Effect of uniaxial strain on characteristic frequency of scaled SiGe HBT with embedded stress raiser

Senlin Kong[1] · Jinxi Wei[1] · Guanyu Wang[1] · Chunyu Zhou[2] · Qi Song[3] · Jianhao Wen[1] · Yingcong Xiang[1] · Peng Ling[1]

Abstract

In order to further improve the high-frequency characteristics of highly scaled SiGe HBT and consider the compatibility with the mature CMOS technology, a new SiGe HBT structure is proposed by introducing an embedded $Si_{1-y}Ge_y$ stress raiser to produce additional uniaxial stress in the bulk collector region. The energy-band configuration of multi-layered emitter has been investigated by the estimation of the strain effect, and then the influence of stress raiser on the electrical properties and frequency response has been studied by employing the SILVACO TCAD tools. The results show that the device performance has been enhanced to different degrees by altering the Ge fraction (y) in the stress raiser. At $y = 0.3$, the current gain is increased by approximately 6% compared to the case without stress in the collector region ($y = 0$). For the case of a uniform $Si_{0.75}Ge_{0.25}$ base, the cut-off frequency (f_T) and the maximum oscillating frequency (f_{max}) are, respectively, peaked at 507.7 GHz and 730.7 GHz. Approximately 29.1% improvement in f_T and 71.5% improvement in f_{max} have been achieved for the proposed HBT device in comparison with an equivalent traditional SiGe HBT. At $y = 0.1$, the frequency characteristic is considered the best due to the maximum of $f_T \times f_{max}$ product.

Keywords SiGe HBT · Uniaxial stress · Stress raiser · Cut-off frequency

1 Introduction

As integrated circuits (ICs) continue to advance to smaller semiconductor process nodes, the cut-off frequency (f_T) and the maximum oscillation frequency (f_{max}) of silicon (Si)-based radio frequency (RF) microelectronic devices are now moving into the terahertz (THz) range. Compared to the traditional III-V devices, Si-based high-frequency devices

Senlin Kong and Jinxi Wei have contributed equally to this work.

✉ Guanyu Wang
sigebicmos@foxmail.com

Chunyu Zhou
zhouchunyu@ysu.edu.cn

[1] College of Electronic Engineering, Chongqing University of Posts and Telecommunications, Chongwen Road, Chongqing 400065, China

[2] Key Laboratory for Microstructural Material Physics of Hebei Province, School of Science, Yanshan University, Hebei Street, Qinhuangdao 066004, Heibei province, China

[3] College of Software Engineering, Chongqing University of Posts and Telecommunications, Chongwen Road, Chongqing 400065, China

have the technical advantages of low cost, easy mass production, and compatibility with the ULSI technology, gradually attracting attention in the application of THz ICs [1]. As the nuclear devices in RF ICs, silicon–germanium (SiGe) heterojunction bipolar transistors (HBTs) offer the advantages of strain engineering and are compatible with CMOS technology. In order to develop f_T of SiGe HBTs in the range of 0.5–0.7 THz for the applications such as space communications and remote sensing, the "DOTFIVE" and "DOTSEVEN" projects [2, 3] have been proposed by STMicroelectronics, Infineon, and other research institutions to give full play to the great potential and technical advantages of SiGe HBTs in the reference circuits and systems. The international semiconductor technology roadmap (ITRS) predicts that the f_T/f_{max} of SiGe HBT will reach 570/610 GHz by 2020 [4]. The improvement of the frequency response of SiGe HBTs is achieved by continuously optimizing the process and scaling the lateral size, which may increase the process cost [5–7].

As the characteristic size of microelectronics devices is becoming smaller and smaller, the challenge of size-scaling is becoming more and more significant. Strain engineering has been considered one of the important technologies

for high-speed devices [8]. Some attempts have been made to introduce the strain effect in SiGe HBTs [9–12]. Metal interconnects are stacked near the SiGe HBTs according to the BiCMOS-9 W technologies of STMicroelectronics [9], because of the different thermal expansion coefficients between each layer, the uniaxial stress is generated into the devices, f_T and f_{max} are then increased by 21% and 12%, respectively. A strained-Si (SSi) HBT structure is proposed [10], using a relaxed SiGe virtual substrate as the collector and an SSi layer as the emitter. The current gain is improved by 11% compared to traditional SiGe HBTs, but the self-heating effect is significant which limits the frequency performance. An external SiGe stress film is used to apply uniaxial stress to the base region of Si bipolar transistor (BJT), and TCAD simulation is carried out based on the fluid dynamics model [11], and the simulation results showed that f_T and f_{max} are increased by 5% and 3%, respectively. In addition, the effect of global additional stress on frequency response is investigated by full-band Monte-Carlo simulation [12], but the methods of stress introduction and stress adjustment are not discussed in detail.

