

PAPER • OPEN ACCESS

High-performance SiGe HBTs for next generation BiCMOS technology

To cite this article: Holger Rücker and Bernd Heinemann 2018 *Semicond. Sci. Technol.* **33** 114003

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

High-performance SiGe HBTs for next generation BiCMOS technology

Holger Rücker¹  and Bernd Heinemann

IHP, Im Technologiepark 25, D-15236 Frankfurt (Oder), Germany

E-mail: ruecker@ihp-micrelectronics.com

Received 28 June 2018, revised 14 August 2018

Accepted for publication 3 September 2018

Published 10 October 2018



CrossMark

Abstract

This paper addresses fabrication aspects of SiGe heterojunction bipolar transistors which record high-speed performance. We previously reported f_T values of 505 GHz, f_{MAX} values of 720 GHz, and ring oscillator gate delays of 1.34 ps for these transistors. The impact of critical process steps on radio frequency performance is discussed. This includes millisecond annealing for enhanced dopant activation and optimization of the epitaxial growth process of the base layer. It is demonstrated that the use of a disilane precursor instead of silane can result in reduced base resistance and favorable device scalability.

Keywords: heterojunction bipolar transistor, SiGe, BiCMOS, radio frequency

(Some figures may appear in colour only in the online journal)

1. Introduction

Today, SiGe heterojunction bipolar transistors (HBT) in BiCMOS technology environment are widely used for applications like automotive radar, high-speed wireless and optical data links, and high-precision analog circuits. While future 5 G communication standards at frequencies up to 40 GHz are likely to be addressed by advanced CMOS technologies, SiGe BiCMOS technologies are a strong contenders for the required high data rate wireless or fiber optic back-haul. Moreover, the advent of highly-integrated mm and sub-mm wave circuits could pave the way to widespread use of these bands for applications such as high-resolution sensors in robotics, imaging and sensing in medicine, industry, and science, or short-range data transmission. These applications take advantage of an ongoing increase of the radio-frequency (RF) performance of SiGe BiCMOS technologies. Recent advancements in technology, modeling, and mm-wave circuit applications of SiGe HBTs achieved in the European research project DOTSEVEN are reviewed in [1].

¹ Author to whom any correspondence should be addressed

BiCMOS technologies with SiGe HBTs featuring transit frequencies f_T up to 320 GHz and maximum oscillation frequencies f_{MAX} up to 400 GHz have reached industrial production level today [2–6]. Recent research results revealed a significant potential for further enhancement of the RF performance. SiGe HBTs with peak f_T values of 505 GHz and peak f_{MAX} values of 720 GHz were demonstrated on the research level [7]. These devices represent the current state of the art for SiGe HBTs. The record HBTs were fabricated in a bipolar-only process flow. The integration of HBTs with a similar performance in industrial BiCMOS technology platforms is currently addressed in the European project TARANTO launched in 2017. With this new HBT generation it becomes conceivable to extend fundamental operating frequencies of integrated silicon circuits up to 300 GHz which is out of reach with existing technologies. The available wide absolute band width at these frequencies will enable, e.g., wireless data-rates in the order of tens of gigabits per second. High-precision radar and imaging systems can take advantage of the corresponding reduction of the carrier wave length. First implementations of such systems in experimental SiGe HBT technologies are reviewed in [1].

This paper addresses advances in device fabrication technology and device scaling that facilitated the remarkable improvement of RF performance. The impact of device geometries, specific material characteristics, and process integration aspects on device performance is discussed in



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

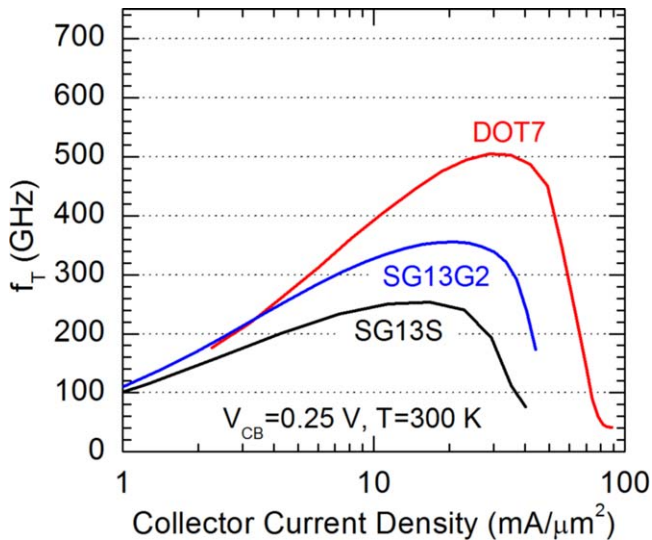


Figure 1. Transit frequency f_T versus collector current density for three HBT generations.

