Non-standard geometry scaling effects in high-frequency SiGe bipolar transistors

M. Schroter\textsuperscript{1,2}, S. Lehmann\textsuperscript{2}, D. Celi\textsuperscript{3}

\textsuperscript{1}ECE Dept., University of California San Diego, La Jolla, CA, USA
\textsuperscript{2}Chair for Electron Devices and Integrated Circuits, Dresden University of Technology, 01062 Dresden, Germany
\textsuperscript{3}STMicroelectronics, Crolles, France

Abstract: The impact of process and geometry effects on important electrical parameters of SiGe HBTs as a function of emitter width is investigated and explained using device simulations. The goal is to provide indications and guidelines for identifying especially those effects that cause non-standard geometry scaling, in order to be able to include them in the parameter extraction for and the generation of geometry-scalable physics-based compact models. Simple extensions of the standard scaling law are suggested, which are suitable for compact modeling. Experimental results exhibiting some of the effects will be shown as demonstration.

1 Introduction

The continuous strive for improved transistor performance leads to an overall increase in device and technology complexity. Not only the device structure itself becomes more sophisticated but often also additional physical effects occur that were negligible or not even existing in a previous generation. All these effects need to be described with sufficient accuracy by compact models in order to employ a developed process technology for production circuit design. Minimizing overall design cost generally requires circuit optimization for achieving the often stringent electrical specifications, which applies especially to RF and high-speed circuits (e.g. [1]).

For a given process technology and circuit topology the electrical performance can usually only be optimized by varying the device layout. RF circuits in particular are very sensitive to the transistor layout used. As a consequence, “geometry scalable” compact models are necessary; i.e. models in which each parameter can be quickly and accurately calculated as a function of the layout. Presently, various approaches exist for including geometry scalable models in design kits (e.g., [2][3]). However, common to all approaches is the use of a "standard" scaling law, which reads for the generic variable $X$ (e.g. current $I$, charge $Q$, capacitance $C$)

$$X = X_A A + X_P P + 4X_C.$$  \hspace{1cm} (1)

Here, $A$ and $P$ are the window area and perimeter of the relevant layer, respectively; $X_A$, $X_P$, $X_C$ represent the component value per area, per perimeter length and per corner, respectively, and are named “process specific” or just “specific” parameters. This standard scaling law has worked well for most junctions and compact model elements. For a structure with a window length $l_{EO}$ much larger than its width $b_{EO}$ (i.e. the 2D case) the corner component is negligible, and the specific parameters $X_A$ and $X_P$ can be easily determined as shown in Fig. 1. For short structures, it was shown in [4], that the corner component can be combined with $X_P$.

In recent SiGe HBT process generations though deviations from the standard scaling behavior (1) have increasingly been observed for certain characteristics such as transfer current, BE depletion capacitance, internal base sheet resistance, transit time, avalanche current. Examples for such “non-standard” scaling behavior are sketched in Fig. 1. The inability to describe the geometry dependence accurately by (1) can severely limit the geometry scaling capability of compact models and, thus, circuit optimization as well as predictive and statistical modeling.

Although it often turns out that non-standard scaling effects could be eliminated or at least significantly reduced by proper changes in the process flow, the problems are often either realized too late or considered to be too inconvenient or expensive to be fixed before qualification. In either case, the burden of making the process cost-efficiently usable for circuit design is then put on compact models. In order to provide compact models for the same layout variations that designers have been used to with a standard-scalable process, parameters now have to be extracted separately for many single-geometry transistors. This can become an extremely time consuming effort which may still not lead to the same layout flexibility and may even result in a loss of statistical modeling capability.

In this paper, a variety of process effects are investigated in terms of their impact on important electrical parameters of SiGe HBTs as a function of emitter width. First, the simulated device structures are introduced, followed by a presentation and explanation of the results. Next, the impact on compact modeling is discussed. Finally, selected experimental results are compared to simulations.

![Fig. 1: Visualization of the geometry scaling behavior: standard scaling according to (1) (solid line) and examples for observed "non-standard" scaling behavior (dashed/dotted line). Using the argument P/A instead of, e.g., width allows to include devices of any size.](image-url)
2 Device structures and investigated effects

For the investigations the simulator DEVICE [5] with a realistic doping profile from a SiGe production process [6] has been employed. The profile was converted to an analytical description in order to allow a flexible variation with emitter width. The nominal profile and certain physical parameters in DEVICE were calibrated based on results from literature, experimental characteristics, and DESSIS. Fig. 2 shows a comparison for the collector current (density). Similar results were obtained for other characteristics such as depletion capacitances and transit frequency. This establishes the basis for the subsequent investigations.

