C_{r1}(T_L)\right)^{\alpha_1} + \left(N_A / C_{r2}(T_L)\right)^{\alpha_2}} + \mu_0(N_D, N_A, T_L) - \frac{\mu_1(N_D, N_A, T_L)}{1 + \left(N_D / C_{s1}(T_L) + N_A / C_{s2}(T_L)\right)^{-2}} \tag{3.20}$$

allows an accurate modeling of the silicon mobility behavior at low and high doping concentration, and includes the dependence upon lattice temperature T_L, and donor and acceptor doping concentration, N_D and N_A respectively. The parameter μ_{max} in (3.20) represents the lattice mobility, while the parameters μ_0 and μ_1 model carrier mobility variations at high and very high doping concentration, respectively. Lattice mobility is calculated by

$$\mu_{\max} = \mu_{\max,0} \cdot \left(\frac{T_L}{T_0}\right)^{\gamma} \tag{3.21}$$

while parameters μ_0 and μ_1 are given by

$$\mu_0\left(N_D,N_A,T_L\right)=\frac{\mu_{0d}\cdot N_D+\mu_{0a}\cdot N_A}{N_D+N_A} \tag{3.22}$$

$$\mu_1\left(N_D,N_A,T_L\right)=\frac{\mu_{1d}\cdot N_D+\mu_{1a}\cdot N_A}{N_D+N_A} \tag{3.23}$$

Equations (3.22) and (3.23) allow distinguishing between majority and minority mobility. Moreover, in order to extend the model to low temperatures (T_L < 300 K), a new formulation for $\mu_{1a,d}$ is introduced here:

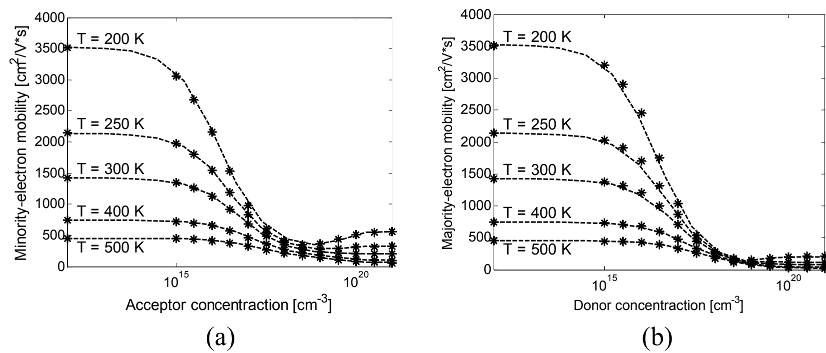
$$\mu_{1d,a}\left(N_D,N_A,T_L\right)=\frac{a\left(N_D,N_A\right)\cdot\left(T_L/T_0\right)^2+b\left(N_D,N_A\right)\cdot\left(T_L/T_0\right)+c\left(N_D,N_A\right)}{\left(T_L/T_0\right)^2} \tag{3.24}$$

The whole set of extracted parameters and equations for pure silicon is reported in Table 3.6, where $T_n = T_L/300$ is the normalized lattice temperature. The new model results and MC data are depicted in Fig. 3.8 ÷ Fig. 3.11. The new model shows a good agreement with experimental data and state-of-art models for silicon. In addition, this model offers a regularity of the results and predictability outside the calibrated range. Majority and minority mobility dependences upon lattice temperature and doping appear to be accurately modeled. Fig. 3.11 compares new model to experimental data taken from [15] and [28] at 200 K for majority electron mobility. Mobility values are reported in a logarithmic scale to emphasise the change of mobility behaviour versus doping for high-doping values.

Table 3.6. Parameters for silicon bulk mobility.

Parameters	Electrons	Holes
$\mu_{max,0}$[cm^2/V·s]	1421.6	485.51
γ	-2.24	-2.49
μ_{0d} [cm^2/V·s]	49	$123.34\cdot T_n^{-1.028}$

Parameters	Electrons	Holes
μ_{0a} [cm^2/V·s]	$205.25 \cdot T_n^{-0.934}$	$46.42 \cdot T_n^{-0.627}$
μ_{1d} [cm^2/V·s]	$\dfrac{24.812 \cdot T_n^2 + 84.6 \cdot T_n - 132.36}{T_n^2}$	$\dfrac{-68.8 \cdot T_n^2 + 300.19 \cdot T_n - 230.53}{T_n^2}$
μ_{1a} [cm^2/V·s]	$\dfrac{-94.533 \cdot T_n^2 + 518.28 \cdot T_n - 419}{T_n^2}$	$\dfrac{-5.783 \cdot T_n^2 + 81.146 \cdot T_n - 76.02}{T_n^2}$
C_{r1} [cm^{-3}]	$8.393 \cdot 10^{16} \cdot T_n^{2.951}$	$1.329 \cdot 10^{17} \cdot T_n^{3.07}$
C_{r2} [cm^{-3}]	$5.42 \cdot 10^{16} \cdot T_n^{3.045}$	$1.631 \cdot 10^{17} \cdot T_n^{3.111}$
C_{s1} [cm^{-3}]	$1.81 \cdot 10^{19}$	$5.1 \cdot 10^{19}$
C_{s2} [cm^{-3}]	$4.2 \cdot 10^{19}$	$5.8 \cdot 10^{19}$
α_1	0.68	0.7
α_2	0.7	0.77



(a) (b)

Fig. 3.8. Minority-electron (a) and majority-hole (b) mobility in pure silicon as a function of doping for different lattice temperatures; proposed model (lines) and MC data (symbols).

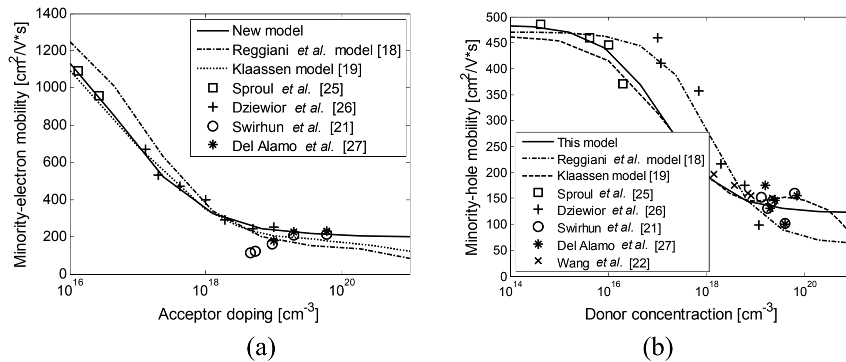


Fig. 3.9. Minority electrons (a) and holes (b) mobility in silicon as a function of doping at 300 K; symbols are experimental data.

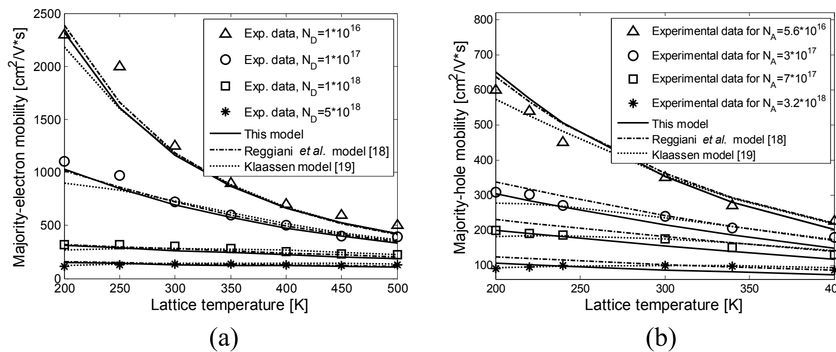


Fig. 3. 10. Majority electrons (a) and holes (b) mobility in silicon as a function of lattice temperature for several doping levels; symbols are experimental data in [28] and [29].

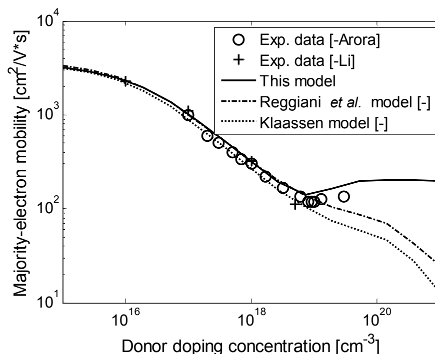


Fig. 3.11. Majority electrons mobility in silicon as a function of doping at 200 K.

Hetero-junction bipolar transistors simulations need suitable models which include the dependence on germanium mole fraction. TCAD tools generally include the mobility dependence on mole-fraction by means of a linear interpolation between silicon and germanium mobility values (see Fig. 3.12) or a piecewise polynomial approximation [4]. However, analytical models are generally preferred for computational reasons and a better physical insight. To our knowledge, the only mobility model including the germanium mole fraction dependence is [30]. In this formulation the dependence on the germanium content is modelled as

$$\frac{1}{\mu_{SiGe}} = \frac{1-x}{\mu_{Si}} + \frac{x}{\mu_{Ge}} + \frac{(1-x)x}{C_\mu} \tag{3.25}$$

where C_μ is a non-linear bowing factor. A value for the bowing factor is given for electrons (a possible value of C_μ for holes is also suggested [31], but it generates negative mobility values for some mole fractions, an shown in Fig. 3.12). Equation (3.25) has been implemented and its bowing parameter has been optimized for both electrons and holes. Even with optimized parameters, this model gives a poor approximation especially for holes (see Fig. 3.13).

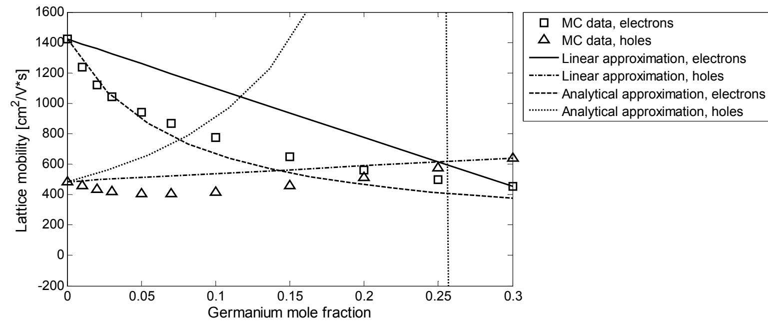


Fig. 3.12. Electrons and hole mobility in $Si_{1-x}Ge_x$ as a function of germanium mole fraction: MC data, linear approximation and model in (3.25).

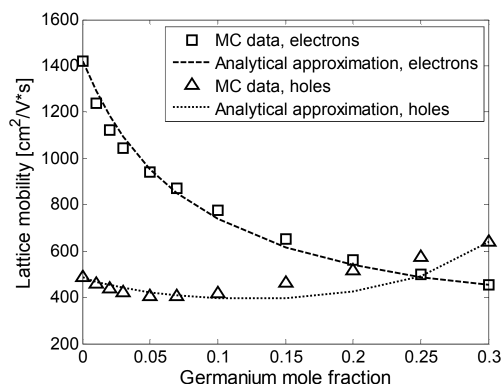


Fig. 3.13. Electrons and hole mobility in $Si_{1-x}Ge_x$ as a function of germanium mole fraction: MC data and optimized results of model in (3.25).

The dependence of electron and hole mobility on germanium content shows a minimum. This minimum is depicted in Fig. 3.12 and Fig. 3.13 for holes, since for electron mobility it appears at a higher mole fraction [32], [33]. The alloy scattering is dominant for mole fraction values lower than the minimum point. After the minimum point, lattice scattering prevails and mobility increases with increasing the mole fraction. The main modeling effort is required for hole mobility. Since in equation (3.25) the mole fraction value corresponding to the minimum mobility is fixed, the equation must be changed in order to reach a good approximation for hole mobility.

Recalling the Matthiessen's role [6], a new model for SiGe mobility has been developed and carefully calibrated for both electrons and holes. The model equation is given by

$$\frac{1}{\mu_{SiGe}} = \frac{1-x_n}{\mu_{Si}} + \frac{x_n}{\mu_{Si_{0.7}Ge_{0.3}}} + \frac{\left(1-x_n^{\alpha}\right)\cdot\left(x_n^{\alpha}\right)}{C_\mu} \quad (3.26)$$

Since in SiGe HBTs the Ge mole fraction never exceeds 0.3 (typically the maximum value is about 0.2) and limiting the mole fraction calibration range improves model accuracy, the germanium content is normalized to 0.3; $x_n = x/0.3$. To implement the model in (3.26), the lattice and bulk mobility model parameters for $Si_{0.7}Ge_{0.3}$ are needed. These parameters are reported in Table 3.7. All parameters in (3.26) have been optimized and are available in Table 3.8. Model results are showed in Fig. 3.14 and in Fig. 3.15. Minority electron mobility model results are reported for different doping values and

compared to those obtained by the Palankovski model [30] in Fig. 3.14. As can be seen, the new model shows a significant improvement in mole fraction dependence approximation. Model results for majority hole are reported in Fig. 3.15, and compared with MC and experimental data [34].

Table 3.7. Parameters for $Si_{0.7}Ge_{0.3}$ bulk mobility

Parameters	Electrons	Holes
$\mu_{max,0}$ [cm^2/V·s]	453.23	641.08
γ	-1.14	-2.118
μ_{0d} [cm^2/V·s]	$91.587\cdot T_n^{-1.0547}$	$130.24\cdot T_n^{-1.3315}$
μ_{0a} [cm^2/V·s]	$191.58\cdot T_n^{-0.92012}$	$40.848\cdot T_n^{-0.63301}$
μ_{1d} [cm^2/V·s]	$\frac{5.494\cdot T_n^2+95.873\cdot T_n-96}{T_n^2}$	$\frac{-90.178\cdot T_n^2+325.95\cdot T_n-214.83}{T_n^2}$
μ_{1a} [cm^2/V·s]	$\frac{49.072\cdot T_n^2+87.32\cdot T_n-131.97}{T_n^2}$	$\frac{3.1834\cdot T_n^2+59.611\cdot T_n-57.091}{T_n^2}$
C_{r1} [cm^{-3}]	$3.8812\cdot 10^{17}\cdot T_n^{2.0845}$	$1.1509\cdot 10^{17}\cdot T_n^{4.2018}$
C_{r2} [cm^{-3}]	$1.913\cdot 10^{17}\cdot T_n^{2.4096}$	$1.3873\cdot 10^{17}\cdot T_n^{3.2117}$
C_{s1} [cm^{-3}]	$6\cdot 10^{19}$	$2\cdot 10^{20}$
C_{s2} [cm^{-3}]	$5.4\cdot 10^{19}$	$7\cdot 10^{19}$
α_1	0.76	0.59
α_2	0.7	0.65

Table 3.8. Parameters for SiGe alloy mobility

Parameters	Electrons	Holes
α	0.487	0.548
C_μ [cm^2/V·s]	2379.4	556.4

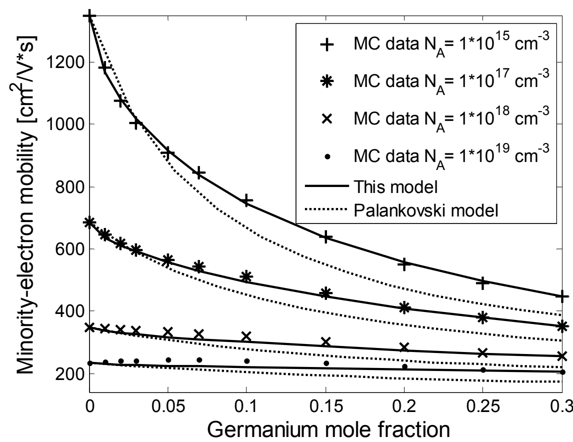


Fig. 3.14. $Si_{1-x}Ge_x$ minority electron mobility as a function of germanium mole fraction (x) for several doping concentrations, $T_L = 300$ K.

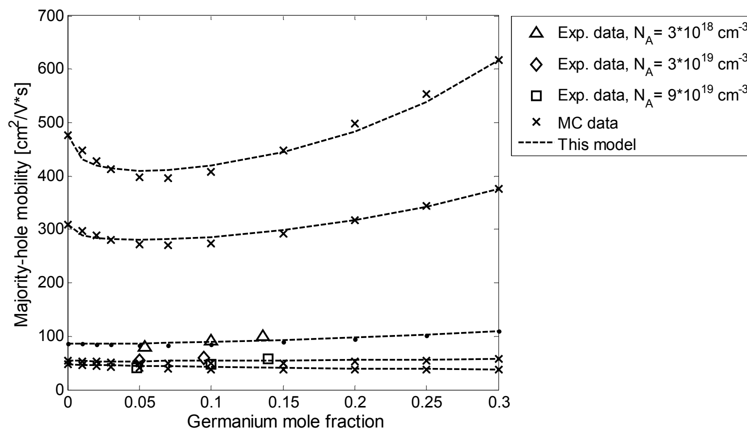


Fig. 3.15. $Si_{1-x}Ge_x$ majority hole mobility as a function of germanium mole fraction (x). Doping values for MC data and model are 10^{15}, 10^{17}, $3\cdot10^{18}$, $3\cdot10^{19}$ and 10^{21} cm^{-3} from top to bottom. Experimental data are from [33].