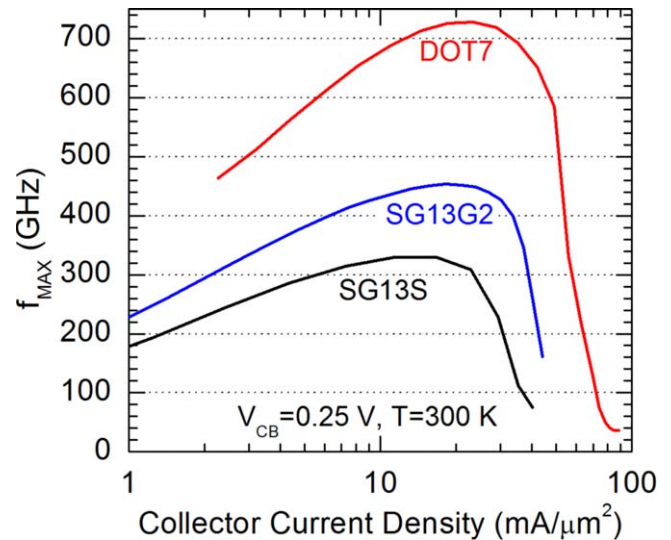


Figure 2. Maximum oscillation frequency f_{MAX} versus collector current density for three HBT generations.

section 2 for three HBT generations. The characteristics of the record-performance HBT of [7] are compared to data for HBTs of IHP's 130 nm BiCMOS technologies SG13S [8] and SG13G2 [9]. Individual process improvements that led to the demonstration of HBTs with 700 GHz f_{MAX} are discussed in section 3. Section 4 addresses the impact of using the higher order precursor gas disilane instead of silane for the epitaxial growth of the base layer stack. Consequences on scalability and device performance are discussed.

2. Progress in RF performance

Figures 1 and 2 show the cutoff frequencies f_T and f_{MAX} as function of the collector current for three HBT generations. The fabrication processes of all three HBT generations are based on the 130 nm BiCMOS technology SG13S of IHP [8] featuring non-selective epitaxial growth of the SiGe base and elevated extrinsic base regions. The devices SG13G2 [9, 10] and DOT7 [7] represent the technology advancements that were achieved based on the results of the two subsequent European research projects DOTFIVE and DOTSEVEN, respectively.

The demonstrated peak f_T/f_{MAX} values of 505 GHz/720 GHz for the DOT7 device are the highest values reported for SiGe HBTs until now. The achieved performance enhancement was made possible by scaling vertical and lateral transistor dimensions and a series of process innovations which helped to reduce non-scaling contributions to device parasitics such as base and emitter resistances. A major feature size for HBT scaling is the emitter width. The emitter widths of the three investigated HBT generations SG13S, SG13G2, and DOT7 are 170 nm, 120 nm, and 105 nm, respectively. These values are still far from the scaling limit indicating room for performance improvement by further device scaling.

Figures 1 and 2 indicate that the roll-off of f_T and f_{MAX} is shifted to higher current densities for the advanced technology generations. This is due to an enhanced dopant concentration in the collector region which suppresses the base-push-out effect. As a consequence, peak values of f_T and f_{MAX} shift to higher current densities for the advanced technology generations. However, it can be seen also that the advanced technologies provide significantly higher values of f_T and f_{MAX} than previous generations for a given collector current density. The higher f_T and f_{MAX} values correspond to higher current gain and power gain for transistor operation at a given application frequency thus enhancing the design margin for circuit applications at this frequency. The reduction of breakdown voltages due to the more aggressive doping profiles of the advanced transistor generations was moderate. The emitter-collector breakdown voltages BV_{EC0} are 1.7 V for the SG13S device and 1.6 V for the SG13G2 and DOT7 devices.