In this work, a new NPN SiGe HBT architecture utilizing a $Si_{1-y}Ge_y$ stress raiser in the bulk-Si collector region is proposed. Due to the mismatch of the lattice constants between the $Si_{1-y}Ge_y$ stress raiser and Si substrate, additional compressive stress is applied along the horizontal axis, and tensile stress is induced along the vertical axis, which enhances electron mobility in the collector region. The physical mechanisms affecting the DC and AC performance of the proposed device will be investigated by theoretical calculation and TCAD simulation.

2 Device structure

The proposed structure of NPN SiGe HBT with an embedded stress raiser is shown in Fig. 1. The emitter region consists of a thin monocrystalline silicon cap layer and a thick polysilicon layer. Both the emitter and the base regions have the same lateral width, which is beneficial for reducing the parasitic effect between the cap layer and base layer, thus improving characteristics frequency. To diminish the base intrinsic resistance, the lateral width of the emitter is reduced to less than 100 nm. In the collector region, the stress raiser generates uniaxial compressive stress, which can improve the electron mobility, and reduce the series resistance, longitudinal transit time, as well as BC junction capacitance. In the emitter region, the thin Si cap layer is also subjected to the same uniaxial compressive stress as the collector region, and the current injected from the base region to the emitter region is effectively decreased by the interfacial energy barrier between the polysilicon layer and strained-Si (SSi) cap layer, which is advantageous for improving current gain.
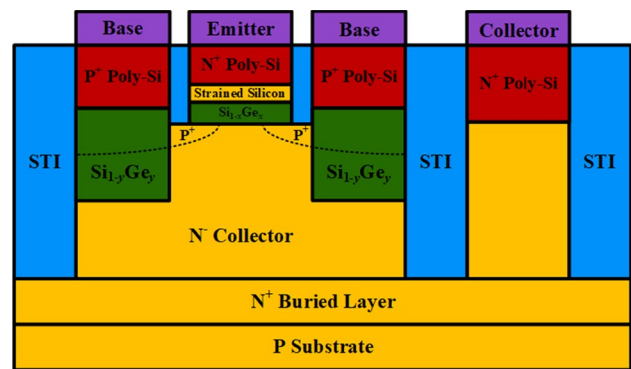

Fig. 1 Device structure of proposed NPN SiGe HBT with an embedded stress raiser

The brief process flow is shown in Fig. 2. The proposed device structure selects a P-type (110)-oriented Si substrate. An N^+-doped layer and an N^--doped layer are epitaxially formed as buried layer (sub-collector) and collector region, respectively, and the right collector region is N^+-doped. The N^- collector region is ion implanted to form the P^+-doped on both sides as two extrinsic base regions. A SiO_2 layer with a thickness of 1–2 μm is deposited on the surface of the device to define the position of an active region. A P-type $Si_{1-x}Ge_x$ layer, an intrinsic SSi cap layer, and an N^+-doped polysilicon layer are selective epitaxially formed in the active region. A nitride layer is deposited on the surface of the device, both the nitride layer and SiO_2 layer are photoetched, the P^+ inner base region is etched, and an embedded $Si_{1-y}Ge_y$ layer is selective epitaxially formed. A polycrystalline $Si_{1-y}Ge_y$ layer is also deposited on the surface of the embedded $Si_{1-y}Ge_y$ layer with the same Ge mole fraction y as a part of the extrinsic base. Two polysilicon electrodes are re-deposited and metal silicide is deposited, to form the electrodes as the emitter, base, and collector.