3.2.3 Energy relaxation time

The macroscopic energy relaxation time is a critical parameter in the HD formulation, appearing in the collision terms of energy balance equation and in high field hydrodynamic mobility model. An accurate verification by MC simulations is needed to achieve consistency of the HD model and the Boltzmann transport equation. A relaxation time model extended to include the mole fraction dependence is given in [30]. This model however, disregards the dependence upon lattice and carrier temperatures. The model proposed in [34] includes the lattice and carrier temperature in silicon and germanium, but it is not extended to silicon-germanium. In addition, relaxation time values reported for silicon do not agree with typical values reported in literature, [35]. In [36] a quadratic dependence on the electron temperature was included for silicon. However, a constant or quadratic relation for the energy relaxation time are to be considered as rough approximations. For accurate and predictive simulations, more refined analytical formulations should be used [37].

Using MC data, a new analytical model for the electron energy relaxation time was developed for silicon and silicon-germanium. This model incorporates the relaxation time variation with electron and lattice temperature, as well as germanium mole fraction. The energy relaxation time for holes can be set to a constant value, since it has no influence on hydrodynamic simulation results. Moreover, the dependence upon doping was found to be negligible by MC results. The model equation for energy relaxation times in bulk material is given by the Gaussian function:

$$\tau_W = \tau_{W,0} + \tau_{W,1} \cdot \exp\left(C_1 \cdot \left(\frac{T_n}{T_0} \right)^2 + C_2 \cdot \left(\frac{T_n}{T_0} \right) + C_3 \cdot \left(\frac{T_L}{T_0} \right) \right) \quad (3.27)$$

where the material composition dependence is included through the following parameters

$$\tau_{W,0} = \tau_{W,0_Si} \cdot (1 - x_n) + \tau_{W,0_Si_{0.7}Ge_{0.3}} \cdot x_n + C_\tau \cdot (1 - x_n) \cdot x_n \quad (3.28)$$

$$C_1 = C_{1_Si} \cdot (1 - x_n) + C_{1_Si_{0.7}Ge_{0.3}} \cdot x_n + C_C \cdot (1 - x_n) \cdot x_n \quad (3.29)$$

where $x_n = x/0.3$ is the normalized mole fraction, C_τ and C_C are referred to as bowing factors and T_0 is the reference temperature of 300 K. The maximum mole fraction considered is still 0.3. All parameters are reported in Table 3.9.

Table 3.9. Parameters for SiGe alloy energy relaxation time.

Parameters	Electrons
$\tau_{w,0_Si}$ [ps]	0.391
$\tau_{w,0_Si_{0.7}Ge_{0.3}}$ [ps]	0.449
$\tau_{w,1}$ [ps]	-0.14434
C_{1_Si}	0.00135
$C_{1_Si_{0.7}Ge_{0.3}}$	0.0028
C_c	-0.00181
C_2	-0.059
C_3	0.0107
C_τ [ps]	-0.05

A comparison between model results and MC data is depicted in Fig. 3.16 and Fig. 3. 17. Note that, although at low carrier temperature the dependence on carrier temperature changes (Fig. 3.16), this behavior needs not to be modeled since when the electron temperature is close to the lattice temperature the term $(T_n-T_L)/\tau_w$ appearing in the energy balance equation is negligible. Model shows a good agreement with MC data for the whole range of temperatures in Fig. 3.16 and of mole fraction values in Fig. 3. 17.

3.2.4 Saturation velocity

The saturation of velocity generally appears in high-field mobility models. Therefore an accurate model embodying all relevant dependences is mandatory. In [39] Ershov *et al.* propose a simple model including the dependence on germanium mole fraction, and the model in [40] adds the lattice temperature influence. However, all available

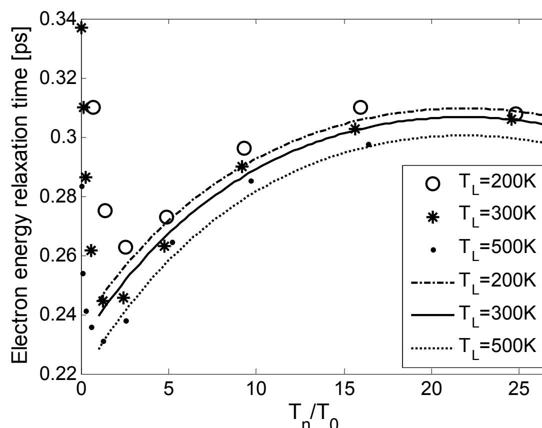


Fig. 3.16. Electron energy relaxation time in silicon as a function of electron temperature for several lattice temperatures, $T_0 = 300$ K. Lines: model; symbols: MC data.

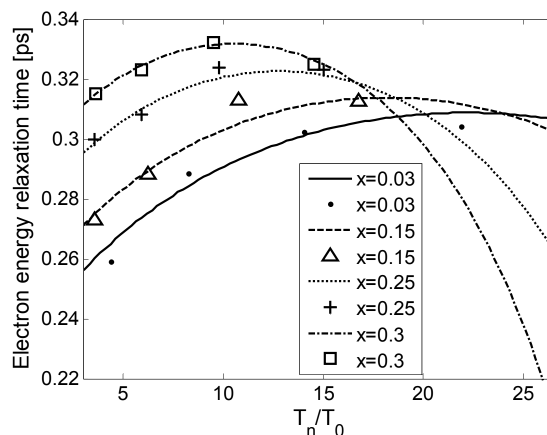


Fig. 3. 17. $Si_{1-x}Ge_x$ electron energy relaxation time as a function of electron temperature for several germanium contents, $T_L = T_0 = 300$ K. Lines: model; symbols: MC data.

models neglect the dependence upon the doping level. By MC investigation we verified that saturation velocity can vary with doping level more than 10% (see Fig. 3.18). Thus, a new model has been developed which includes the dependence upon lattice temperature, germanium mole fraction and doping value. Moreover, it has been

verified that saturation velocity values are not influenced by the doping type. The new model for bulk silicon is given by

$$v(N,T_L)=\frac{v_{300}(N)}{1\text{-}A\cdot\left(1\text{-}(T_L/T_0)^b\right)} \tag{3.30}$$

$$v_{300}(N)=v_1\cdot(N/N_{REF})^g \tag{3.31}$$

where N_{REF} is fixed to 10^{16} cm^{-3} and T_0 represents the room temperature (300 K). Model extension to include the dependence upon the germanium mole fraction is given by

$$v_{SiGe}=v_{Si}\cdot(1\text{-}x_n)+v_{Si_{0.7}Ge_{0.3}}\cdot x_n+\left(1\text{-}x_n^{\alpha}\right)\cdot x_n^{\alpha}\cdot C_v \tag{3.32}$$

where C_v is again a bowing factor. The germanium content x_n is still normalized to 0.3.

Model parameters are summarized in Table 3.10 for both electrons and holes. Model results and a comparison with MC data are depicted in Fig. 3.18 and Fig. 3.19.

Table 3.10. Saturation velocity model parameters.

Parameter	Electrons	Holes
	Si (x_n = 0)	
v_1 [cm/s]	$9.954\cdot10^6$	$1.1412\cdot10^7$
g	0.0127	0.0084
A	0.3163	0.218
b	0.8091	1.0693
	$Si_{0.7}Ge_{0.3}$ (x_n = 1)	
v_1 [cm/s]	$6.093\cdot10^6$	$8.1459\cdot10^6$
g	0.004	0.0062
A	0.175	0.1284
b	1.1048	1.24
C_v[cm/s]	$-5.262\cdot10^6$	$-2.22\cdot10^6$
α	0.74665	0.8223

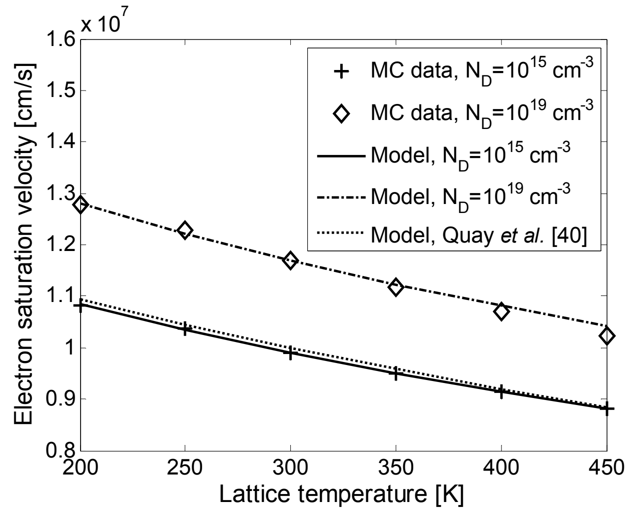


Fig. 3.18. Electron saturation velocity in silicon as a function of lattice temperature for several doping concentrations.

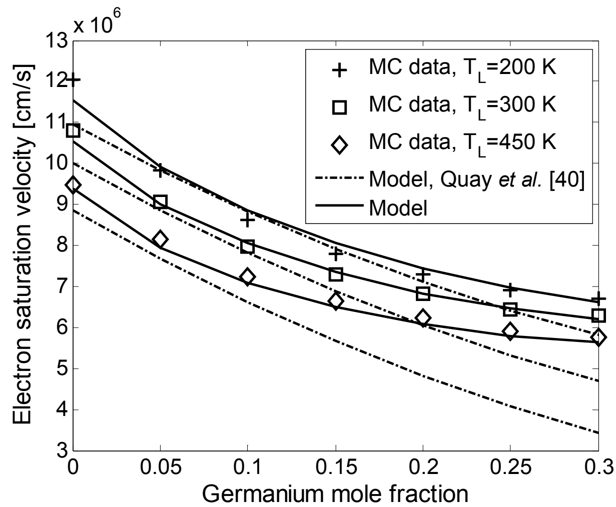


Fig. 3.19. Electron saturation velocity in $Si_{1-x}Ge_x$ as a function of germanium mole fraction for several lattice temperatures. $N = N_D = 10^{18}$ cm^{-3}.

3.2.5 High-field mobility

The well known Canali *et al.* model [41] has been widely used for modeling high field mobility in drift-diffusion simulations. It relates

high-field mobility (μ_{HIGH}) to low-field mobility (μ_{LOW}), electric field (E) and carrier saturation velocity (v_{SAT}) using the simple relation

$$\mu = \frac{\mu_{LOW}}{\left[1+\left(\frac{E\cdot\mu_{LOW}}{v_{sat}}\right)^{\beta}\right]^{\frac{1}{\beta}}} \tag{3.33}$$

where β is a fitting parameter.
The Canali model has been extended later [3] with an additional parameter α

$$\mu = \frac{(\alpha+1)\,\mu_{LOW}}{\alpha+\left[1+\left(\frac{(\alpha+1)\,E\cdot\mu_{LOW}}{v_{sat}}\right)^{\beta}\right]^{\frac{1}{\beta}}} \tag{3.34}$$

However, for modern heterojunction transistors the Canali model becomes inaccurate. This is due to the fact that it is based on the assumption of local equilibrium between electrons and the lattice, which yields a saturation behavior for carrier velocity. With the ongoing downscaling of devices, the use of ET and HD models has become mandatory, in order to capture nonlocal effects. It is well known that in the presence of a strong electric field electrons gain energy and their temperature is highly increased [42]. The carrier temperature gradient introduces an additional driving force and HD simulations replace the electric field with an effective driving force [4], [43]. The effective driving force is derivable from the homogeneous steady state energy balance equation and is given by

$$\mathrm{F}=\sqrt{\frac{\max(\mathrm{W_C}\text{-}\mathrm{W_0},0)}{\left(\tau_E\cdot q\cdot\mu_{HIGH}\right)}} \tag{3.35}$$

where τ_E is the carrier energy relaxation time, W_0 and W_C are the lattice and carriers (electrons/holes) energies, respectively, and are expressed as

$$\mathrm{W_0}=\frac{3}{2}\cdot k_B\cdot T_L,\qquad \mathrm{W_C}=\frac{3}{2}\cdot k_B\cdot T_C \tag{3.36}$$

Several high-field mobility models for HD simulations include a direct dependence upon the carrier temperature [31], [44]. However,

these models are derived by replacing the high-field mobility μ_{HIGH} in (3.35) with the low-field value μ_{LOW}. As a consequence, consistency between the electric field and the effective driving force in the homogenous case is lost. This has been verified by applying (3.35) to MC data (see Fig. 3.20). In this calculation the low field mobility and energy relaxation time are computed using the models described in §3.2.2 and §3.2.3.

A more general approach is available in commercial device simulators [3], which combines the effective driving force (3.35) with the model by Canali *et al.* [46]:

$$\mu_{HIGH} = \frac{\mu_{LOW}}{\left[1 + \left(\frac{F \cdot \mu_{LOW}}{v_{sat}}\right)^{\beta}\right]^{1/\beta}} \tag{3.37}$$

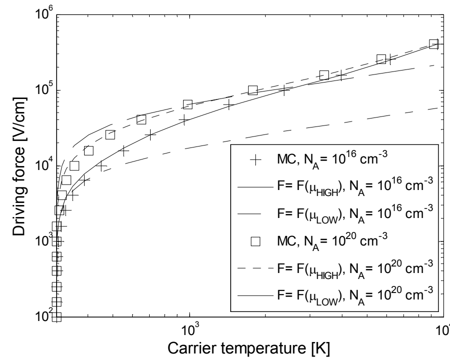


Fig. 3.20. Electric field in homogeneous silicon (MC data) and effective driving force computed by using μ_{HIGH} and μ_{LOW} MC data in (3.35).

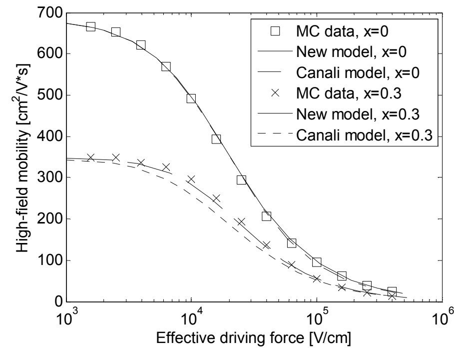


Fig. 3.21. High field mobility as a function of effective driving force as given in (3.35) at $T_L = 300$ K and $N = N_A = 10^{18}$ cm^{-3}.

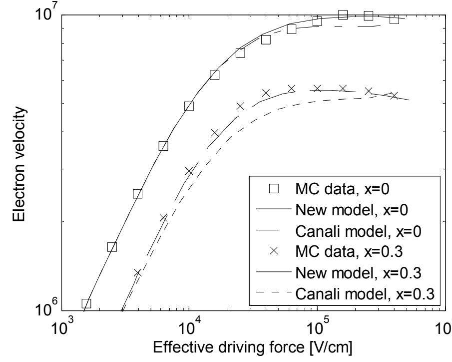


Fig. 3.22. Electron velocity as function of effective driving force computed by using (3.35) at T_L = 300 K and $N = N_A = 10^{18}$ cm^{-3}. Velocity is calculated as the product between mobility and driving force.

However, this formulation has been verified to be not suitable for SiGe, since it doesn't provide an accurate dependence on all relevant variables (see Fig. 3.21); moreover, it does not account for the negative differential electron drift velocity (Gunn effect) as shown in Fig. 3.22.