The high-speed performance of a technology for switching operation can be efficiently benchmarked by ring oscillator gate delays. Gate delays of current-mode-logic (CML) ring oscillators are plotted in figure 3 as a function of collector current density for the three technology generations. The oscillators consist of 31 stages and a 1:16 frequency divider. Currents per stage were adjusted to a single-ended voltage swing of 300 mV at a supply voltage of -2.5 V. Figure 3 indicates that the improved f_T and f_{MAX} values of the advanced HBT generations are associated with significantly reduced gate delay times. The minimum gate delay of 1.34 ps for the DOT7 device represents the shortest gate delay demonstrated for a SiGe HBT technology so far [7]. Shorter gate delays have not been reported for any other integrated circuit technology.

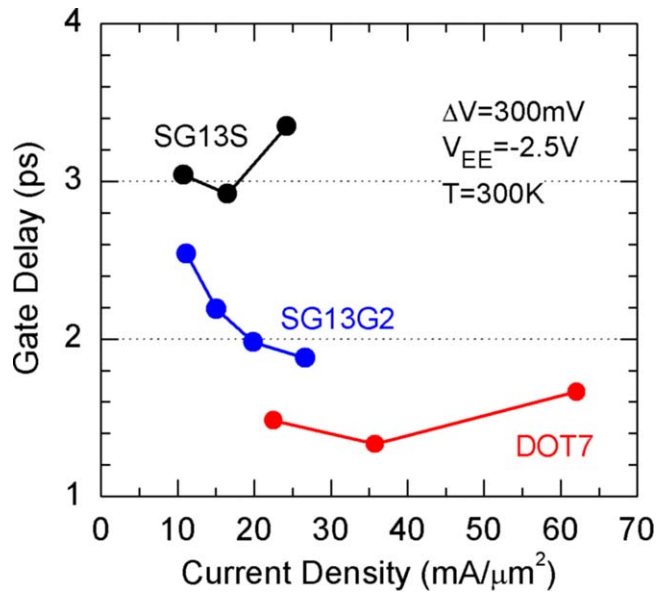


Figure 3. CML ring oscillator gate delay versus collector current density for three HBT generations.

3. Process optimization leading to 700 GHz f_{MAX}

In this section, we review the process developments that facilitated the improved RF performance of the DOT7 device with respect to the SG13G2 reference process. The introduced process modifications addressed the reduction of device parasitics by reducing lateral device dimensions as well as by improving the control of the doping profile and the conductivity of critical device regions. Further details of this process optimization and results for individual device parameters are given in [1, 7].

A schematic cross section of the HBT is given in figure 4. Basic features common to all three HBT generations are: (i) elevated extrinsic base regions self-aligned to the emitter window resulting in low base resistance, (ii) the formation of the HBT in a single active area without shallow trench isolation (STI) between emitter and collector contacting regions resulting in low collector resistance and small collector-substrate junction areas, and (iii) the absence of epitaxially-buried subcollectors and deep trenches in order to limit process complexity. Low-resistive collector wells isolated by standard STI are formed by ion implantation and rapid thermal annealing at the beginning of the HBT process module. Critical lateral dimensions such as the width of the emitter window (w_E), the width of the collector window (w_C), the width of the emitter poly-silicon (w_{EP}) and the width of the base-emitter spacer (d_{Sp}) are indicated. Details of the fabrication process and depth profiles of the doping and germanium concentrations of the advanced transistor generation can be found in chapters 1.3.1 and 1.4 of [1], respectively. The fabrication process includes the following four epitaxial steps. (1) After patterning the oxide layer that forms the isolation between collector and extrinsic base link the opened collector window (w_C) is filled by selective epitaxy of un-doped Si. Subsequently, masked ion implantation is used to form the selectively-implanted collector (SIC) in the inner transistor

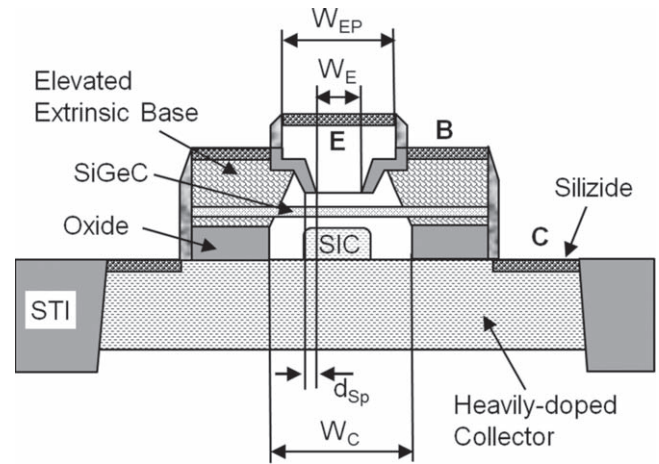


Figure 4. Schematic cross section of the HBT. Key device dimensions are indicated.

region. (2) The base layer stack consisting of an un-doped Si buffer layer, the SiGe:C layer containing the boron-doped base, and an un-doped Si cap layer is grown by non-selective epitaxy. (3) The emitter is formed by epitaxial growth of *in situ* arsenic-doped silicon. (4) After emitter structuring, extrinsic base regions are elevated by selective growth of an *in situ* boron-doped layer.