![Graph showing collector current density vs collector-emitter voltage](image)

**Fig. 2:** Comparison of collector current density (emitter window area $A_{E0} = 0.65\times12.65\text{mm}^2$) obtained from measurements (symbols, $V_{BC}/V = 0.5, 0, -1.241$), DESSIS (dashed lines, $V_{BC}/V = 0.4, 0, -1.717$), and DEVICE (solid lines, $V_{BC}/V = 0.4, 0.5, 0, -1.241, -1.717$).

Fig. 3 visualizes schematically the profile variations for the investigated cases characterized by the acronyms in the l.h.s. column of Table 1, which provides a brief description. The reference structure and profile (for each width) is indicated by REF.

The doping changes investigated here can result from a variety of effects and structural conditions during fabrication. EJD, EJI, and PED [7] can be caused by certain conditions for, e.g., emitter poly thickness and grain size, or the poly-mono interface. Furthermore, an extrinsic base implant close to the emitter window can cause TED [8,9]. The doping of the selectively implanted collector (SIC) and its lateral spread underneath the window may vary with the emitter window size [10,11]. In particular, scattering of implanted doping atoms already at layers above the surface can lead to a laterally non-equidistributed SIC doping concentration (cf. lower part of Fig. 3a).

There is certainly a variety of effects impacting geometry scaling, which can be roughly subdivided into (i) purely layout dependent effects with layout independent process specific parameters and (ii) effects resulting from layout dependent specific parameters. The first case can be described by a standard-scaling law such as (1), while the second case can lead to non-standard scaling behavior. Both cases can be further subdivided into emitter width and length dependent effects. For a systematic investigation and in order to separate the various effects, it is useful to start with the emitter width dependence by assuming sufficiently long devices ($l_{E0} >> b_{E0}$), which are also being employed during compact model parameter extraction. As a consequence, 2D device simulations are sufficient. The investigated structures have an emitter window widths of $b_{E0}/\mu m = 0.25, 0.45, 0.65, 1.05, 1.45$, and a unit emitter length $l_{E0} = 1\mu m$.

![Diagram showing investigated structures and conditions](image)

**Fig. 3:** Investigated structures and conditions: (a) contours showing the variation of the planar BE junction (EJD, EJI), the SIC extension with emitter width (CWI), and SIC doping change under the emitter due to implant scattering (CIS); (b) contours showing the effect of TED and edge diffusion (PED) on emitter edge and perimeter junction region; (c) SIC doping variation vs. emitter width (CDL, CDQ).

<table>
<thead>
<tr>
<th>label</th>
<th>meaning</th>
<th>fig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EJD</td>
<td>Emitter Junction depth Decrease with emitter width</td>
<td>3a</td>
</tr>
<tr>
<td>EJI</td>
<td>Emitter Junction depth Increase with E width</td>
<td>3a</td>
</tr>
<tr>
<td>TED</td>
<td>Transient-Enhanced-Diffusion (TED)-like effect</td>
<td>3b</td>
</tr>
<tr>
<td>PED</td>
<td>emitter Poly Edge Diffusion effect</td>
<td>3b</td>
</tr>
<tr>
<td>CIS</td>
<td>Collector Implant Scattering</td>
<td>3a</td>
</tr>
<tr>
<td>CDL</td>
<td>Collector implant Doping increase: Linear with $b_{E0}$</td>
<td>3c</td>
</tr>
<tr>
<td>CDQ</td>
<td>Collector implant Doping decrease below $b_{E0,REF}$</td>
<td>3c</td>
</tr>
<tr>
<td>CWI</td>
<td>Collector implant Width Increase with E width</td>
<td>3a</td>
</tr>
</tbody>
</table>

**Table 1:** Quick reference guide for the various investigated cases and their labels. In this paper, results for CDQ and CWI are not displayed, but only briefly discussed in the text.
3 Simulation results