One can try to optimize the parameter β in equation (3.37) by introducing some non-constant formulation for it. However, with tis simple fix a reasonable approximation cannot be reached because the analytical formulation can never describe the carrier velocity decay at high driving forces. The extended Canali model is based on the saturation assumption for carrier velocity. Therefore, the resulting carrier velocity, given by the product between carrier mobility and effective driving force, cannot describe the actual behavior. Although the Gunn effect is well known for III-V semiconductors, it has been verified that the transition of electrons from the twofold valley to the fourfold valley can occur in SiGe as well [47],[48]. However this behavior has been never included in any analytical mode. Thus, we propose a new analytical model where the high-field mobility is calculated from the equations:

$$\mu_{HIGH} = \frac{\mu_{LOW}}{\left\{\left[1+\left(\frac{F}{F_R}\right)^{\alpha}\right]\cdot\left[1+\left(\frac{F\cdot\mu_{LOW}}{v_{sat}}\right)^{\beta}\right]^{\frac{1}{\beta}}\right\}} \qquad (3.38)$$

$$\beta=\beta_0+\beta_1\cdot\log_{10}\left(1+x_n\right)+\left(\beta_2\cdot\log_{10}\left(\frac{N}{N_{REF}}\right)\right)^{\beta_3} \tag{3.39}$$

$$\alpha=\alpha_0-\alpha_1\cdot\log_{10}\left(1+x_n\right) \tag{3.40}$$

$$F_R=\frac{T_0}{T_L}\cdot\exp\left(F_{R0}-F_{R1}\cdot\log_{10}\left(N/N_{REF}\right)\right) \tag{3.41}$$

where (3.35) is used to calculate the driving force. This model provides an accurate description of mobility as a function of lattice temperature, carrier temperature, doping and mole fraction. Model parameter are listed in Table 3.11 (N_{REF} in (3.39) and (3.41) is fixed to 10^{15} cm^{-3}).

Table 3.11. High-field mobility model parameters.

Parameter	Electrons	Holes
β_0	1.26	1.1
β_1	1	0
β_2	0.07	0.09
β_3	1	2
α_0	1	-
α_1	1.4	-
F_{R0}	-16.8112	-
F_{R1}	0.3979	-

The development of this new model required a significant effort, since it was not simple to identify an analytical formulation able to reproduce correctly high field mobility variations with doping, mole fraction and carrier temperature, as these dependences are also included in different quantities (low-field mobility, driving force and saturation velocity). The new formulation consists of a nonlinear equations system ((3.35) and (3.38) ÷ (3.41)), whose unknown quantities F and μ_{HIGH} can be obtained from the numerical solution of the nonlinear equations. Using this approach we can calculate the effective driving force as depicted in Fig. 3.23, which is in good agreement with MC data. The new model includes the dependences on carrier temperature, lattice temperature and mole fraction (see Fig. 3.24 and Fig. 3.25). The accuracy is strongly improved compared to the extended Canali model (see Fig. 3.21 and Fig. 3.22). Additionally, it predicts the negative differential slope of electron velocity versus electric field (Fig. 3.22). Since holes have no "hot effects", the model (3.37) combined with (3.39) can still be used for holes and the electric field is the driving force. A good accordance with MC data was found by introducing an adequate dependence on doping value for the β parameter, (3.39).

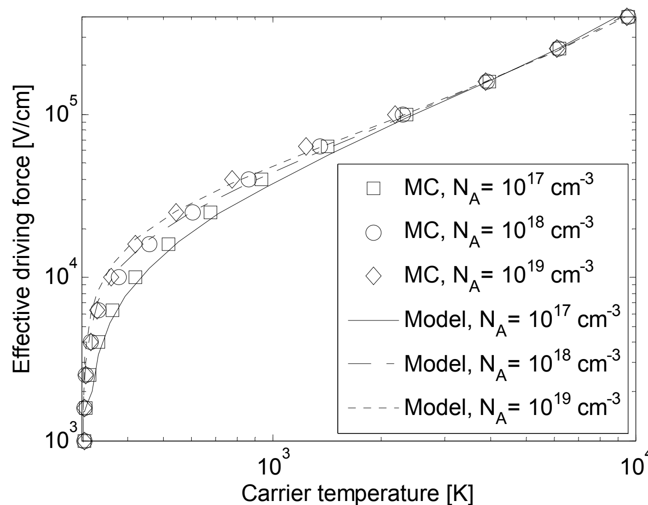


Fig. 3.23. Effective driving force in silicon for minority electrons as a function of electron temperature for several doping concentrations at $T_L = 300$ K.

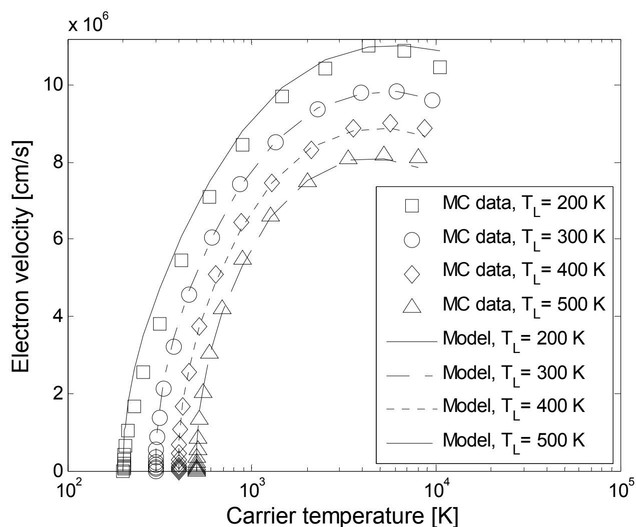


Fig. 3.24. Electron drift velocity in silicon ($\mu_{HIGH} \cdot F$) as a function of electron temperature for several lattice temperatures and $N = N_A = 10^{18}$ cm^{-3}.

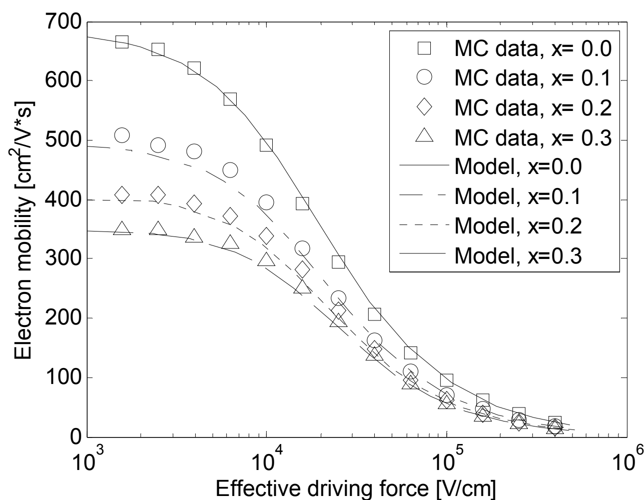


Fig. 3.25. Out-of-plane minority electron mobility in $Si_{1-x}Ge_x$ as a function of effective driving force for several germanium mole fraction; $N = N_A = 10^{17}$ cm^{-3}, T_L = 300 K.

3.3 References

[1] R.W. Dutton, A.J. Strojwas, "Perspectives on Technology and Technology-Driven CAD," *IEEE Transaction on Computer-Aided Design* vol. 19 n. 12, pp. 1544-1560, 2000.
[2] M. Al-Sa'di, V. d'Alessandro, S. Fregonese, S.-M. Hong, C. Jungemann, C. Maneux, I. Marano, A. Pakfar, N. Rinaldi, G. Sasso, M. Schröter, A. Sibaja-Hernandez, C. Tavernier, and G. Wedel, "TCAD simulation and development within the European DOTFIVE project on 500GHz SiGe:C HBT's", *European Microwave Integrated Circuits Conference*, pp. 29-32*,* September 2010.
[3] T. Grasser, S. Selberherr, "Technology CAD: Device simulation and characterization," *Vacuum Science and Technology B*, 20, pp. 407-413, 2002.
[4] Synopsys TCAD Software, Release 2007.03, Synopsys, Montain View, CA, 2007.
[5] C. Jungemann, B. Neinhus, and B. Meinerzhagen, "Comparative Study of Electron Transit Times Evaluated by DD, HD, and MC Device Simulation for a SiGe HBT," *IEEE Transaction On Electron Devices*, vol. 48, no.10, pp. 2216-2220, October 2001.
[6] P. Ashburn, SiGe Heterojunction Bipolar Transistors, John Wiley & Sons, 2003.
[7] D. B. M. Klaassen, J. W. Slotboom and H. C. de Graaff, "Unified apparent bandgap narrowing in n- and p-type silicon", *Solid-State Electronics*, vol. 35, no. 2, pp. 125-129, 1992.
[8] R. J. E. Hueting, J. W. Slotboom, A. Pruijmboom, W. B. de Boer, C. E. Timmering and N. E. B. Cowern, "On the optimization of SiGe-Base Bipolar Transistors", *Transaction on Electron Devices*, vol. 43, no. 9, pp. 1518-1524, 1996.
[9] G. Sasso, G. Matz, C. Jungemann, and N. Rinaldi, "Accurate Mobility and Energy Relaxation Time Models for SiGe HBTs Numerical Simulation", *International Conference on Simulation of Semiconductor Processes and Devices*, pp. 241-244, September 2009.
[10] G. Sasso, G. Matz, C. Jungemann, and N. Rinaldi, "Analytical Models of Effective Dos, Saturation Velocity and High-Field

Mobility for SiGe HBTs Numerical Simulation", *International Conference on Simulation of Semiconductor Processes and Devices*, pp. 279-282*,* September 2010.

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

[12] M. A Green, “Intrinsic concentration, effective densities of states, and effective mass in silicon,” *Journal of Applied Physics*, vol. 67, no. 6, pp. 2944-2954, June 1990.

[13] B. Pejcinovic, L. E. Kay, T. Tang, and D. H. Navon, “Numerical simulation and comparison of Si BJT’s and $Si_{1-x}Ge_x$ HBT’s,” *IEEE Transaction on Electron Devices*, vol. 36, no. 10, pp. 2129-2137, October 1989.

[14] R. People, “Physics and applications of Ge_xSi_{1-x}/Si strained-layer heterostructures,” *IEEE Jurnal of Quantum Electronics*, Vol. 22, No. 9, pp. 1696-1710, September 1986.

[15] D. M. Caughey, and R.F. Thomas, “Carrier Mobilities in Silicon Empirically related to Doping and Field,” *Proc. IEEE*, vol. 55, pp. 2192-2193, December 1967.

[16] N. D. Arora, J. R. Hauser, and D. J. Roulston, “Electron and Hole Mobilities in Silicon as a Function of Concentration and Temperature,” *IEEE Transaction on Electron Devices*, vol. 29, no. 2, pp. 292-295, February 1982.

[17] G. Masetti, M. Severi, and S. Solmi, “Modeling of Carrier Mobility against Carrier Concentration in Arsenic, Phosphorus and Boron-Doped Silicon,” *IEEE Transaction on Electron Devices*, vol. 30, no. 7, pp. 764-769, July 1983.

[18] S. Reggiani, M. Valdinocci, L. Colalongo, M. Rudan, G. Baccarani and A.D. Stricker et al., “Electron and Hole Mobility in Silicon at Large Operating Temperatures – Part I: Bulk Mobility,” *IEEE Transaction on Electron Devices*, vol. 49, no. 3, pp. 490-499, March 2002.

[19] D. B. M. Klaassen, “A unified mobility model for device simulation,” *International Electron Devices Meeting*, pp. 357-360, September 1990.

[20] J. D. Cressler, “Emerging application opportunities for SiGe technology”, *IEEE Custom Integrated Circuits Conference*, pp. 57-64, September 2008.

[21] S. E. Swirhun, Y. H. Kwark, and R. M. Swanson, “Temperature Dependence of Minority Electron Mobility and Bandgap narrowing in p^+ Si,” *International Electron Devices Meeting*, pp. 298-301, December 1988.

[22] C.H. Wang, K. Misiakos, and A. Neugroschel, “Temperature Dependence of Minority Hole Mobility in Heavily Doped Silicon,” *Applied Physics Letters*, pp. 159-161, July 1990.

[23] I. Y. Leu, and A. Neugroschel, “Minority-Carrier Transport Parameters in Heavily Doped p-Type Silicon at 296 and 77 K,” *IEEE Transaction on Electron Devices*, vol. 40, no. 10, pp.1872-1875, October 1993.

[24] M. Finetti and A. M. Mazzone, “Impurity Effects on Conduction in Heavily Doped n-Type Silicon,” *Journal of Applied Physics*, vol. 48, no. 11, pp. 4597-4600, November 1977.

[25] A. B. Sproul, M. A. Green, and A. W. Stephens, “Accurate Determination of Minority Carrier and Lattice Scattering Mobility in Silicon from Photoconductance Decay,” *Journal of Applied Physics*, vol. 72, no. 9, pp. 4161-4171, November 1992.

[26] J. Dziewior, and D. Silber, “Minority-Carrier Diffusion Coefficients in Highly Doped Silicon,” *Applied Physics Letters*, vol. 35, no. 2, pp. 170-172, July 1979.

[27] J. A. Del Alamo, S. E. Swirhun, and R. M. Swanson, “Measurement and Modeling of Minority Carrier Transport in Heavily-Doped Silicon,” *Solid-State Electronics*, vol.28, no. 1-2, pp. 47-52, January-February 1985.

[28] S. S. Li, and W. R. Thurber, “The Dopant Density and Temperature Dependence of Electron Mobility and Resistivity in n-Type Silicon,” *Solid State Electronics*, vol. 20, no. 7, pp. 609-616, July 1977.

[29] S. S. Li, “The Dopant Density and Temperature Dependence of Hole Mobility and Resistivity in Boron-Doped Silicon,” *Solid State Electronics*, vol. 21, no. 9, pp. 1109-1117, September 1978.

[30] V. Palankovski, G. Roehrer, T. Grasser, S. Smirnov, H. Kosina, ans S. Selberherr, “Rigorous modeling approach to numerical simulation of SiGe HBTs,” *Applied Surface Science*, vol. 224, no. 1-4, pp. 361-364, March 2004.

[31] V. Palankovski and R. Quay, *Analysis and Simulation of Heterostructure Devices*. Berlin, Germany: Springer-Verlag, 2004.

[32] S. Smirnov, H. Kosina, and S. Selberherr, “Investigation of the electron mobility in strained $Si_{1-x}Ge_x$ at high Ge composition,” *International Conference on Simulation of Semiconductor Processes and Devices*, pp. 29-32, 2002.

[33] P. Gaworzewski, K. Tittelbach-Helmrich, and U. Penner, “Electrical Properties of Lightly Doped p-Type Silicon-Germanium Single Crystals,” *Journal of Applied Physics*, vol. 83, no. 10, pp. 5258-5263, May 1998.

[34] T. K. Carns, S. K. Chun, M. O. Tanner, K. L. Wang, T. I. Kamins, J. T. Turner, D. Y. C. Lie, M.-A. Nicolet, and R. G. Wilson, “Hole mobility measurements in heavily doped $Si_{1-x}Ge_x$ strained layers,” *IEEE Transaction on Electron Devices*, vol. 41, no. 7, pp.1273-1281, July 1994.

[35] B. Gonzales, V. Palankovski, H. Kosina, A. Hernandez, and S. Selberherr, “An energy relaxation time model for device simulation,” *Solid State Electronics*, vol. 43, no. 9, pp. 1791-1795, September 1999.

[36] M. V. Fischetti, “Monte Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zinc-Blende Structures-Part I: Homogeneous Transport,” *IEEE Transaction on Electron Devices*, vol. 38, no. 3, pp. 634-649, March 1991.

[37] V. M. Agostinelli, T. J. Bordelon, X. L. Wang, C. F. Yeap, C. M. Mazir, and A. F. Tasch, “An energy-dependent two-dimensional substrate current model for the simulation of submicrometer MOSFETs,” *IEEE Electron Device Letters*, vol. 13, no. 11, pp.554-556, November 1992.

[38] B. Neinhues, S. Decker, P. Graf, F. M. Bufler, and B. Meinerzhagen, “Consistent Hydrodynamic and Monte-Carlo Simulations of SiGe HBTs Based on Table Models for the Relaxation Times,” *VLSI Design*, pp. 387-391, 1998.

[39] M. Ershov, and V. Ryzhii, “High-field electron transport in SiGe alloy,” *Japanese Journal of Applied Physics*, vol. 33, pp. 1365-1371, 1994.