The realization of high f_{MAX} values requires high cutoff frequencies f_T together with low base resistance R_B and low base-collector capacitance C_{BC} as indicated by the relation

$$f_{MAX} \cong \sqrt{\frac{f_T}{8\pi R_B C_{BC}}}. \quad (1)$$

High f_T values are facilitated by steep vertical doping profiles with short transit times through the base and the base-emitter and base-collector junctions combined with short charging times of the base-emitter and base-collector capacitances. Short transit times were addressed by an optimized SiGe base profile with an increased peak Ge concentration of 32% after deposition and 28% after the full process and by reduced thicknesses of the low-doped base-emitter and base collector transitions. An enhanced arsenic concentration and a reduced thickness of the emitter helped to reduce the emitter resistance. Consequently, low base-emitter and base-collector charging times could be maintained despite of the enhanced capacitances due to reduced junction widths. The broadening of the doping profiles was minimized by reducing the spike temperature for the final rapid thermal process. All three technology generations take advantage of the suppressed diffusion of boron due to carbon doping. The SiGe:C layers are doped with about 10^{20} cm^{-3} carbon.

A further process modification that helped to reduce the excess resistances of base, emitter, and collector was the introduction of a millisecond flash-annealing step combined with a low-temperature back-end-of-line process with nickel silicide. Heating the wafer surface to temperatures close to the melting point at a millisecond time scale results in strongly enhanced dopant activation in heavily doped device regions without noticeable diffusion. The thermal budget after ms-

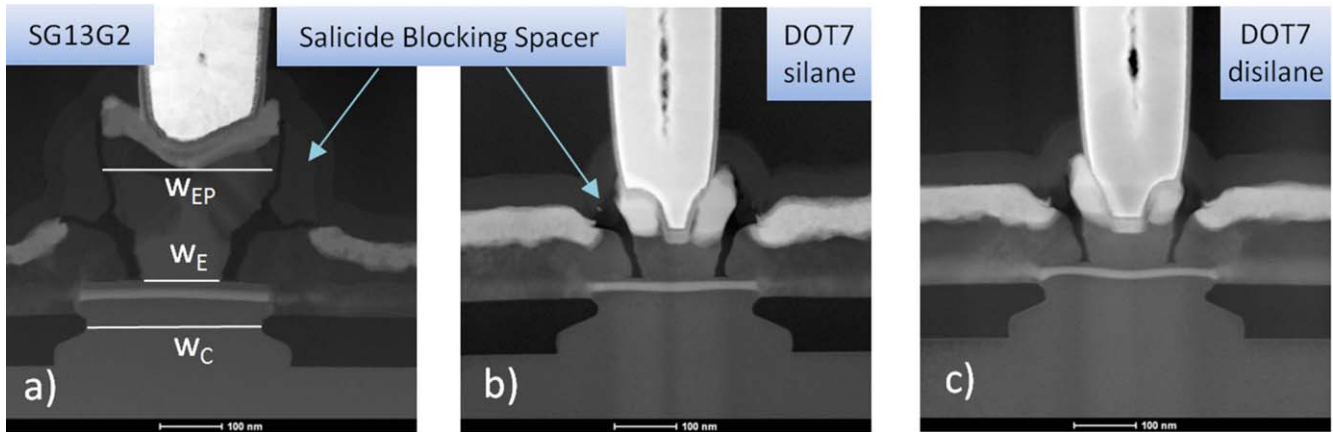


Figure 5. TEM cross sections of SiGe HBTs: (a) SG13G2, (b) DOT7 HBT with silane-based epitaxy (c) DOT7 HBT with disilane-based epitaxy.

annealing has to be minimized in order to avoid dopant deactivation. This is supported by the replacement of the CoSi_2 process by a NiSi process with lower thermal budget. An additional reduction of R_B was achieved by optimizing the process for the selective epitaxial growth and by increasing the dopant concentration of the elevated external base regions.