The parameters selected here for comparison and discussion mostly correspond to typical process control monitors. The scaling behavior of the collector current \( I_C \) at low injection is shown in Fig. 4. Changes in collector doping (CDL, CDQ, CIS) and SIC width (CWI) have negligible influence here due to the high base doping concentration in SiGe HBTs. Thus a straight line very close to that of REF is obtained. PED also follows (1) but with a larger slope that indicates a higher injection due to the longer junction path at the perimeter; however, the area specific value \( I_{CA} \) is still the same as for the reference structure, since the perimeter profile influence disappears for large widths. This also holds for TED. Here, a slightly smaller electron injection results from the decreasing BE junction depth towards the emitter edge. Be \( I_{AE} < I_A \) the average transfer current injected across the edge region width \( \Delta b \) (cf. Fig. 3b), then the total current can be written as (2D case)

\[
I = I_A(b-\Delta b) + I_{AE}\Delta b + 2I_P ,
\]

which explains the smaller \( I_C \) values. At large enough widths \( b > \Delta b \), the area component \( I_A \) always exists, with a constant edge and perimeter component. Hence, a linear dependence on \( b \) results with the same extrapolated value \( I_A \) as REF. However, once \( b < \Delta b \) the planar junction under the window and the associated \( I_A \) component have completely disappeared. Also, \( I_{AE} \) now becomes a function of \( b \) so that the total current is

\[
I = I_A(b) + 2I_P .
\]

Writing the equations for the above two cases in standard form and using the generic variable \( X = (I, Q, C) \) gives for the 2D case

\[
X = \begin{cases} X_A + \left[ (X_{AE} - X_A) \frac{\Delta b}{2} + X_P \right]^2 , & b > \Delta b \\ X_{AE} + X_P^2 , & b \leq \Delta b \end{cases}
\]

Depending on whether \( X_{AE} \) is smaller or larger than \( X_A \) the slope at larger widths is smaller or larger than for REF.

The most pronounced deviation from standard scaling is observed for EJI. Here, the encroachment of the high emitter doping into a region of already significant base doping reduces the Gummel number for larger emitter widths which leads to an increase of \( I_C \) with width. The opposite case occurs for EJD. However, the nonlinearity is (much) less pronounced since in the (optimized) reference profile the emitter and base doping are already sufficiently separated (to reduce the BE depletion capacitance) leading to a smaller change in the Gummel number.

As expected, the BE profile peculiarities are much more visible in the BE depletion capacitance as shown in Fig. 5. Except for the changes in collector doping and width, which have no influence at all on \( C_{JE} \), more or less strongly nonlinear dependences are observed. The results for EJI behave similarly as for \( I_C \) since for larger widths the emitter junction is located at a larger base doping. The nonlinear trend is now more pronounced for the opposite case (EJD) since \( C_{JE} \) depends on (base) doping only and not on the Gummel number.

For PED, (3) can be used with \( X = C_{JE} \) and \( C_{JE,AE} > C_{JE,A} \) due to higher doping and a larger junction path at the perimeter. Hence, the slope is larger than for REF, but can slightly decrease towards narrow widths. The area specific value \( C_{JE,A} \) is the same as for REF.

Eq. (3) also applies to \( C_{JE} \) in case of TED. For SiGe HBT processes, the base doping towards the surface under the spacer is very small, so that the associated perimeter capacitance \( C_{JP} \) is fairly small and \( C_{JE,A} - C_{JE,P} < 0 \), leading even to a negative slope towards smaller structures as can be observed in Fig. 5. For sufficiently small widths, when the area related portion has disappeared, the ratio of the still constant perimeter portion to the now width dependent edge portion increases since the retracting junction depth is located at lower base doping concentrations. Thus, the slope decreases.
\[ C_{JC} = \frac{C_{JC,A}(b_{E0} + b_{SIC}) + 2C_{JCX,A}b_{BCx}}{b_{E0} + b_{SIC}} \]  

(4)

For CIS, deviations from REF are expected, but are - despite significant doping variation - fairly small as shown in Fig. 6. Since the result is still a straight line, the doping variation cannot be detected in practice (i.e. without knowing the reference line). In contrast, a nonlinear dependence is observed for CDL, clearly indicating a doping change with emitter width. Note, that CWI does not differ significantly from REF, which also holds for CDQ for \( b_{E0} > b_{E0,ref} \). Hence, they have been omitted in Fig. 6.

Constant specific parameters \( C_{JC,A} \) and \( C_{JCX,A} \) could also be determined from the standard scaling form

\[ \frac{C_{JC}}{b_{E0} + b_{SIC}} = \frac{C_{JC,A}}{b_{E0} + b_{SIC}} + 2C_{JCX,A} \]

(5)

In practice, the problem here is to determine \( b_{SIC} \) properly. If \( C_{JCX,A} \) is extracted from a large area test structure, then \( b_{SIC} \) and \( C_{JC,A} \) can be determined from (4) and be used later for calculating scalable model parameters. (Note, that the total BC junction width is known.)