[40] R. Quay, C. Moglestue, V. Palankovski, and S. Selberherr, “A temperature dependent model for the saturation velocity in semiconductor materials,” *Materials Science in Semiconductor Processing*, vol. 3, no. 1-2, pp. 149-155, March 2000.
[41] C. Canali, G. Majni, R. Minder, and G. P. Ottaviani, “Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature,” *IEEE Transaction on Electron Devices*, vol. 22, no. 11, pp. 1045-1047, November 1975.
[42] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, “A review of hydrodynamic and energy-transport models for semiconductor device simulation,” *Proceedings IEEE*, vol. 91, no. 2, pp. 251-274, 2003.
[43] Atlas Users Manual, SILVACO, 2000.
[44] Y. Apanovich, E. Lyumkis, B. Polsky, A. Shur, and P. Blakey, “Steady-state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models,” *IEEE Transaction on CAD*, vol. 13, no. 6, pp. 702-711, June 1994.
[45] G. Sasso, G. Matz, C. Jungemann, and N. Rinaldi, “Accurate mobility and energy relaxation time models for SiGe HBTs numerical simulation,” *International Conference on Simulation of Semiconductor Processes and Devices*, pp. 241-244, September 2009.
[46] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberighi-Quaranta, “Electron drift velocity in silicon,” *Physical Review B*, vol. 12, no. 4, pp. 2265-2284, April 1975.
[47] B. K. Ridley, and T. B. Watkins, “The possibility of negative resistance effects in semiconductors,” *Proceedings of the Physical Society*, vol. 78 n. 2, pp. 293-304, 1961.
[48] F. M. Bufler, P. Graf, S. Keith, and B. Meinerzhagen, “Full band Monte Carlo investigation of electron transport in strained Si grown on $Si_{1-x}Ge_x$ substrates,” *Applied Physics Letters*, vol. 70 no. 16, pp. 2144-2146, April 1997.

Chapter 4

Verification of transport models

In Chapter 3 we presented a derivation of new analytical models for transport parameters and comparison with state-of-art models. In this chapter we present a detailed description of models' implementation in a commercial device simulator [1] and report hydrodynamic device simulations based on the novel analytical transport models for several one-dimensional and two-dimensional structures with a different f_T. HD simulation results are compared to results obtained using a standard set of models for silicon-germanium and to more reliable results (i.e. MC and SPRING simulation results and experimental data) with the aim of validating proposed models and clarifying their reliability and accuracy over different technologies.

4.1 Implementation of analytical models in *Sdevice*

The full set of developed models has been successfully implemented in the device simulator (*Sdevice)* of the *Sentaurus TCAD* by *SYNOPSYS* [1] using its *Physical Model Interface* (*PMI*). This interface provides a direct access to the transport parameters models to be used in the semiconductor transport equations. The user can provide new *C++* functions to compute these models, and *Sdevice* loads the functions at run-time using the dynamic loader. No access to the *Sdevice* source code is necessary. For each *PMI* model, the user must implement a *C++* subroutine to evaluate the model. Additional subroutines are necessary to evaluate the derivatives of the model with respect to all the input variables. More specifically, the user must implement a *C++* class that is derived from a base class declared in the header file *PMIModels.h*. In addition, a so-called virtual

constructor function must be provided, which allocates an instance of the derived class.

As instances, the low field mobility and energy relaxation time models implementation are quoted at the end of this chapter.

4.2 Reference state-of-art models

Based on Monte Carlo simulations [2], an *Sdevice* parameter file is available in the software library of materials. This file contains transport parameters at 300 K for silicon-germanium under biaxial compressive strain present. Indeed, a biaxial compressive strain is occurs in the base of npn-SiGe HBTs which is grown on top of a relaxed silicon substrate. The parameters for electrons refer to the out-of-plane direction (that is perpendicular to the SiGe/Si interface) and the parameters for holes to the in-plane direction (that is parallel to the SiGe/Si interface). The transport parameters have been obtained from the full-band Monte Carlo simulations in [2]. The band gap in relaxed SiGe alloys has been extracted from the measurements in [3] and the values in strained SiGe are calculated according to the model solid theory of C. G. Van de Walle [4].

Simulations have been performed using this parameter file for evaluation and comparison with the novel developed analytical models. The models used for this comparison are detailed in Table 4.1 and will be referred to as “Bufler models” henceforth.

Table 4.1. Bufler models

PARAMETER	MODEL
Low-field mobility	*Philips unified mobility model*
High-field mobility	*Extended Canali model*
Bandgap narrowing	*Braunstein et Al.*
Effective density of states	*Slotboom*
Energy relaxation time	*Constant for Silicon*
	Linear dependent on x for $Si_{1-x}Ge_x$

4.3 Verification of models

Having described the implementation of the novel models in *Sdevice*, we discuss their accuracy for devices of different technology nodes.

4.3.1 One-dimensional 100 GHz device

The verification of the new models has been performed for a 1D reference structure whose maximum cut-off frequency is about 100 GHz and whose profile is depicted in Fig. 4.1.

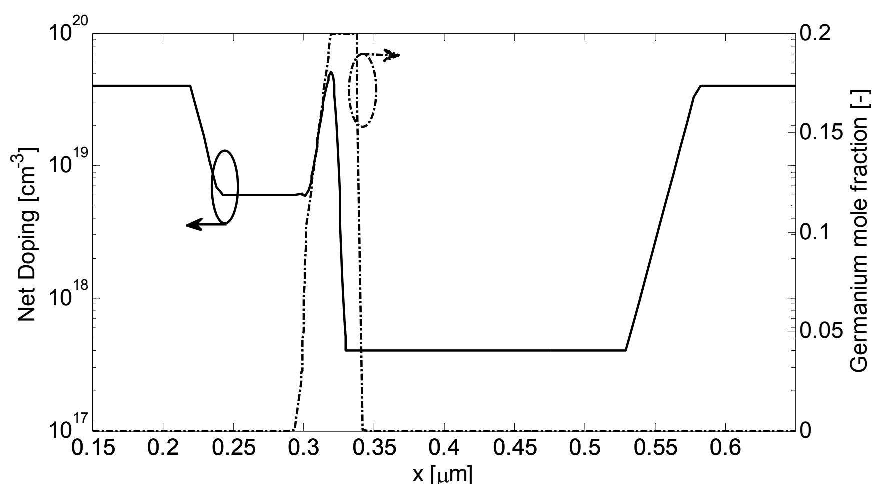


Fig. 4.1. Doping profile of the 100-GHz SiGe HBT.

Simulation results for the new developed model set are reported in Fig. 4.2 and Fig. 4.3. Model results are compared to MC data and to results obtained for the same structure using Bufler models; MC data used for verification have been generated by Bundesweher University. As can be seen in Fig. 4.2 and Fig. 4.3, using the new models we achieve a marked improvement in simulation results for the collector current. On the other hand, the cut-off frequency versus collector current density plot depicted in Fig. 4.3 shows that, with the exception

of high current region where f_T plummets due to the Kirk effect, the results obtained using Bufler models' also provide an adequate description.

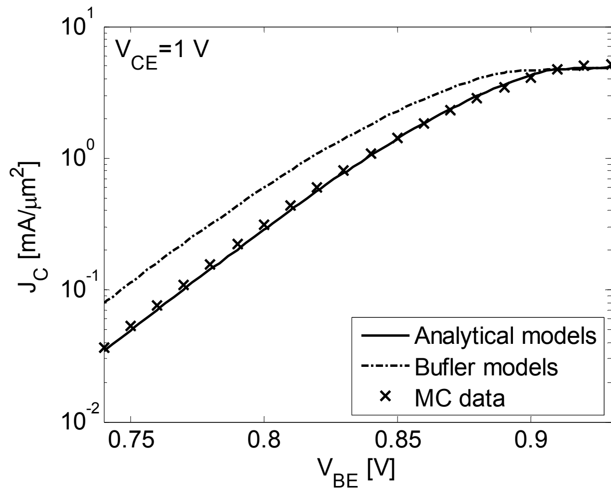


Fig. 4.2. Transfer characteristics at V_{CE} = 1 V: comparison between MC data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

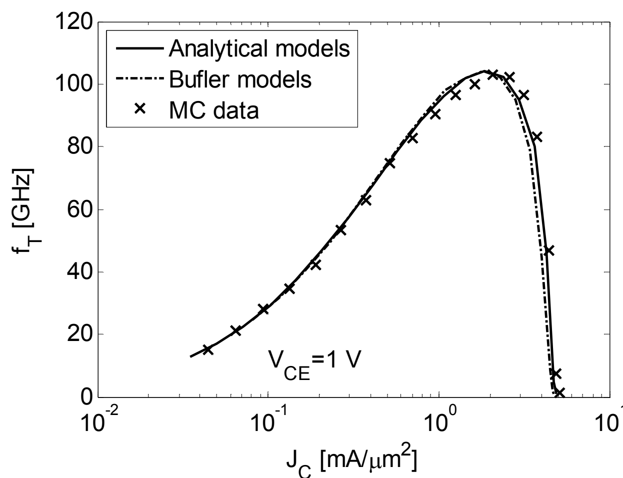


Fig. 4.3. Cut-off frequency at V_{CE} = 1 V: comparison between MC data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

4.3.2 One-dimensional 450 GHz device

The proposed models have been applied to the simulation of a 1D structure whose maximum cut-off frequency is about 450 HD simulations with the new models have been compared to the results obtained by *SPRING*. Device profile is depicted in Fig. 4.4.

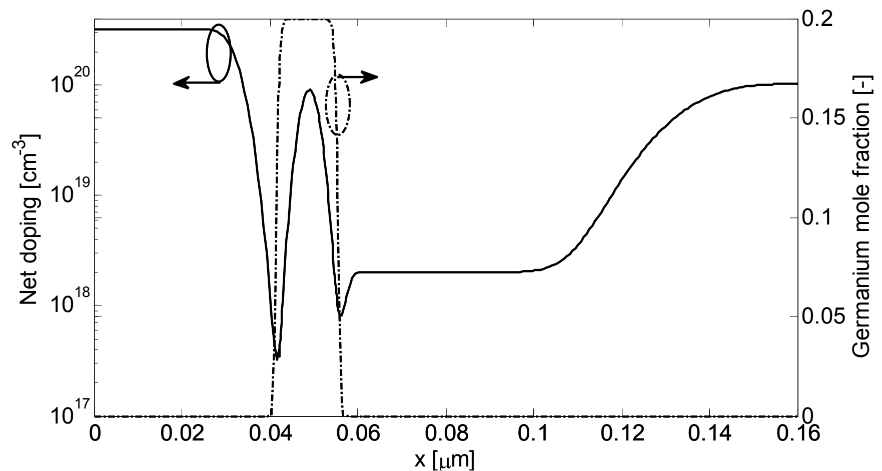


Fig. 4.4. Doping profile of the 450-GHz SiGe HBT.

HD simulations have been performed using the proposed analytical parameters, Bufler models and reliable look-up tables described in §3.1. The results of this comparison are depicted in Fig. 4.5 ÷ Fig. 4.7 for the collector current density, base current density and cut-off frequency respectively.

An inspection of the Gummel plot in Fig. 4.5 confirms that new models offer an improved accuracy with respect to the Bufler models. A comparison for the maximum cut-off frequency yields the same conclusions as in §4.3.1. Also shown are the simulation results for the calibrated look-up table models used for analytical models development.

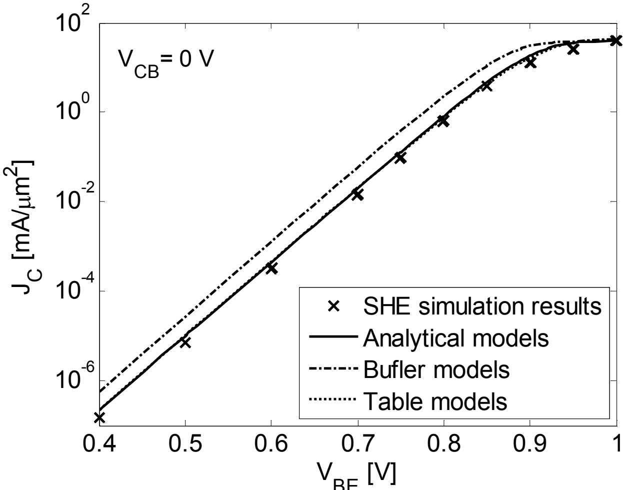


Fig. 4.5. Transfer characteristic at V_{CB} = 0 V: comparison between SHE data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

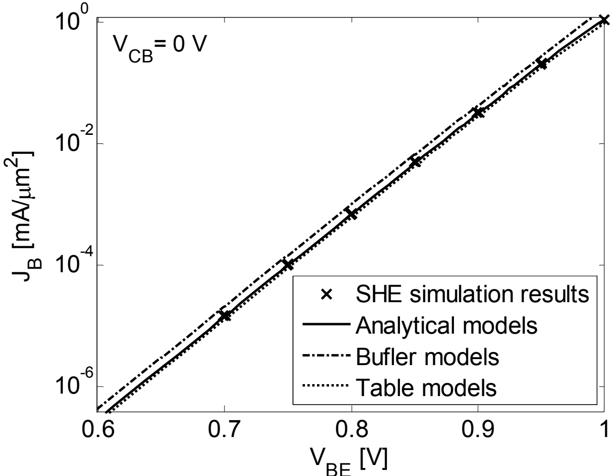


Fig. 4.6. Base current density in transfer characteristic at V_{CB} = 0 V: comparison between SHE data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

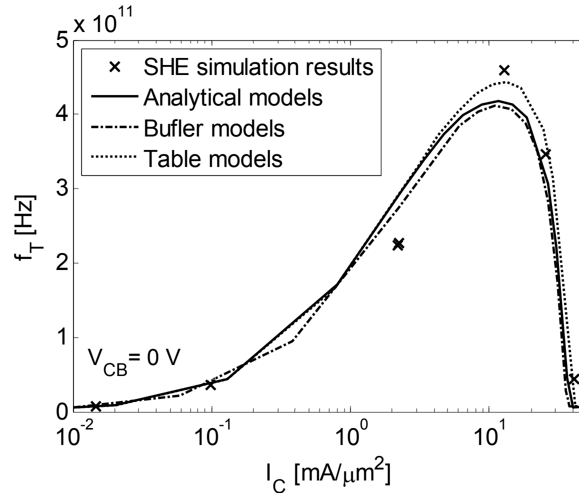


Fig. 4.7. Cut-off frequency at V_{CB} = 0 V: comparison between SHE data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

4.3.3 Two-dimensional 230 GHz device.

Models verification has been repeated for the 2D HBT simplified structure depicted in Fig. 4.8. The vertical doping profile is shown in Fig. 4.9. The 2D doping concentration is described by analytical expressions derived as an approximation of a complete structure obtained by process simulation and provided by *STMicroelectronics*. The fabrication process is described in [5] and is intended for millimeter-wave applications. The technology has a 230-GHz f_T / 280-GHz f_{max} HBT, which is achieved with a fully self-aligned (FSA) architecture using selective epitaxial growth of the base. This technology is mainly intended for 60-GHz WLAN, 77-GHz automotive radars and 100-Gb/s optical communications.

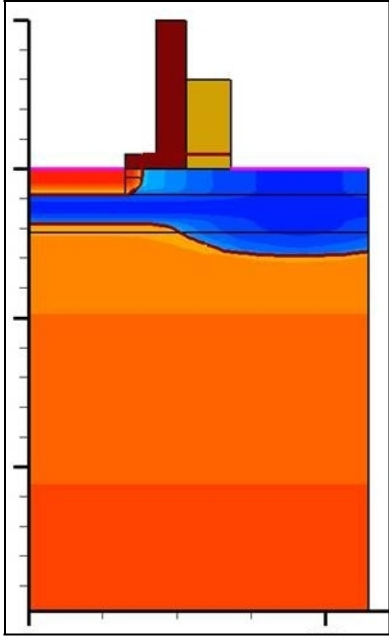


Fig. 4.8. Two-dimensional doping concentration of the 230 GHz SiGe HBT.

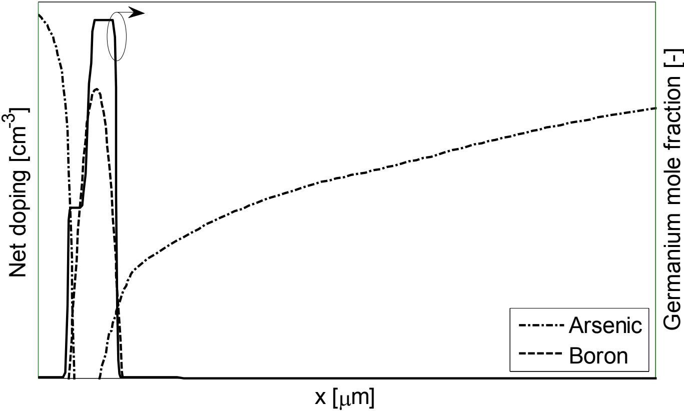


Fig. 4.9. One-dimensional doping profile of the 230 GHz SiGe HBT for the intrinsic region.