Scaling of lateral device dimensions is essential for simultaneous realization of low R_B and low C_{BC} that is required for high f_{MAX} values. Figures 5(a) and (b) show transmission electron microscopy cross sections of SG13G2 and DOT7 HBTs, respectively. Critical lateral dimensions such as w_E , w_C , and w_{EP} were reduced for the DOT7 devices (see figure 5) by reducing the corresponding lithographic dimensions. However, the possibilities for scaling the emitter window by lithographic measures are limited by the resolution of the used 248 nm DUV scanner. Therefore, w_E could be scaled only moderately from 120 to 105 nm. The width of base-emitter spacers was reduced for the DOT7 device by modifying the corresponding etching and deposition processes. In addition, the width of the silicide blocking spacers at the outside of the emitter poly was reduced for the DOT7 device. Moreover, the process for the SIC that is formed before base epitaxy in the collector region below the emitter window was changed from a resist mask in SG13G2 to a hard mask process in DOT7 allowing for an improved control of the width and the doping concentration of the SIC region.

The described measures for lateral device scaling resulted in a significant reduction of R_B and of contributions to C_{BC} due to device edges for the DOT7 HBT in comparison to the SG13G2 device [7]. In combination with the enhanced f_T of the DOT7 device, this resulted in the strong increase of f_{MAX} shown in figure 2.

4. Study of two precursor gases for base epitaxy

A crucial step of the HBT fabrication is the epitaxy of the base layer stack. At the beginning of this epitaxial process, the wafers are covered with an about 80 nm thick oxide layer except for small collector windows of width w_C . These

collector windows were filled with Si up to the level of the oxide surface by a preceding selective epitaxial process. The following epitaxy of the Si/SiGe:C/Si base layer stack is performed in a differential manner, i.e., single-crystalline layers are grown on Si areas whereas poly-crystalline layers are grown on oxide areas. We have investigated two processes options for the epitaxy of the base layer stack which use silane and disilane as Si precursor gases, respectively. The replacement of silane by the higher order precursor disilane results in significantly higher growth rates. As a consequence, lower deposition temperatures are applicable for the disilane process.

Next, the two processes are analyzed with respect to their potential for lowest base resistance and device scaling. Figure 6 shows scanning electron microscopy (SEM) images of HBTs after deposition of the base layer stack using silane-based and disilane-based epitaxy. Silane epitaxy (figure 6(a)) results in crystalline growth on Si areas and poly-crystalline growth on oxide areas. The grain structure of the poly-crystalline layer causes a rough transition line between the poly-crystalline and single-crystalline areas as indicated in figure 6(a). Disilane epitaxy results in crystalline growth on Si areas and amorphous growth on oxide areas due to the reduced temperature ($T < 550^\circ\text{C}$). The amorphous layer is crystallized in a subsequent annealing step. Figure 6(b) shows an SEM top view after annealing. The resulting poly-crystalline layer is smoother than in the case of silane epitaxy. Moreover, a sharp transition line is formed between the poly and single-crystalline areas. For ideal transistor operation, the emitter window has to be securely positioned inside the unperturbed crystalline Si area within the collector window. The bumpy border line of the crystalline area in the case of silane epitaxy restricts the usable active transistor area in this case. A sharp straight border of the crystalline area as indicated in figure 6(b) is advantageous for minimum overlap of the collector window (w_C) over the emitter window (w_E) and thus supports the reduction of the contribution of the device edges to the base-collector capacitance.

Both epitaxy variants were evaluated in process splits of the HBT fabrication. Device cross sections for the two cases