![Fig. 6: Scaling behavior of the zero-bias BC depletion capacitance. EJJD, EJI, TED and PED have no impact.](image)

The scaling behavior of the zero-bias base resistance of a tetrode structure [12,13] is shown in Fig. 7. The width dependence is given by

\[ R_{BC} = r_{SBIO}b_{E0} + 2(r_{SBIO}r_{S} + R_{x}) \]  

(6)

Here, \( r_{SBIO} \) is the zero-bias internal base sheet resistance, \( r_{S} \) is the spacer (or base link) sheet resistance, \( b_{s} \) is the spacer width, and \( R_{x} \) contains the remaining contributions of the external base. Plotting \( R_{BC} \) vs. \( b_{E0} \) permits to extract \( r_{SBIO} \) from the slope and \( r_{S} \) from the intercept, since \( R_{x} \) can be determined from separate test structures. According to Fig. 7, the most visible deviation from a straight-line occurs for EJJD: the slope (i.e. \( r_{SBIO} \)) increases towards larger widths since the emitter junction depth penetrates the higher doped region. In the opposite case (EJJD) the deviation appears much smaller since higher values of \( r_{SBIO} \) now occur for narrow widths, where \( r_{SBIO} \) does not have much influence. Due to the high base doping concentration in SiGe HBTs variations in collector doping have negligible influence on \( R_{BC} \).

![Fig. 7: Scaling behavior of the zero-bias base resistance obtained from a tetrode structure [13] (2D simulation, \( l_{E0}=1\mu m \)).](image)

Geometry scaling of the low-injection minority charge \( Q_{f0} \) can be described by the sum of an (internal) area and a perimeter related component:

\[ Q_{f0} = Q_{f0i} + Q_{f0p} = Q_{f0iA} + Q_{f0iP}P_{E0} \]  

(7)

The area specific and perimeter length specific component are given by

\[ Q_{f0A} = \tau_{f0}I_{CA} \quad \text{and} \quad Q_{f0P} = \tau_{f0}I_{CP} \]  

(8)

with \( \tau_{f0} \) and \( \tau_{f00} \) as bottom and perimeter related storage time, which are the parameters to be extracted for a compact model. Writing (7) in the usual way results in the standard form

\[ \frac{Q_{f0}}{A_{E0}} = \frac{Q_{f0A}}{A_{E0}} + \frac{Q_{f0P}}{P_{E0}} \]  

(9)

with the general relation

\[ Q_{f0}(b_{E0}) = \tau_{f0}(b_{E0})I_{C}(b_{E0}) \]  

(10)

\( \tau_{f0} \) is the low-current transit time and can be determined from the transit frequency.

As shown in Fig. 8, the expected straight line is obtained for REF. Standard scaling behavior is also observed for variations in collector profile (CDL, CIS). Notice, that the variation of the charge with width (or \( P_{E0}/A_{E0} \)) consists of the contribution from the specific collector current component and from the storage time components.

For PED a straight line is obtained (with the same y-axis intercept \( Q_{f0A} \) as for REF), not allowing to detect the profile change but making it at least easy to model. The larger slope than REF results from a larger perimeter charge. The opposite is the case for TED, where the perimeter charge is smaller than in REF. Its relative contribution towards narrow widths increases and leads to a compensation of the negative slope.

For EJJD and EJJD, the change of the BE junction depth with width leads to a change of both the specific current components and storage time under the emitter and at the perimeter. Regional analysis shows that base and emitter storage time component changes al-
most compensate each other with junction depth variation.

Fig. 8: Scaling behavior of the low-injection minority charge.

4 Discussion

According to the results, the geometry dependence of certain electrical parameters still follows the standard scaling form (1) even though the doping profile varies with width. Therefore, those parameters do not allow to electrically detect such profile changes. They only lead to a different set of specific parameters by using the standard scaling law. However, by looking at several parameters simultaneously (e.g. $I_C$, $C_{JE0}$, $C_{CD0}$, $r_{SB0}$, BV-$CEO$, $Q_0$ (or peak $f_1$)) it is possible to detect a profile change if at least one of those parameters deviates from standard scaling. For instance, the depletion capacitances are significantly more sensitive to the corresponding doping changes than $I_C$ or $r_{SB0}$. The data and associated trends of $X(b)$ provided in this work can serve as a guideline for identifying the specific type of profile change with width.