Simulation results for the simplified 2D structure are depicted in Fig. 4.10 and Fig. 4.11. In this figure the new transport models are compared to Bufler set and to calibrated results provided by *STMicroelectronics*. The latter results are obtained using confidential transport parameters' models, empirically calibrated with measurements. In order to correctly evaluate the differences between results, it should be recalled that STMicroelectronics results refer to a complete 3D structure (i.e. including buried layer, shallow trench, substrate region, parasitic capacitances at Via0 level etc). The results pertaining the Gummel plot in Fig. 4.10 confirm that the proposed model set improves the accuracy of HD simulation. The significant differences in maximum cut-off frequency values at low current injection levels can be justified by the difference in the simulated structure.

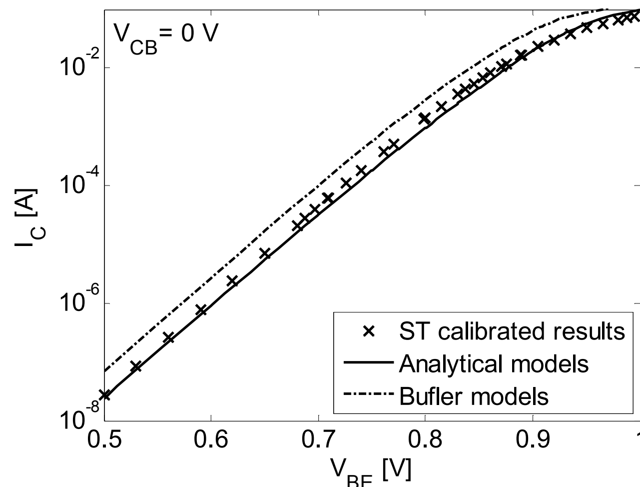


Fig. 4.10. Transfer characteristic at V_{CB} = 0 V: comparison between calibrated data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

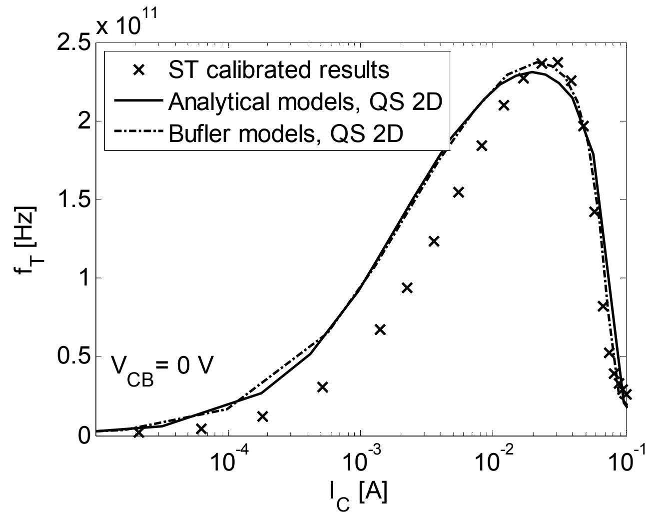


Fig. 4.11. Cut-off frequency at $V_{CB} = 0$ V: comparison between calibrated data and HD simulation results obtained with default (Bufler) and proposed analytical models for transport parameters.

4.4 *C++* code examples for *Sdevice* implementation

In this section we report the details of the implementation, implementation of the low field mobility model and the energy relaxation time model.

4.4.1 Low field mobility model

```
#include "PMIModels.h"
#include <cmath>

#ifndef Max
#define Max( a, b )    ((a) < (b) ? (b) : (a))
#endif
```

```
class sasso_DopingDepMobility : public PMI_DopingDepMobility {
protected:
 const double T0;
 double umax0_Si, gamma_Si, u0d0_Si, u0d1_Si, u0a0_Si, u0a1_Si,
Cr10_Si, Cr11_Si, Cr20_Si, Cr21_Si, Cs1_Si, Cs2_Si, alfa1_Si,
alfa2_Si, u1da_Si, u1db_Si, u1dc_Si, u1aa_Si, u1ab_Si, u1ac_Si;
 double umax0_Ge, gamma_Ge, u0d0_Ge, u0d1_Ge, u0a0_Ge,
u0a1_Ge, Cr10_Ge, Cr11_Ge, Cr20_Ge, Cr21_Ge, Cs1_Ge, Cs2_Ge,
alfa1_Ge, alfa2_Ge, u1da_Ge, u1db_Ge, u1dc_Ge, u1aa_Ge,
u1ab_Ge, u1ac_Ge;
 double Cu, alfa;

public:
 sasso_DopingDepMobility (const PMI_Environment& env,
                          const PMI_AnisotropyType anisotype );
 ~sasso_DopingDepMobility () {}

 void Compute_m
     (const double n, const double p,
      const double t, double& m);

 void Compute_dmdn
     (const double n, const double p,
      const double t, double& dmdn);

 void Compute_dmdp
     (const double n, const double p,
      const double t, double& dmdp);

 void Compute_dmdt
     (const double n, const double p,
      const double t, double& dmdt);
};

sasso_DopingDepMobility::
sasso_DopingDepMobility (const PMI_Environment& env,
                         const PMI_AnisotropyType anisotype) :
    PMI_DopingDepMobility (env, anisotype),
```

```
    T0 (300.0)
{
}

void sasso_DopingDepMobility::
Compute_m    (const double n, const double p,
         const double t, double& m)
{ const double umax_Si = umax0_Si * pow ((t/T0), gamma_Si);
  const double u1d_Si = (u1da_Si * pow ((t/T0),2)+ u1db_Si * (t/ T0)
+ u1dc_Si)/(pow((t/T0),2));
  const double u1a_Si = (u1aa_Si * pow ((t/T0),2)+ u1ab_Si * (t/ T0)
+ u1ac_Si)/(pow((t/T0),2));
  const double u0d_Si = u0d0_Si * pow ((t/T0), u0d1_Si);
  const double u0a_Si = u0a0_Si * pow ((t/T0), u0a1_Si);
  const double Cr1_Si = Cr10_Si * pow ((t/T0), Cr11_Si);
  const double Cr2_Si = Cr20_Si * pow ((t/T0), Cr21_Si);
  const double ND = Max (ReadDoping (PMI_Donor), 1.0);
  const double NA = Max (ReadDoping (PMI_Acceptor), 1.0);
  const double u0_Si = (u0d_Si*ND+u0a_Si*NA)/(ND+NA);
  const double u1_Si = (u1d_Si*ND+u1a_Si*NA)/(ND+NA);
  const double m_Si = (umax_Si-u0_Si)/(1 +
pow((ND/Cr1_Si),alfa1_Si)+pow((NA/Cr2_Si),alfa2_Si))+u0_Si-
u1_Si/(1+pow(((ND/Cs1_Si)+(NA/Cs2_Si)),-2.0));
  const double umax_Ge=umax0_Ge * pow ((t/T0), gamma_Ge);
  const double u1d_Ge = (u1da_Ge * pow ((t/T0),2)+ u1db_Ge * (t/
T0) + u1dc_Ge)/(pow((t/T0),2));
  const double u1a_Ge = (u1aa_Ge * pow ((t/T0),2)+ u1ab_Ge * (t/
T0) + u1ac_Ge)/(pow((t/T0),2));
  const double u0d_Ge = u0d0_Ge * pow ((t/T0), u0d1_Ge);
  const double u0a_Ge = u0a0_Ge * pow ((t/T0), u0a1_Ge);
  const double Cr1_Ge = Cr10_Ge * pow ((t/T0), Cr11_Ge);
  const double Cr2_Ge = Cr20_Ge * pow ((t/T0), Cr21_Ge);
  const double u0_Ge = (u0d_Ge*ND+u0a_Ge*NA)/(ND+NA);
  const double u1_Ge = (u1d_Ge*ND+u1a_Ge*NA)/(ND+NA);
  const double m_Ge = (umax_Ge-u0_Ge)/(1 +
pow((ND/Cr1_Ge),alfa1_Ge)+pow((NA/Cr2_Ge),alfa2_Ge))+u0_Ge-
u1_Ge/(1+pow(((ND/Cs1_Ge)+(NA/Cs2_Ge)),-2.0));
  m =pow((1-
ReadxMoleFraction()/0.3)/m_Si+(ReadxMoleFraction()/0.3)/m_Ge+(
```

```
1-
pow(ReadxMoleFraction()/0.3,alfa))*pow((ReadxMoleFraction()/0.3)
,alfa)/Cu,-1);
}

void sasso_DopingDepMobility::
Compute_dmdn (const double n, const double p,
            const double t, double& dmdn)
{ dmdn = 0.0;
}

void sasso_DopingDepMobility::
Compute_dmdp (const double n, const double p,
            const double t, double& dmdp)
{ dmdp = 0.0;
}

void sasso_DopingDepMobility::
Compute_dmdt (const double n, const double p,
            const double t, double& dmdt)
{ const double umax_Si=umax0_Si * pow ((t/T0), gamma_Si);
  const double u1d_Si = (u1da_Si * pow ((t/T0),2)+ u1db_Si * (t/ T0)
+ u1dc_Si)/(pow((t/T0),2));
  const double u1a_Si = (u1aa_Si * pow ((t/T0),2)+ u1ab_Si * (t/ T0)
+ u1ac_Si)/(pow((t/T0),2));
  const double u0d_Si = u0d0_Si * pow ((t/T0), u0d1_Si);
  const double u0a_Si = u0a0_Si * pow ((t/T0), u0a1_Si);
  const double Cr1_Si = Cr10_Si * pow ((t/T0), Cr11_Si);
  const double Cr2_Si = Cr20_Si * pow ((t/T0), Cr21_Si);
  const double ND = Max (ReadDoping (PMI_Donor), 1.0);
  const double NA = Max (ReadDoping (PMI_Acceptor), 1.0);
  const double u0_Si = (u0d_Si*ND+u0a_Si*NA)/(ND+NA);
  const double u1_Si = (u1d_Si*ND+u1a_Si*NA)/(ND+NA);
  const double u01_Si=(ND*u0d1_Si+NA*u0a1_Si)/(NA+ND);
  const double dm_b_Si= (u0_Si+u01_Si)/t;
  const double dm_c_Si=
(ND*(u1db_Si+2.0*u1dc_Si*T0/t)+NA*(u1ab_Si+2.0*u1ac_Si*T0/t)
)/(T0*(NA+ND)*(1+pow((ND/Cs1_Si+NA/Cs2_Si),-2)));
```

```
  const double dm_a_Si= (1/t)*((gamma_Si*umax_Si-u0_Si-
u01_Si)*(1+pow((ND/Cr1_Si),alfa1_Si)+pow((NA/Cr2_Si),alfa2_Si))
+(umax_Si-
u0_Si)*(alfa1_Si*Cr11_Si*pow((ND/Cr1_Si),alfa1_Si)/ND+alfa2_Si
*Cr21_Si*pow((NA/Cr2_Si),alfa2_Si)/NA))/pow((1+pow((ND/Cr1_Si
),alfa1_Si)+pow((NA/Cr2_Si),alfa2_Si)),2);
  const double m_Si = (umax_Si-u0_Si)/(1 +
pow((ND/Cr1_Si),alfa1_Si)+pow((NA/Cr2_Si),alfa2_Si))+u0_Si-
u1_Si/(1+pow(((ND/Cs1_Si)+(NA/Cs2_Si)),-2.0));
  const double dmdt_Si= dm_a_Si+dm_b_Si-dm_c_Si;
  const double umax_Ge=umax0_Ge * pow ((t/T0), gamma_Ge);
  const double u1d_Ge = (u1da_Ge * pow ((t/T0),2)+ u1db_Ge * (t/
T0) + u1dc_Ge)/(pow((t/T0),2));
  const double u1a_Ge = (u1aa_Ge * pow ((t/T0),2)+ u1ab_Ge * (t/
T0) + u1ac_Ge)/(pow((t/T0),2));
  const double u0d_Ge = u0d0_Ge * pow ((t/T0), u0d1_Ge);
  const double u0a_Ge = u0a0_Ge * pow ((t/T0), u0a1_Ge);
  const double Cr1_Ge = Cr10_Ge * pow ((t/T0), Cr11_Ge);
  const double Cr2_Ge = Cr20_Ge * pow ((t/T0), Cr21_Ge);
  const double u0_Ge = (u0d_Ge*ND+u0a_Ge*NA)/(ND+NA);
  const double u1_Ge = (u1d_Ge*ND+u1a_Ge*NA)/(ND+NA);
  const double u01_Ge=(ND*u0d1_Ge+NA*u0a1_Ge)/(NA+ND);
  const double dm_b_Ge= (u0_Ge+u01_Ge)/t;
  const double dm_c_Ge=
(ND*(u1db_Ge+2.0*u1dc_Ge*T0/t)+NA*(u1ab_Ge+2.0*u1ac_Ge*T
0/t))/(T0*(NA+ND)*(1+pow((ND/Cs1_Ge+NA/Cs2_Ge),-2)));
  const double dm_a_Ge= (1/t)*((gamma_Ge*umax_Ge-u0_Ge-
u01_Ge)*(1+pow((ND/Cr1_Ge),alfa1_Ge)+pow((NA/Cr2_Ge),alfa2_
Ge))+(umax_Ge-
u0_Ge)*(alfa1_Ge*Cr11_Ge*pow((ND/Cr1_Ge),alfa1_Ge)/ND+alfa
2_Ge*Cr21_Ge*pow((NA/Cr2_Ge),alfa2_Ge)/NA))/pow((1+pow((N
D/Cr1_Ge),alfa1_Ge)+pow((NA/Cr2_Ge),alfa2_Ge)),2);
  const double m_Ge = (umax_Ge-u0_Ge)/(1 +
pow((ND/Cr1_Ge),alfa1_Ge)+pow((NA/Cr2_Ge),alfa2_Ge))+u0_Ge-
u1_Ge/(1+pow(((ND/Cs1_Ge)+(NA/Cs2_Ge)),-2.0));
  const double dmdt_Ge= dm_a_Ge+dm_b_Ge-dm_c_Ge;

  const double mtot=pow((1-
ReadxMoleFraction()/0.3)/m_Si+(ReadxMoleFraction()/0.3)/m_Ge+(
```

```
1-
pow(ReadxMoleFraction()/0.3,alfa))*pow((ReadxMoleFraction()/0.3)
,alfa)/Cu,-1);

  dmdt=pow(mtot,2)*((1-
ReadxMoleFraction()/0.3)*dmdt_Si/pow(m_Si,2)-
(ReadxMoleFraction()/0.3)*dmdt_Ge/pow(m_Ge,2));
}

class sasso_e_DopingDepMobility : public sasso_DopingDepMobility
{
public:
  sasso_e_DopingDepMobility (const PMI_Environment& env,
                             const PMI_AnisotropyType anisotype);
  ~sasso_e_DopingDepMobility () {}
};
sasso_e_DopingDepMobility::
sasso_e_DopingDepMobility(const PMI_Environment& env,
                             const PMI_AnisotropyType anisotype) :
 sasso_DopingDepMobility (env, anisotype)

{umax0_Si=InitParameter ("umax0_Si_e", 1421.6);
 gamma_Si=InitParameter ("gamma_Si_e", -2.24);
 u0d0_Si=InitParameter ("u0d0_Si_e", 49);
 u0d1_Si=InitParameter ("u0d1_Si_e", 0);
 u0a0_Si=InitParameter ("u0a0_Si_e", 205.25);
 u0a1_Si=InitParameter ("u0a1_Si_e", -0.934);
 Cr10_Si=InitParameter ("Cr10_Si_e", 8.393e16);
 Cr11_Si=InitParameter ("Cr11_Si_e", 2.951);
 Cr20_Si=InitParameter ("Cr20_Si_e", 5.42e16);
 Cr21_Si=InitParameter ("Cr21_Si_e", 3.045);
 Cs1_Si=InitParameter ("Cs1_Si_e", 1.81e19);
 Cs2_Si=InitParameter ("Cs2_Si_e", 4.2e19);
 alfa1_Si=InitParameter ("alfa1_Si_e", 0.68);
 alfa2_Si=InitParameter ("alfa2_Si_e", 0.7);
 u1da_Si=InitParameter ("u1da_Si_e", 24.812);
 u1db_Si=InitParameter ("u1db_Si_e", 84.06);
 u1dc_Si=InitParameter ("u1dc_Si_e", -132.36);
```

```
 u1aa_Si=InitParameter ("u1aa_Si_e", -94.533);
 u1ab_Si=InitParameter ("u1ab_Si_e", 518.28);
 u1ac_Si=InitParameter ("u1ac_Si_e", -419.15);

umax0_Ge=InitParameter ("umax0_Ge_e", 453.23);
 gamma_Ge=InitParameter ("gamma_Ge_e", -1.14);
 u0d0_Ge=InitParameter ("u0d0_Ge_e", 91.587);
 u0d1_Ge=InitParameter ("u0d1_Ge_e", -1.0547);
 u0a0_Ge=InitParameter ("u0a0_Ge_e", 191.58);
 u0a1_Ge=InitParameter ("u0a1_Ge_e", -0.92012);
 Cr10_Ge=InitParameter ("Cr10_Ge_e", 3.8812e17);
 Cr11_Ge=InitParameter ("Cr11_Ge_e", 2.0845);
 Cr20_Ge=InitParameter ("Cr20_Ge_e", 1.913e17);
 Cr21_Ge=InitParameter ("Cr21_Ge_e", 2.4096);
 Cs1_Ge=InitParameter ("Cs1_Ge_e", 6e19);
 Cs2_Ge=InitParameter ("Cs2_Ge_e", 5.4e19);
 alfa1_Ge=InitParameter ("alfa1_Ge_e", 0.76);
 alfa2_Ge=InitParameter ("alfa2_Ge_e", 0.7);
 u1da_Ge=InitParameter ("u1da_Ge_e", 5.4938);
 u1db_Ge=InitParameter ("u1db_Ge_e", 95.873);
 u1dc_Ge=InitParameter ("u1dc_Ge_e", -96.001);
 u1aa_Ge=InitParameter ("u1aa_Ge_e", 49.072);
 u1ab_Ge=InitParameter ("u1ab_Ge_e", 87.321);
 u1ac_Ge=InitParameter ("u1ac_Ge_e", -131.97);

 Cu=InitParameter ("Cu_e", 2379.4);
 alfa=InitParameter ("alfa_e", 0.487);
}

class sasso_h_DopingDepMobility : public sasso_DopingDepMobility
{
public:
  sasso_h_DopingDepMobility (const PMI_Environment& env,
                             const PMI_AnisotropyType anisotype);
  ~sasso_h_DopingDepMobility () {}
};
sasso_h_DopingDepMobility::
sasso_h_DopingDepMobility(const PMI_Environment& env,
                             const PMI_AnisotropyType anisotype) :
```

```
sasso_DopingDepMobility (env, anisotype)

{umax0_Si=InitParameter ("umax0_Si_h", 485.51);
 gamma_Si=InitParameter ("gamma_Si_h", -2.49);
 u0d0_Si=InitParameter ("u0d0_Si_h", 123.34);
 u0d1_Si=InitParameter ("u0d1_Si_h", -1.028);
 u0a0_Si=InitParameter ("u0a0_Si_h", 46.42);
 u0a1_Si=InitParameter ("u0a1_Si_h", -0.627);
 Cr10_Si=InitParameter ("Cr10_Si_h", 1.329e17);
 Cr11_Si=InitParameter ("Cr11_Si_h", 3.07);
 Cr20_Si=InitParameter ("Cr20_Si_h", 1.631e17);
 Cr21_Si=InitParameter ("Cr21_Si_h", 3.111);
 Cs1_Si=InitParameter ("Cs1_Si_h", 5.1e19);
 Cs2_Si=InitParameter ("Cs2_Si_h", 5.8e19);
 alfa1_Si=InitParameter ("alfa1_Si_h", 0.7);
 alfa2_Si=InitParameter ("alfa2_Si_h", 0.77);
 u1da_Si=InitParameter ("u1da_Si_h", -68.801);
 u1db_Si=InitParameter ("u1db_Si_h", 300.19);
 u1dc_Si=InitParameter ("u1dc_Si_h", -230.53);
 u1aa_Si=InitParameter ("u1aa_Si_h", -5.7832);
 u1ab_Si=InitParameter ("u1ab_Si_h", 81.146);
 u1ac_Si=InitParameter ("u1ac_Si_h", -76.025);

umax0_Ge=InitParameter ("umax0_Ge_h", 641.08);
 gamma_Ge=InitParameter ("gamma_Ge_h", -2.118);
 u0d0_Ge=InitParameter ("u0d0_Ge_h", 130.24);
 u0d1_Ge=InitParameter ("u0d1_Ge_h", -1.3315);
 u0a0_Ge=InitParameter ("u0a0_Ge_h", 40.848);
 u0a1_Ge=InitParameter ("u0a1_Ge_h", -0.63301);
 Cr10_Ge=InitParameter ("Cr10_Ge_h", 1.1509e17);
 Cr11_Ge=InitParameter ("Cr11_Ge_h", 4.2018);
 Cr20_Ge=InitParameter ("Cr20_Ge_h", 1.3873e17);
 Cr21_Ge=InitParameter ("Cr21_Ge_h", 3.2117);
 Cs1_Ge=InitParameter ("Cs1_Ge_h", 2e20);
 Cs2_Ge=InitParameter ("Cs2_Ge_h", 7e19);
 alfa1_Ge=InitParameter ("alfa1_Ge_h", 0.59);
 alfa2_Ge=InitParameter ("alfa2_Ge_h", 0.65);
 u1da_Ge=InitParameter ("u1da_Ge_h", -90.178);
```

```
 u1db_Ge=InitParameter ("u1db_Ge_h", 325.95);
 u1dc_Ge=InitParameter ("u1dc_Ge_h", -214.83);
 u1aa_Ge=InitParameter ("u1aa_Ge_h", 3.1834);
 u1ab_Ge=InitParameter ("u1ab_Ge_h", 59.611);
 u1ac_Ge=InitParameter ("u1ac_Ge_h", -57.908);

 Cu=InitParameter ("Cu_h", 556.4);
 alfa=InitParameter ("alfa_h", 0.548);
}

extern "C"
PMI_DopingDepMobility* new_PMI_DopingDep_e_Mobility
   (const PMI_Environment& env, const PMI_AnisotropyType
anisotype)
{ return new sasso_e_DopingDepMobility (env, anisotype);
}

extern "C"
PMI_DopingDepMobility* new_PMI_DopingDep_h_Mobility
   (const PMI_Environment& env, const PMI_AnisotropyType
anisotype)
{ return new sasso_h_DopingDepMobility (env, anisotype);
}
```