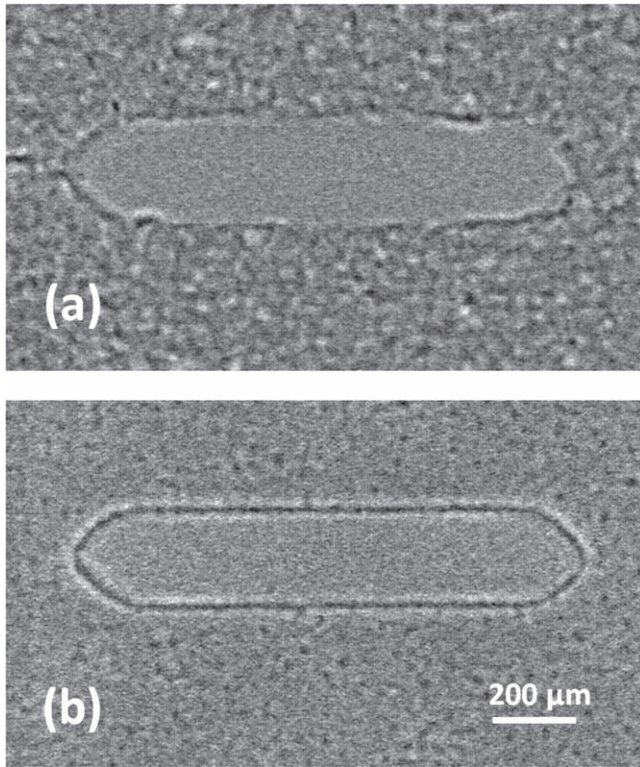


Figure 6. SEM top view of two HBTs with identical collector window size after epitaxial growth of the Si/SiGe/Si base layer stack. (a) Silane-based epitaxy, (b) disilane-based epitaxy.

are shown in figures 5(b) and (c). Figure 5(b) shows a flat and smooth SiGe base layer (bright area below emitter window) within the active transistor area for the silane-based epitaxy. In contrast, the SiGe layer is wavy in the device with disilane-based epitaxy and the thickness of the SiGe layer is varying across the active transistor area. Further optimization of the disilane-based process is needed to overcome this disadvantage.

The HBT with best RF performance [7] reviewed in section 2 was fabricated with silane epitaxy of the base layer stack. The process with disilane epitaxy resulted in lower peak f_T and f_{MAX} values for the most scaled device. This is

attributed to a not yet fully optimized process in the case of disilane epitaxy. Nevertheless, the present experiment indicates a potential advantage of the disilane-based process for the realization of lowest base resistance values as discussed below.

In the following, we analyze electrical results for HBTs with three different emitter window widths fabricated with silane and disilane epitaxy of the base layer stack. Measured device parameters are summarized in table 1. The highest f_T is obtained in both process splits for the device with the intermediate emitter width of $0.14 \mu\text{m}$. Peak f_T values are about 15 GHz lower for the smallest emitter width of $0.105 \mu\text{m}$. This is attributed to the relative increase of non-scaling contribution from the device edges to delay times and to C_{BC} and C_{BE} . The disilane-based devices show about 50 GHz lower peak f_T values than the corresponding silane-based devices. This is attributed to imperfectly matched Ge and B profiles of the two base layers. The disilane-based devices exhibit a lower Ge concentration and a higher base doping resulting in a lower transconductance at collector currents corresponding to peak f_T . Peak f_{MAX} values are highest for the devices with minimum emitter widths of $0.105 \mu\text{m}$ due to the reduction of R_B with decreasing w_E . Note that the devices analyzed here show somewhat lower f_{MAX} values than the record device discussed in section 2 due to a more relaxed device layout with slightly larger w_C and w_{EP} .

It is remarkable that the HBTs with disilane epitaxy show significantly higher f_{MAX} than the devices with silane epitaxy despite of about 50 GHz lower f_T values. The higher f_{MAX} values of the disilane devices are mainly caused by lower values of R_B and to a smaller extend by lower C_{BC} (see table 1). The base resistance R_B is about 3Ω lower for the devices with disilane epitaxy for all three emitter geometries. This cannot be attributed to the 14% lower intrinsic base sheet resistance R_{sbi} alone but also to a lower external base resistance R_{Bx} in the case of disilane epitaxy. Contributions to R_{Bx} that are relevant here include the contact resistance between silicide an external base poly-silicon, the sheet resistance of the external base poly-silicon layer, and the contact resistance between external base poly-silicon and the single-crystalline base region. The silane and disilane processes resulted in the

Table 1. HBT parameters for process splits with silane and disilane base epitaxy. Three devices with different emitter widths w_E and common emitter length l_E are shown. C_{BC} and C_{BE} were extracted from S -parameter measurements. R_E was measured with the open collector method. $R_B + R_E$ was derived from circle fit of s_{11} . R_{sbi} was derived from tetrode structures. $R_s(\text{BP})$ is the sheet resistance of the un-silicized external base poly-silicon layer.