For doping changes that lead to a nonlinear behavior of an electrical parameter the scaling equation has to be extended. A simple approach is to write

$$X = \frac{X_w}{(2/b)} + X_n + X_n^2 + X_n^3, \quad l \gg b.$$  \hfill (11)

$X_w$ and $X_n$ are additional (fitting) parameters, that represent a deviation from the standard scaling (i.e. the two middle terms) for wide (index “w”) and narrow (index “n”) devices. Usually only one of the additional parameters has to be determined. In order to be able both to fit the parameters sufficiently accurate and to distinguish a doping change related trend from variations due to measurement uncertainty, a sufficient number of different widths needs to be measured in any way. The results obtained by applying (11) to the $C_{JE0}$ data of devices with larger variations are shown as solid lines in Fig. 9 along with the parameter values. Excellent agreement is obtained in all cases using just the $X_w$ parameter.

The drawback of the proposed extension though is that (11) has to be applied to every specific parameter that is affected by the same width dependent profile change. A more efficient and physics-based approach would be to apply (11) only to those technology parameters (TPs), such as base or collector doping, that are also used for predictive and statistical modeling (cf. [14]). This way, all specific parameters related to the particular TP would automatically become width dependent if the proper relations have been established as described in e.g. [14]. This approach requires though an identification of the affected TPs from the observed scaling behavior.

5 Experimental results

An example encountered for measured transistors was the observed decrease of the avalanche current $I_{aval}$ (and multiplication factor $M$) associated with a decrease of peak $f_1$ along with the critical current density characterizing the $f_1$ drop with larger emitter widths. Such behavior is not compatible with standard scaling. In contrast, the associated $C_{JE0}$ data still follow the standard scaling, so that this parameter cannot be used for detecting the cause of the behavior.

Figs. 10 and 11 show the measured data along with selected simulated cases. As it turns from evaluating the geometry dependence of the complete set of simulated parameters, only for CIS the trends of measured and simulated data agree, strongly pointing towards a collector doping change with width. The difference in absolute values of peak $f_1$ between measured and CIS data in Fig. 10 results from the lack of accurate information on the width dependence of the collector doping profile. The multiplication factor of CIS in Fig. 11 significantly decreases with width in a similar way as the measured data, while REF only shows a very small width dependence like most other cases. For CDL, the width dependence is also significant, but goes in the opposite direction. Due to the smaller (per area) edge and perimeter current contribution $I_{CE}$, a similar behavior is observed for TED but with a smaller dependence.

More detailed investigations using process simulations [11] provided strong evidence that there is indeed a change in the collector doping similar to that modeled in CIS. Scattering of the collector implant atoms causes for wider widths the collector doping to decrease under the emitter and to increase toward the perimeter, which
leads to the observed \( f_T \) drop and geometry dependence of \( B V_{CEO} \).

How can this effect be taken into account in a scalable compact model such as HICUM [15]? Since \( C_{CEO} \) data follow the standard scaling, the corresponding extracted parameter \( C_{CEO, A} \), which enters both \( f_T \) and \( I_{aval} \), remains width independent. Hence, it cannot be used to describe the observed behavior. The physical reason for the latter is a laterally inhomogeneous BC space-charge width and electric field, which lead to width dependent specific parameters \( \tau_0 \) and \( q_{AVL} \). In other words, the (effective) emitter width or area for the minority charge and avalanche current is not the same as for the transfer current. Therefore, \( \tau_0 \) and \( q_{AVL} \) need to be modeled using, e.g., (11).

6 Conclusions

Based on extensive use of device simulation, the impact of process related geometry effects on important electrical parameters of SiGe HBTs as a function of emitter width has been investigated and explained. It was shown that width dependent doping profile changes can still lead to standard scaling behavior in certain electrical parameters and, hence, cannot be detected by observing those parameters only. However, within a more complete set of electrical parameters usually one or more can be found with non-standard scaling behavior. Thus, the combined observation of such a set of electrical parameters often allows to identify the type of profile change. This information can then be used to develop a physics-based description of the associated geometry dependent model parameters in order to be able to still generate geometry scalable compact models. A simple generic extension of the standard scaling law has been suggested that is suitable for compact modeling.

Finally, the simulated characteristics have been compared to experimental results exhibiting non-standard scaling. For the process under consideration, similar trends were obtained confirming the assumption about a SIC profile change.

It is certainly possible that a superposition of different structural changes can lead to compensations and difficulties in their separate detection from electrical characteristics. More investigations are required here, as also for the impact of changes in actual vs. drawn width and 3D effects in small devices.

Acknowledgments: The authors MS and SL would like to thank STMicroelectronics and the German Research Foundation (DFG) for financial support.

References