4.4.2 Energy relaxation time model

```
#include "PMIModels.h"
#include <cmath>

class Analytical_EnergyRelaxationTime : public
PMI_EnergyRelaxationTime {
 protected:
  const double T0;
  double Tl;
  double tw0Si, tw0Ge, Ctau, tw1, C1Si, C1Ge, Cc, C2, C3;
```

```
 public:
  Analytical_EnergyRelaxationTime (const PMI_Environment& env);

  ~Analytical_EnergyRelaxationTime ();

  void Compute_tau (const double ct, double& tau) ;
  void Compute_dtaudct (const double ct, double& dtaudct) ;
};

Analytical_EnergyRelaxationTime ::
Analytical_EnergyRelaxationTime (const PMI_Environment& env)  :
  PMI_EnergyRelaxationTime (env), T0 (300.0){
   Tl = InitParameter ("lattice_temperature", 300.);
  }

Analytical_EnergyRelaxationTime ::
~Analytical_EnergyRelaxationTime (){
}

void Analytical_EnergyRelaxationTime::
Compute_tau (const double ct, double& tau)
{  const double tw0 = tw0Si * (1 - ReadxMoleFraction()/0.3) + tw0Ge
* ReadxMoleFraction()/0.3 + Ctau * (1 - ReadxMoleFraction()/0.3) *
ReadxMoleFraction()/0.3;
   const double C1 = C1Si * (1 - ReadxMoleFraction()/0.3) + C1Ge *
ReadxMoleFraction()/0.3 + Cc * (1 - ReadxMoleFraction()/0.3) *
ReadxMoleFraction()/0.3;
   tau = tw0 + tw1 * exp (C1 * pow (ct/T0, 2)+ C2 * (ct/T0) + C3 *
(Tl/T0));
}

void Analytical_EnergyRelaxationTime::
Compute_dtaudct (const double ct, double& dtaudct)
{  const double tw0 = tw0Si * (1 - ReadxMoleFraction()/0.3) + tw0Ge
* ReadxMoleFraction()/0.3 + Ctau * (1 - ReadxMoleFraction()/0.3) *
ReadxMoleFraction()/0.3;
```

```
  const double C1 = C1Si * (1 - ReadxMoleFraction()/0.3) + C1Ge *
ReadxMoleFraction()/0.3 + Cc* (1 - ReadxMoleFraction()/0.3) *
ReadxMoleFraction()/0.3;
  dtaudct = tw1 * exp (C1 * pow (ct/T0, 2)+ C2 * (ct/T0) + C3 *
(Tl/T0)) * (2 * C1 * (ct/T0)+ C2)/ T0;
}



class Analytical_e_EnergyRelaxationTime : public
Analytical_EnergyRelaxationTime {
public:
 Analytical_e_EnergyRelaxationTime (const PMI_Environment&
env);
 ~Analytical_e_EnergyRelaxationTime () {}
};

Analytical_e_EnergyRelaxationTime ::
Analytical_e_EnergyRelaxationTime (const PMI_Environment& env):
Analytical_EnergyRelaxationTime (env)
{ tw0Si=InitParameter ("tw0Si_e" , 0.39102e-12);
 tw0Ge=InitParameter ("tw0Ge_e" , 0.44944e-12);
 Ctau=InitParameter ("Ctau_e" , -0.05e-12);
 tw1=InitParameter ("tw1_e" , -0.14434e-12);
 C1Si=InitParameter ("C1Si_e" , 0.0013509);
 C1Ge=InitParameter ("C1Ge_e" , 0.0028);
 Cc=InitParameter ("Cc_e" , -0.00181);
 C2=InitParameter ("C2_e" , -0.059019);
 C3=InitParameter ("C3_e" , 0.010688);
}

class Analytical_h_EnergyRelaxationTime : public
Analytical_EnergyRelaxationTime {
public:
 Analytical_h_EnergyRelaxationTime (const PMI_Environment&
env);
 ~Analytical_h_EnergyRelaxationTime () {}
};
```

```
Analytical_h_EnergyRelaxationTime ::
Analytical_h_EnergyRelaxationTime (const PMI_Environment& env):
Analytical_EnergyRelaxationTime (env)
{ tw0Si=InitParameter ("tw0Si_h" , 0.29e-12);
  tw0Ge=InitParameter ("tw0Ge_h" , 0.88e-12);
  Ctau=InitParameter ("Ctau_h" , 0);
  tw1=InitParameter ("tw1_h" , 0);
  C1Si=InitParameter ("C1Si_h" , 0);
  C1Ge=InitParameter ("C1Ge_h" , 0);
  Cc=InitParameter ("Cc_h" , -0.00181);
  C2=InitParameter ("C2_h" , 0);
  C3=InitParameter ("C3_h" , 0);
}

extern "C"
PMI_EnergyRelaxationTime* new_PMI_e_EnergyRelaxationTime
  (const PMI_Environment& env)
{ return new Analytical_e_EnergyRelaxationTime (env);
}

extern "C"
PMI_EnergyRelaxationTime* new_PMI_h_EnergyRelaxationTime
  (const PMI_Environment& env)
{ return new Analytical_h_EnergyRelaxationTime (env);
}
```

4.5 References

[1] Synopsys TCAD Software, Release 2007.03, Synopsys, Montain View, CA, 2007.

[2] F. M. Bufler, “Full-Band Monte Carlo Simulation of Electrons and Holes in Strained Si and SiGe”, München: Herbert Utz Verlag, 1998.

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

[4] C. G. Van de Walle, “Band lineups and deformation potentials in the model-solid theory”, *Physical Review B*, vol. 39, no. 3, pp. 1871-1883, March 1989.

[5] 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.T. Nicolson, K.H. K. Yau, N. Revil, D. Gloria, D. Dutartre, S. P. Voinigescu, and A. Chantre, “0.13 µm SiGe BiCMOS Technology Fully Dedicated to mm-Wave Applications”, *IEEE Journal of solid-state circuits*, vol. 44, no. 9, pp. 2312-2321, September 2009.

Chapter 5

Avalanche multiplication measurements and modeling

As demonstrated by recent published results [1], the aggressive technology development effort of the DOTFIVE project [2] has been significantly advancing the performance of SiGe:C HBTs towards the Terahertz range. Since the breakdown voltage introduces a trade-off with high frequency figures-of-merit in SiGe HBTs, the definition of Safe Operating Area (SOA) limits is a pivotal issue for designers and technologists. SOA limits specification is not a trivial task, as the maximum usable output voltages and currents depend on the driving conditions at the input port [3]. An additional difficulty is related to the fact that impact ionization mechanisms concur and interact with other limiting mechanisms, i.e. self-heating and hot-carrier degradation.

A commonly quoted parameter used to define the operating limits of the bipolar transistor is the open-base breakdown voltage BV_{CEO}. However, it has been noted that the maximum operating voltage in practical circuit configurations is actually larger than BV_{CEO} [4], see Chap. 1. In fact, common circuit configurations are not characterized by a large impedance on the base, so that SiGe HBTs can be safely biased above open-base breakdown voltage. Practical circuit topologies using a common-base configuration include cascode stages, output buffers and differential pairs. Unfortunately, when the device is biased above BV_{CEO} the operation is more complex, since avalanche multiplication effects cause a base current reversal. This in turn results in a distributed ohmic drop across the base region yielding a current focusing in the centre of the base, known as the pinch-in effect [4-6]. Therefore, in order to predict the SOA boundaries encountered in realistic circuit applications, an accurate model for the pinch-in effect is required. To this aim, an accurate model for multiplication factor (M) is required [7]. In this section we present an accurate calibration of avalanche generation models for device simulation and a review of main models for M in SiGe HBTs. Next, we propose a new complete model for M, which provides a good accuracy for a wide range of

values of the collector voltage, as required for a correct description of the pinch-in effect. The new model, combined with the analytical model for the base current-dependent base resistance [5], is suitable for being incorporated into HBT compact models to properly describe device operation above BV_{CEO}.

5.1 Device simulation of the avalanche multiplication

5.1.1 Simulation approaches

Several approaches are possible to include impact ionization phenomena in device simulation, namely:

- Newton approach.
- Gummel approach.
- Equivalent collector current control

Newton's method is a coupled procedure which solves the semiconductor equations (ref semicond. equs) simultaneously, through a generalization of the Newton-Raphson method for determining the roots of an equation [8]. Gummel's method [8], [9] solves the coupled set of semiconductor equations together with the Poisson equation via a decoupled procedure (see [10] for details on both methods). In general, the Gummel's method is preferred at low bias because of its faster convergence and low cost for iteration. At medium and high bias the Newton's method becomes more convenient, since the convergence rate of Gummel's method becomes worse as the coupling between equations becomes stronger at higher bias.

When impact ionization phenomena areincludedin device simulation, Gummel's method becomes unfeasible because convergence is too weak. Newton's approach works better, but convergence is lost close to the breakdown point (i.e. the point where the output conductance reaches infinity) or before. Therefore, we must resort to the equivalent collector current control method by introducing a resistor at the collector node. Using this method we can perform the simulation of the output characteristics of a bipolar

transistor where snapback behaviour occurs. For a collector voltage lower than the snapback voltage the simulation results are identical to those obtained by using until voltage control. However, the current control method is able to go over the breakdown point and describe the portion of the output characteristic with a negative conductance. Moreover, by choosing a proper value for collector resistance, the simulation time is reasonable and not longer compared to a simulation performed with a voltage control. Therefore this last technique has been chosen for breakdown simulations.

5.1.2 Avalanche generation models for HD simulations

Electron–hole pair production due to avalanche generation (impact ionization) requires a certain threshold field strength and the possibility of acceleration, that is, wide space charge regions. If the width of a space charge region is greater than the mean free path between two ionizing impacts, charge multiplication occurs, which can cause electrical breakdown. The reciprocal of the mean free path is called the ionization coefficient. The generation rate is expressed in device simulators as

$$G^{ii} = \alpha_n n v_n + \alpha_p p v_p \tag{5.1}$$

where α_n and α_p are the ionization coefficients for electrons and holes.

The ionization coefficients strongly depend on the field strength. Several models are available in *Sdevice*, [11], for both DD and HD simulations, but only the following ones have been considered for HD simulations:

-Van Overstraeten – de Man model [12].

-Okuto-Crowell model [13].

-Lackner model [14].

Van Overstraeten model is based on the Chynoweth law [15]:

$$\alpha\left(F_{ava}\right) = \gamma a \exp\left(-\frac{\gamma b}{F_{ava}}\right) \tag{5.2}$$

with:

$$\gamma = \frac{\tanh\left(\frac{\hbar\omega_{op}}{2kT_0}\right)}{\tanh\left(\frac{\hbar\omega_{op}}{2kT}\right)} \tag{5.3}$$

The coefficients a, b, and $\hbar\omega_{op}$, have been measured by van Overstraeten and de Man [12]. This model, however, does not work at all for HD simulations, since convergence is lost in the starting point, i.e. $V_C = 0$ V.