	Unit	Silane			Disilane		
w_E	μm	0.105	0.14	0.24	0.105	0.14	0.24
l_E	μm		5			5	
Peak f_T	GHz	486	500	481	433	451	433
Peak f_{MAX}	GHz	560	519	368	604	561	389
C_{BC}	fF	9.3	10.3	13.3	9	10	13
C_{EB}	fF	12.1	14.4	19.7	11.7	13.8	19.3
R_E	Ω	2.1	1.3	1.1	3.2	2.1	1.6
R_B	Ω	11.0	11.6	15.4	7.7	8.8	12.0
R_{sbi}	Ω		2.9			2.5	
$R_s(\text{BP})$	Ω		66			65	

same low sheet resistance of the un-silicized external base poly-silicon layer $R_s(\text{BP})$ (table 1). The lower external base resistance R_{Bx} of the disilane process results most likely from a lower resistance between the internal base and the external base poly-silicon at the edge of the collector window due to favorable structural properties in this case.

In conclusion, the use of disilane instead of silane as precursor gas for the epitaxy of the base layer stack opens up attractive additional degrees of freedom for device optimization. This concerns the morphology of the base layer stack which is favorable for device scaling as well as the potential for reduction of the base resistance. However, further optimization of the disilane-based process is required for the realization of flat and homogeneous SiGe layers down to smallest collector windows.

5. Summary

The fabrication of a new generation of high-performance SiGe HBTs with peak f_T/f_{MAX} values of 505 GHz/720 GHz and CML ring oscillator gate delays of 1.34 ps has been reviewed. The integration of such HBTs in an RF BiCMOS technology platform will significantly enhance its capability for mm-wave and THz applications. A route to further enhancement of the RF performance of SiGe HBTs is scaling of vertical and lateral device dimensions. One of the crucial process steps that define the scaling capability of the technology is the epitaxial growth of the SiGe base layer stack. It has been demonstrated here that the use of a disilane precursor instead of silane can result in reduced base resistance and favorable device scalability.

Acknowledgments

This work was supported by the European Commission and the German Federal Ministry of Education and Research (BMBF) under the contract No. 737454—ECSEL TARANTO. The authors gratefully acknowledge numerous contributions of the HBT team of IHP, in particular the support of

Y Yamamoto and B Tillack for the disilane process development.

ORCID iDs

Holger Rucker  <https://orcid.org/0000-0001-7407-959X>

References

- [1] Rinaldi N and Schröter M (ed) 2018 *Silicon-Germanium Heterojunction Bipolar Transistors for mm-Wave Systems: Technology, Modeling and Circuit Applications* (Gstrup and Delft: River Publishers)
- [2] Chevalier P *et al* 2014 A 55 nm triple gate oxide 9 metal layers SiGe BiCMOS technology featuring 320 GHz f_T /370 GHz f_{MAX} HBT and high-Q millimeter-wave passives *IEEE Int. Electron Devices Meeting (IEDM)* pp 77–9
- [3] Trivedi V P *et al* 2016 A 90 nm BiCMOS technology featuring 400 GHz f_{MAX} SiGe:C HBT *IEEE Bipolar/BiCMOS Circuits and Technology Meeting* pp 60–3
- [4] Pekarik J J *et al* 2014 A 90 nm SiGe BiCMOS technology for mm-wave and high-performance analog applications *IEEE Bipolar/BiCMOS Circuits and Technology Meeting* pp 92–5
- [5] Hurwitz P *et al* 2016 Advances in RF foundry technology for wireless and wireline communications *IEEE 16th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)* pp 5–8
- [6] Böck J *et al* 2015 SiGe HBT and BiCMOS process integration optimization within the DOTSEVEN project *IEEE Bipolar/BiCMOS Circuits and Technology Meeting* pp 121–4
- [7] Heinemann B *et al* 2016 SiGe HBT with f_T/f_{MAX} of 505 GHz/720 GHz *IEEE Int. Electron Devices Meeting (IEDM)* pp 1–4
- [8] Rucker H *et al* 2010 A 0.13 μm SiGe BiCMOS technology featuring f_T/f_{MAX} of 240/330 GHz and gate delays below 3 ps *IEEE J. Solid-State Circuits* **45** 1678–86
- [9] Rucker H, Heinemann B and Fox A 2012 Half-terahertz SiGe BiCMOS technology *12th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)* pp 133–6
- [10] Heinemann B *et al* 2010 SiGe HBT technology with f_T/f_{MAX} of 300 GHz/500 GHz and 2.0 ps CML gate delay *IEEE Int. Electron Devices Meeting (IEDM)* pp 1–4