Lackner [14] derived a pseudo-local ionization rate in the form of a modification to the Chynoweth law, assuming stationary conditions. The temperature-dependent factor was introduced to the original model:

$$\alpha\left(F_{ava}\right) = \frac{\gamma a}{Z}\exp\left(-\frac{\gamma b}{F_{ava}}\right) \tag{5.4}$$

with:

$$Z = 1 + \frac{\gamma b_n}{F_{ava}}\exp\left(-\frac{\gamma b_n}{F_{ava}}\right) + \frac{\gamma b_p}{F_{ava}}\exp\left(-\frac{\gamma b_p}{F_{ava}}\right) \tag{5.5}$$

and:

$$\gamma = \frac{\tanh\left(\frac{\hbar\omega_{op}}{2kT_0}\right)}{\tanh\left(\frac{\hbar\omega_{op}}{2kT}\right)} \tag{5.6}$$

Device simulations were performed to compared the Okuto and Lackner. It was found that the simulation time for the Lackner model is almost twice higher compared to the Okuto model. For this reason the Okuto model has been chosen for calibration.

5.1.3 Okuto-Crowell model calibration

MC simulation results have been provided by *Bundeswehr University* for the 100 GHz reference structure already described in §4.3.1. In these simulations the multiplication factor, $\xi = M$-1 is extracted from base current and collector current versus collector-base voltage at a fixed emitter-base bias, using (5.7)

$$\begin{gathered} \xi(V_{CB}) = \frac{\Delta I_B(V_{CB})}{I_C(V_{CB}) - \Delta I_B(V_{CB})} \quad , \\ \Delta I_B(V_{CB}) = I_B(0) - I_B(V_{CB}) \end{gathered} \tag{5.7}$$

Equation (5.7) is derived under the assumption that the variation of the base current is only due to impact ionization (II). In [16] it was suggested to improve the accuracy of this method by including the Early effect in the base current equation. This can be done, by computing variations of the base current due to base width modulation and substituting $I_B(0)$ in (5.7) with a base current linearly dependent on collector bias. The slope of the I_B vs. V_{CB} dependence should be extracted at low V_{CB} (where II is negligible). However, the extraction of the base current slope failed for the simulated device because it appears independent from collector bias in the analyzed range, V_{BC} = 0 ÷ 0.5 V. As a matter of fact, the Early effect influence on (5.7) is negligible for the reference structure.

Equation (5.7) is valid at low collector current densities, where high injection effects are negligible. ξ is the ratio of the generated electron-hole pairs to the injected current and M is the multiplication factor. ξ equals M-1 only at low current densities, when ΔI_C and ΔI_B are comparable:

$$\begin{gathered} \xi(V_{CB}) = \frac{\Delta I_B(V_{CB})}{I_C(V_{CB}) - \Delta I_B(V_{CB})} \\ \Updownarrow \Delta I_B = \Delta I_C \\ \xi(V_{CB}) = \frac{\Delta I_C(V_{CB})}{I_C(V_{CB}) - \Delta I_C(V_{CB})} = \frac{I_C(V_{CB}) - I_C(0)}{I_C(0)} = M_n - 1 \end{gathered} \tag{5.8}$$

Moreover, it was found that the contribution of secondary holes to the total multiplication is more than two orders of magnitude lower than the contribution of primary electrons [17] therefore hole II calibration has been neglected.

The Okuto-Crowell empirical model [13] for avalanche generation is given by

$$\alpha\left(F_{ava}\right)=a\cdot\left(1+c\left(T-T_0\right)\right)F_{ava}^{\gamma}\exp\left[-\left(\frac{b\left[1+dT-T_0\right]}{F_{ava}}\right)^{\delta}\right] \quad (5.9)$$

In (5.9) T_0 is the reference temperature, fixed at 300 K, T is the lattice temperature and a, b, c, d, γ and δ are user-adjustable parameters, whose default values are reported in Table 5.1.

If a local carrier temperature-dependent impact ionization model is selected, the driving force F_{ava} equals an effective field E^{eff} calculated from the carrier temperature using the conversion formulas

$$nv_{sat,n}E_n^{eff}=n\frac{3k}{2q}\frac{T_n-T}{\lambda_n\tau_{en}}+\frac{\Upsilon_n}{q}\left(E_g+\delta_n kT_n\right)\alpha_n nv_{sat,n} \quad (5.10)$$

$$pv_{sat,p}E_p^{eff}=p\frac{3k}{2q}\frac{T_p-T}{\lambda_p\tau_{ep}}+\frac{\Upsilon_p}{q}\left(E_g+\delta_p kT_p\right)\alpha_p pv_{sat,p} \quad (5.11)$$

where the default parameters' values for silicon are reported in Table 5.2.

Table 5.1. Default coefficients for Okuto-Crowell model.

Parameter	Electrons	Holes
a [V^{-1}]	0.426	0.243
b [V/cm]	$4.81\cdot10^{5}$	$6.53\cdot10^{5}$
c [K^{-1}]	$3.05\cdot10^{-4}$	$5.35\cdot10^{-4}$
d [K^{-1}]	$6.86\cdot10^{-4}$	$5.67\cdot10^{-4}$
γ	1	1
δ	2	2

Table 5.2. Hydrodynamic avalanche model: default parameters for silicon.

Parameter	Default
λ_n	1
λ_p	1
Y_n	1
Y_p	1
δ_n	1.5
δ_p	1.5

Based on MC data for V_{BE} = 0.8, parameter calibration was performed. Since isothermal simulations have been performed, parameters *c* and *d* in (5.9) have no influence on results. An inverse calibration procedure has been carried out: parameters *a* and *b* have been changed in order to get the best fit between ξ, as computed by using (5.7) from terminal quantities and MC data. The results of this parameter extraction procedure are displayed in Fig. 5.1 and the corresponding parameter values are in Table 5.3. It can be noted that parameters for holes were scaled according to the values for electrons, keeping constant at the default value in Table 5.1 the ratio between them.

Table 5.3. Calibrated coefficients for Okuto-Crowell model.

Parameter	Electrons	Holes
a [V^{-1}]	0.39	0.222
b [V/cm]	$4.4 \cdot 10^5$	$5.95 \cdot 10^5$
c [K^{-1}]	$3.05 \cdot 10^{-4}$	$5.35 \cdot 10^{-4}$
d [K^{-1}]	$6.86 \cdot 10^{-4}$	$5.67 \cdot 10^{-4}$
γ	1	1
δ	2	2

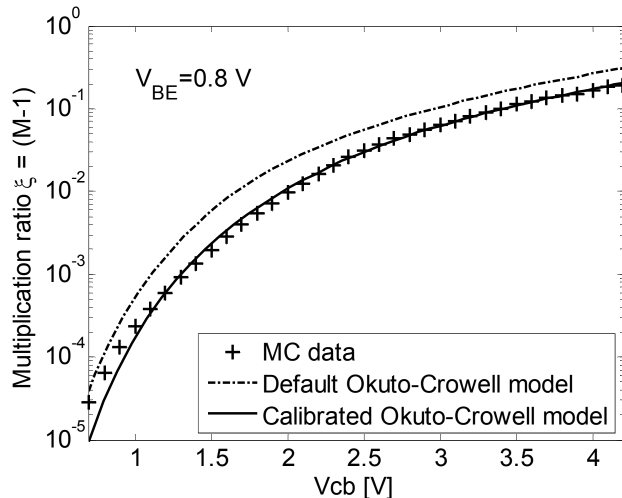


Fig. 5.1. Multiplication ratio vs. collector voltage; comparison between MC data and simulation results for $V_{BE} = 0.8$ V.

5.2 Avalanche multiplication factor model

5.2.1 Main models review

Several formulations for the multiplication factor have been proposed in the literature (e.g. [18-21]). These models retain a reasonable accuracy only for a limited range of the collector voltage. On the other hand, a correct description of the pinch-in effect [4] requires a model valid for a wide range of V_{CB} values ($>BV_{CEO}$), possibly close to the open-emitter breakdown voltage BV_{CBO}. Multiplication factor values have been generated using a MC simulator by Bundeswehr University for the 100 GHz reference structure described in §2.2, whose BV_{CBO} is about 12.5 V. Thus, using MC data, a comparison between available multiplication factor models has been carried out.

The impact-ionization multiplication coefficient, defined as the ratio of electron-hole pairs generated via impact ionization to the electron current entering the base-collector space-charge region, is obtain from [22] as

$$M-1=\xi=\frac{\Delta I_B}{I_C-\Delta I_B}=\frac{I_B\left(V_{CB}=0\right)-I_B\left(V_{CB}\right)}{I_C-\left[I_B\left(V_{CB}=0\right)-I_B\left(V_{CB}\right)\right]} \tag{5.12}$$

In the simplest form, it can be expressed using Miller's empirical model [19]:

$$M=\frac{1}{1-\left(V_{CB}/BV_{CBO}\right)^N} \tag{5.13}$$

where BV_{CBO} denotes the collector-base breakdown voltage under open-emitter conditions and represents the base-collector junction breakdown voltage. Miller formulation is very simple and includes only one fitting parameter, *N*, determining the slope of the double-logarithmic plot of 1-1/*M* versus V_{CB} (see Fig. 5.2). Since depicted data are not on a straight line, Miller formula cannot provide a reasonable accuracy over a wide V_{CB} range. For this reason, Reisch [20] proposed a "double Miller model"

$$M=\frac{1}{1-\dfrac{1}{\left(f_1\left(V-V_{CR}\right)\left(\dfrac{BV_1}{\left|V_{CB}\right|+\delta}\right)^{N_1}+f_2\left(V_{CR}-V\right)\left(\dfrac{BV_2}{\left|V_{CB}\right|+\delta}\right)^{N_2}\right)}} \tag{5.14}$$

This formulation combines two limiting formulas similar to (5.13) by suitable weighting factors f_1 and f_2, and allowing two independent fittings for low and high current levels. The weighting factors are expressed as

$$f_{1,2}\left(V-V_{CR}\right)=\frac{1}{1+\alpha\exp\left[\beta_{1,2}\left(V-V_{CR}\right)\right]} \tag{5.15}$$

In (5.14) δ is a small positive constant (e. g. $\delta = 10^{-8}$) introduced to avoid singularities in the case $V_{CB} = 0$. The value of α is determined so as to obtain the correct value of $1-1/M$ at $V_{CB} = V_{CR}$:

$$\alpha = 2\left(1-\frac{1}{M}\right)\left(\frac{BV_1}{V_{CR}}\right)^{N_1} - 1 \tag{5.16}$$

where the crossover voltage V_{CR} is given by

$$V_{CR} = \frac{BV_1^{N_1/(N_1-N_2)}}{BV_2^{N_2/(N_1-N_2)}} \tag{5.17}$$

A comparison between MC data and (optimized) Miller and Reisch models is depicted in Fig. 5.2 and Fig. 5.3. These results indicate that, although the Reisch model show a better accuracy compared to the Miller model, it is not suitable for SiGe HBTs, despite all the involved parameters.

The model proposed by Rickelt and Rein [18] provides acceptable results for weak avalanche [see Fig. 5.4 and Fig. 5.5]:

$$M = 1 + c_1 \cdot \left(V_{CB}/V\right)^{m_1} + c_2 \cdot \left(V_{CB}/V\right)^{m_2} \tag{5.18}$$

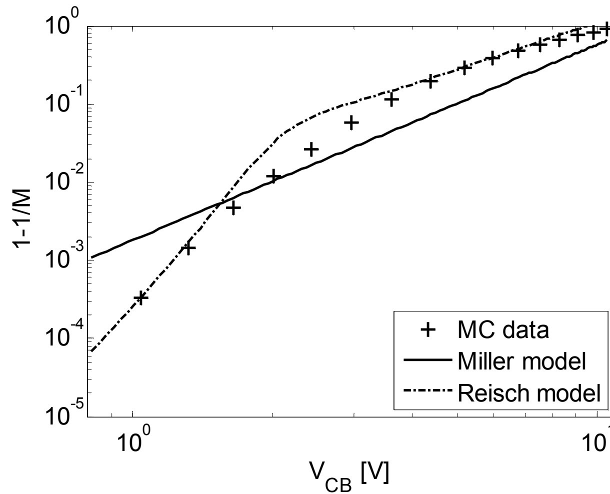


Fig. 5.2. Parameters extraction of the carrier multiplication factor M for Miller and Reisch models by fitting MC simulation data, BV_{CBO} = 12.5 V.

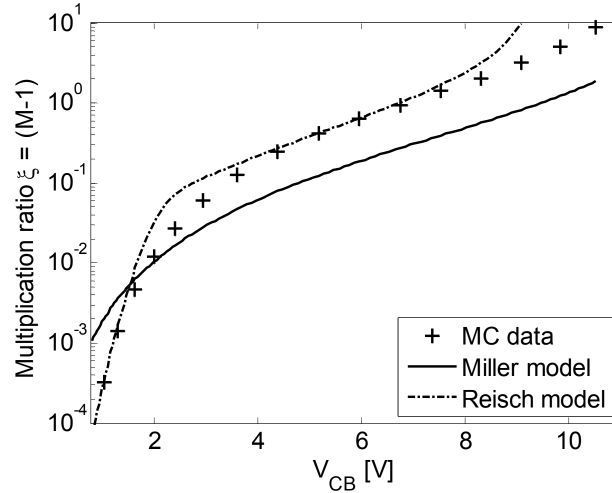


Fig. 5.3. Avalanche multiplication ratio for Miller and Reisch models over a wide V_{CB} range (BV_{CBO} = 12.5 V).

However, in this formulation the multiplication factor does not tend to infinity as V_{CB} approaches the breakdown voltage BV_{CBO} and does not include high current effects, (5.18).

A different approach has been proposed in [21]. The multiplication factor is given by

$$M = 1 + m \cdot \tan\left[f_I \cdot \left(\frac{V_{CB}}{f_T \cdot BV_{CBO}} \right)^n \cdot \frac{\pi}{2} \right] \tag{5.19}$$

where

$$f_I = \exp\left(-\frac{I_C}{I_{T0}} \right) \tag{5.20}$$

and

$$f_T = \exp\left(\alpha_{BV} \Delta T \right) \tag{5.21}$$

This formulation is more accurate for high V_{CB} values and accounts for the current dependence of the multiplication factor [see (5.20)], whereas it suffers from a lack of accuracy at low V_{CB} values, as shown in Fig. 5.4 and Fig. 5.5. The empirical correction factor f_I (5.20) introduces an additional fitting parameter and allows to account for the decrease of M at high-current levels due to the ohmic drop across the collector resistance and the Kirk effect. The correction

parameter f_T is introduced to account for the temperature dependence of the breakdown voltage BV_{CBO}. Since this effect was estimated to be negligible for the devices considered in this study, which are essentially thermally limited, we assumed $\alpha_{BV} \approx 0$ (i.e., $f_T = 1$). For devices with a lower thermal resistance, this effect may be quite significant and should be taken into account.

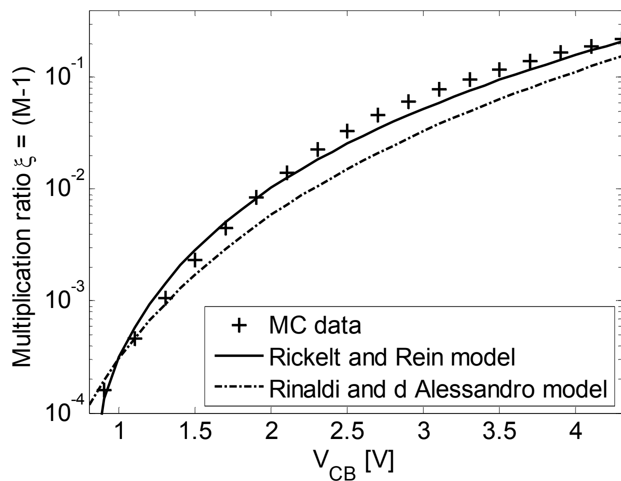


Fig. 5.4. Avalanche multiplication ratio over a wide V_{CB} range (BV_{CBO} = 12.5 V): comparison between models and MC results.

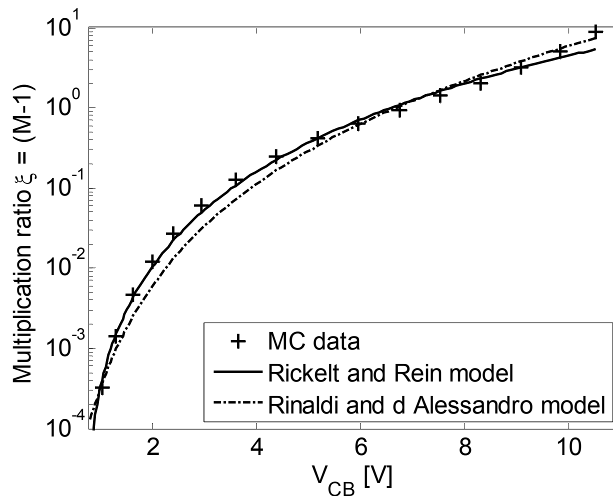


Fig. 5.5. Avalanche multiplication ratio for weak avalanche (BV_{CBO} = 12.5 V): comparison between models and MC results.

5.2.2 A new multiplication factor model

Bearing in mind models limitations emphasized in §5.2.1, a novel analytical formulation has been developed. Model parameters were calibrated by comparison with MC data. The model was successfully verified by measurements, allows an accurate fitting for low and high V_{CB} values. The multiplication factor is given by

$$M = 1 + \frac{a \cdot x}{1 - x} \exp\left(-b \cdot x^{c}\right) \tag{5.22}$$

where

$$x = f_I \cdot \left(\frac{V_{CB}}{BV_{CBO}} \right) \tag{5.23}$$

The new formulation has been developed starting from weak avalanche models available in industrial compact model for high frequency bipolar devices, HICUM [23] and MEXTRAM [24], which have been extended in order to guarantee an infinite value for M as V_{CB} approaches the breakdown voltage BV_{CBO}. In addition, the model can include the collector current dependence by adding the parameter f_I which is defined as

$$f_I = \exp\left(-\frac{I_C}{I_{T0}} \right) \tag{5.24}$$

In Figs. 5.6 ÷ 5.9 the new model results are compared with MC simulation data for the 100 GHz SiGe HBT reference structure and with optimized results obtained from previous formulations described in §5.2.1 for a narrow and a wide V_{CB} range. As can be seen, the new model provides an accurate fit despite the limited number of parameters.

The model proposed in [18] shows a maximum percentage error of about 40%, while the new model with optimized parameters reduces the maximum percentage error below 10%.

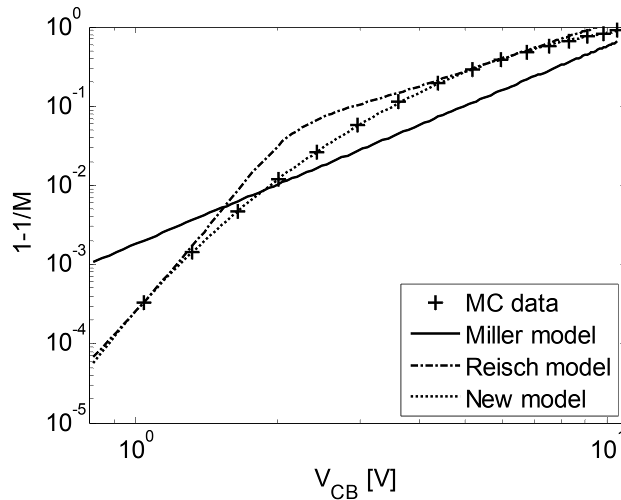


Fig. 5.6. Parameters extraction of the carrier multiplication factor *M* for Miller, Reisch and proposed model by fitting MC simulation data, BV_{CBO} = 12.5 V

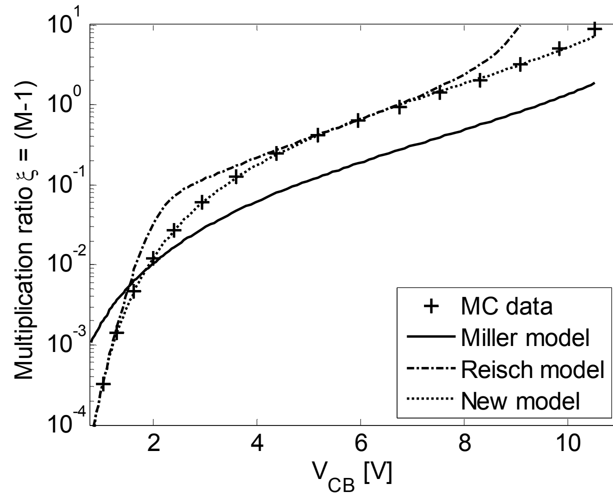


Fig. 5.7. Avalanche multiplication ratio over a wide V_{CB} range (BV_{CBO} = 12.5 V): comparison of Miller, Reisch and proposed model with MC results.

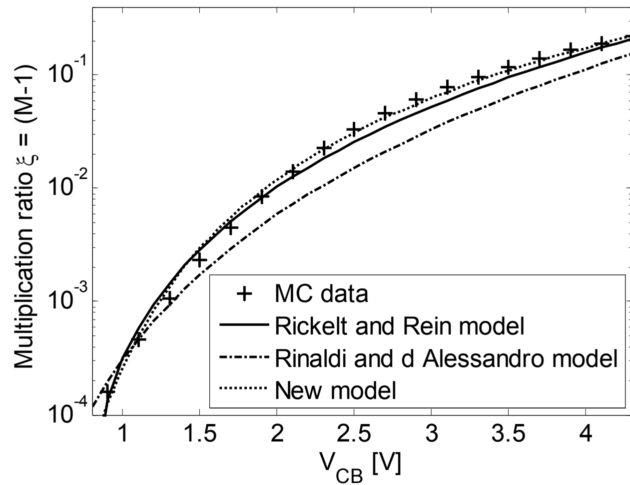


Fig. 5.8. Avalanche multiplication ratio for weak avalanche (BV_{CBO} = 12.5 V): comparison between models and MC results.

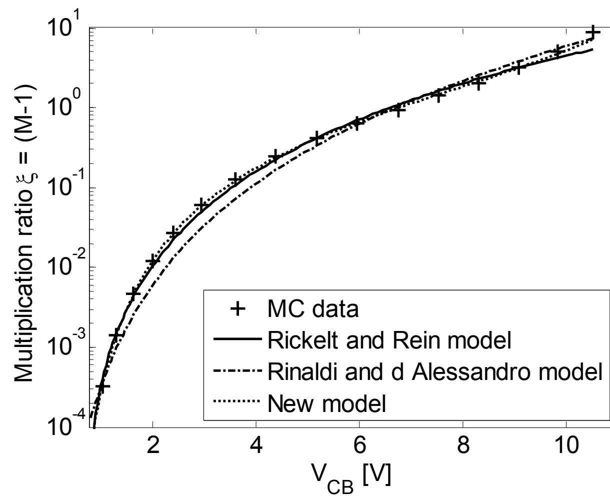


Fig. 5.9. Avalanche multiplication ratio over a wide V_{CB} range (BV_{CBO} = 12.5 V): comparison between models and MC results.

5.2.3 Measurements and model parameters' extraction

Parameter extraction has been also performed using experimental data for two different technologies. The first technology is an IBM's 300-GHz SiGe HBT process. For this technology the measured multiplication factor data are published in [25]. The second technology is a ST double-polysilicon FSA-SEG SiGe:C HBT process with f_{MAX} of about 300 GHz [1]. The multiplication factor was extracted from measurements by means of equation (5.12). The collector current was measured as a function of the collector voltage with a grounded base and a fixed V_{EB}, as described in [19]. Model results are depicted in Fig. 5.10. The optimized model parameters values are given in Table 5.4, together with the relevant parameters of the corresponding process. These results indicate that the new proposed model provides accurate results for different technological nodes. Moreover, parameters values' regularity with BV_{CBO} assures the reliability and accuracy of the proposed multiplication factor model.

Table 5.4. HBT Parameters Comparison.

Parameter	MC results	IBM	ST
BV_{CBO} [V]	12.5	6	5.5
f_T [GHz]	100	250	230
f_{MAX} [GHz]	-	315	290
a	31.9	79.92	118.05
b	2.34	4.18	4.3
c	-0.54	-0.47	-0.46

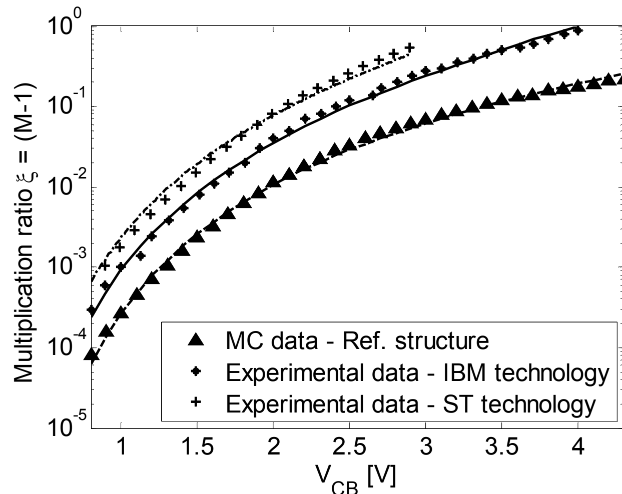


Fig. 5.10. Avalanche multiplication ratio as a function of V_{CB} at low V_{CB} values. Comparison of proposed model (lines) with experimental data and MC results (symbols).

5.3 References

[1] P. Chevalier , F. Pourchon, T. Lacave, G. Avenier, Y. Campidelli, L. Depoyan, G. Troillard, M. Buczko, D. Gloria, D. Céli, C. Gaquière, and A. Chantre, "A conventional Double-Polysilicon FSA-SEG Si/SiGe:C HBT Reaching 400 GHz fmax," *BIPOLAR/BiCMOS Circuits and Technology Meeting Proceedings*, Capri, Italy, pp. 1-4, October 2009.

[2] http://www.dotfive.eu/

[3] J. D. Cressler, "Emerging SiGe HBT Reliability Issues for Mixed-Signal Circuit Applications," *IEEE Transactions on Device and Materials Reliability*, vol. 4, no. 2, pp. 222-236, June 2004.

[4] C. M. Grens, J. D. Cressler, and A. J. Joseph, "On Common-Base Avalanche Instabilities in SiGe HBTs", *IEEE Transactions on Electron Devices*, vol. 55, no. 6, pp. 1276-1285, June 2008.

[5] M. Costagliola, and N. Rinaldi, "Theoretical analysis and modeling of bipolar transistor operation under reversal base current conditions", *BIPOLAR/BiCMOS Circuits and Technology Meeting Proceedings*, Capri, Italy, pp. 25-28, October 2009.

[6] M. Rickelt, and H. M. Rein, "A Novel Transistor Model for Simulating Avalanche-Breakdown Effects in Si Bipolar Circuits", *IEEE Journal of Solid-State Circuits*, vol. 37, no. 9, pp. 1184-1197, September 2002.

[7] G. Sasso, M. Costagliola, and N. Rinaldi, "Avalanche multiplication and pinch-in models for simulating electrical instability effects in SiGe HBTs", *Microelectronics Reliability*, Volume 50, Issues 9-11, pp. 1577-1580, September 2010.

[8] S. Selberherr, Simulation of Semiconductor Devices and Processes, Springer-Verlag Wien, New York.

[9] H. K. Gummel, "A self-consistent iterative scheme for one-dimensional steady state transistor calculation," *IEEE Transactions on Electron Devices*, vol. 11, no. 10, pp. 455-465, October 1964.

[10] D. Vasileska, and S. M. Goodnick, Computational Electronics, Morgan & Claypool, 2006.

[11] Synopsys TCAD Software, Release 2007.03, Synopsys, Mountain View, CA, 2007.

[12] R. van Overstraeten, and H. de Man, "Measurement of the Ionization Rates in Diffused Silicon p-n Junctions," *IEEE Journal of Solid-State Electronics*, vol. 13, no. 1, pp. 583-608, January 1970.

[13] Y. Okuto, and C. R. Crowell, "Threshold Energy Effect on Avalanche Breakdown Voltage in Semiconductor Junctions," *IEEE Journal of Solid-State Electronics*, vol. 18, no. 2, pp. 161-168, February 1975.

[14] T. Lackner, "Avalanche Multiplication in Semiconductors: A Modification of Chynoweth's Law," *IEEE Journal of Solid-State Electronics*, vol. 34, no. 1, pp. 33-42, January 1991.

[15] A. G. Chynoweth, "Ionization Rates for Electrons and Holes in Silicon," *Physical Review*, vol. 109, no. 5, pp. 1537–1540, March 1958.

[16] G. Niu, J. D. Cressler, S. Zhang, U. Gogineni, and D. C. Ahlgren, "Measurement of Collector-Base Junction Avalanche Multiplication Effects in Advanced UHV/CVD SiGe HBT's", *IEEE Transactions on Electron Devices*, vol. 46, no. 5, pp. 1007-1015, May 1999.

[17] W. Maes, K. De Meyer, and R. Van Overstraeten, "Impact Ionization In Silicon: a review and update", *IEEE Journal of Solid-State Electronics*, vol. 33, no. 6, pp. 705-718, June 1990.

[18] M. Rickelt, and H.M. Rein, "A Novel Transistor Model for Simulating Avalanche-Breakdown Effects in Si Bipolar Circuits," *IEEE Journal of Solid-State Electronics*, vol. 37, no. 9, pp. 1184-1197, September 2002.

[19] S. L. Miller, "Ionization Rates for Holes and Electrons in Silicon," *Physical Review*, vol. 105, no. 4, pp. 1246-1249, February 1957.

[20] M. Reisch, High-frequency Bipolar Transistors: Physics, Modelling, Applications. Springer, 2003.

[21] N. Rinaldi, and V. d'Alessandro, "Theory of Electrothermal Behavior of Bipolar Transistors: Part III—Impact Ionization," *IEEE Transactions on Electron Devices*, vol. 53, no. 7, pp. 1683-1697, July 2006.

[22] P.F. Lu, and T.C. Chen, "Collector-base Junction Avalanche Effects in Advanced Double-poly Self-aligned Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 36, no. 6, pp.1182-1188, June 1989.

[23] M. Schroeter, Z. Yan, T.Y. Lee, and W. Shi, "A Compact Tunneling Current and Collector Breakdown model," *BIPOLAR/BiCMOS Circuits and Technology Meeting Proceedings*, pp. 203-206, 1998.

[24] W.J. Kloosterman, and H.C. De Graaff, "Avalanche Multiplication in a Compact Bipolar Transistor Model for Circuit Simulation", *IEEE Transactions on Electron Devices*, vol. 36, no. 7, pp. 1376-1380, July 1989.

[25] 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, September 2005.

Conclusions and outlook

The basic purpose of this thesis was to set capabilities and limits of HD models and to develop physical models for transport parameters, in order to allow reliable and predictive HD simulation of advanced SiGe heterostructure devices in commercial TCADs.

The issues related to device simulation with HD models have been studied over different technological nodes. The analysis highlights that the influence of each parameter is similar for different technologies. For different scaled devices the unphysical effects in the output characteristics can be removed by using optimizing parameters, and the best results are achieved with the Blotekjær model. Therefore, the maximum cut-off frequency overestimation of HD models can be strongly mitigated for frequencies under the terahertz range. However, the optimization must be repeated for each different technological node. Since the overestimation of the diffusive component of the energy density increases with scaling, HD models are too inaccurate for modeling carrier transport in devices within the terahertz range. For HBTs with a maximum cut-off frequency above 500 GHz, HD simulation becomes too inaccurate and more reliable approaches, such as Monte Carlo and Spherical Harmonic Expansion are needed.

However, analytical transport models for DD/HD simulation available in the literature refer to silicon, and do not include the dependence on all relevant parameters, in particular germanium content. Therefore, based on the MC data, a full set of transport parameters have been generated and new analytical models have been developed which include the dependence upon all relevant quantities. ; In addition, new models have been compared with state-of-art models and experimental results.

Novel transport models have been successfully implemented in a commercial device simulator and hydrodynamic simulation results have been reported for several one-dimensional and two-dimensional structures with a different f_T maximum. HD simulation results have been compared to results obtained using a standard set of models for silicon-germanium and to more reliable results (i.e. MC and SPRING simulation results and experimental data), validating proposed models

and clarifying their reliability and accuracy over different technologies.

Finally, after an accurate calibration of model by Okuto and Crowell for avalanche generation in HD device simulation and a review of main models for multiplication factor (M) in SiGe HBTs, a new complete model for M has been proposed. New model provides a good accuracy for a wide range of values of the collector voltage. The new model is suitable for being incorporated into HBT compact models to properly describe device operation above BV_{CEO}.