3 Estimation of strain mechanism

Since the lattice constant of SiGe is greater than that of Si, the embedded $Si_{1-y}Ge_y$ stress raiser generates a longitudinal uniaxial compressive stress σ_{xx} in the collector region. The lattices of Si collector are further compressed, and the lattice constant decreases. The conventional biaxially strained $Si_{1-x}Ge_x$ base is superimposed with uniaxial compressive stress to form the uni-biaxially strained base region. The Si cap layer in the emitter region is also affected by the uniaxial compressive stress σ_{xx}. Considering the possibility of integrating with uniaxially strained Si pMOSFETs, the additional stress is usually applied along the [110] direction, and
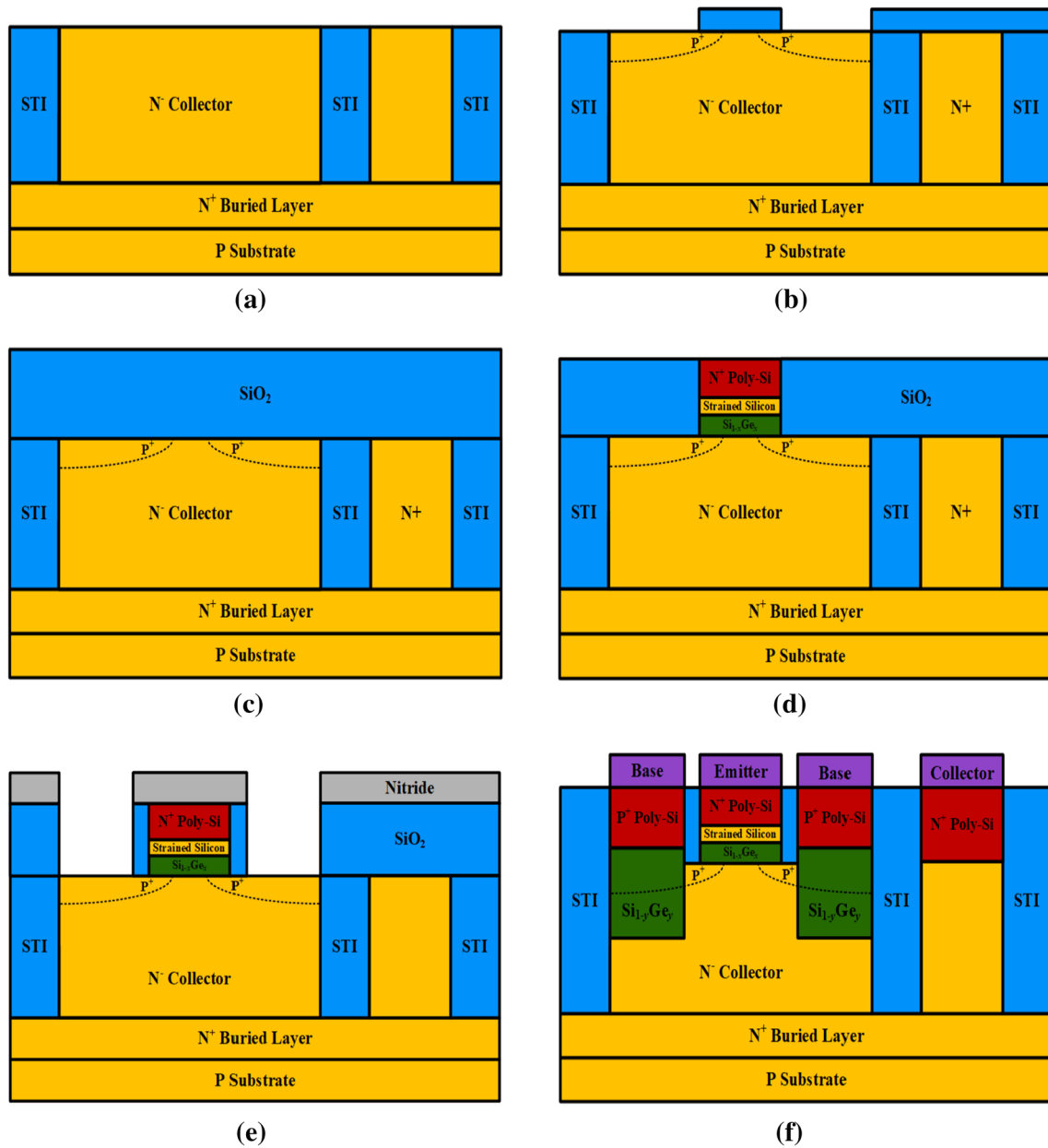
Fig. 2 Brief process flow of proposed NPN SiGe HBT structure

the strain tensors of the Si cap layer and bulk-Si collector are then expressed as the same form [12]

$$\varepsilon_{Si} = \begin{bmatrix} s_{12}\sigma_{xx} & 0 & 0 \\ 0 & \frac{1}{2}(s_{11}+s_{12})\sigma_{xx} & \frac{1}{4}s_{44}\sigma_{xx} \\ 0 & \frac{1}{4}s_{44}\sigma_{xx} & \frac{1}{2}(s_{11}+s_{12})\sigma_{xx} \end{bmatrix}, \quad (1)$$

s_{11}, s_{12} and s_{44} are the elastic compliance coefficients, and the longitudinal strain tensor is

$$\varepsilon_{||,Si} = \frac{1}{2}(s_{11}+s_{12})\sigma_{xx}. \quad (2)$$

Once the biaxially strained SiGe base region is subjected to additional uniaxial stresses from the collector region, the lattice of $Si_{1-x}Ge_x$ will be further compressed and the lattice constant will be reduced. In order to maintain the same strain level as the longitudinal strain tensor $\varepsilon_{||,Si}$ in Eq. (2), the longitudinal strain tensor in the unibiaxially strained base can be equated to

$$\varepsilon_{||,SiGe} = (1 + \varepsilon_{||,Si})\frac{a_0}{a(x)} - 1, \quad (3)$$

where a_0 and $a(x)$ represent the lattice constants of relaxed Si and relaxed $Si_{1-x}Ge_x$, respectively. If the longitudinal strain tensor of biaxially strained base is consistent with $\varepsilon_{\parallel,SiGe}$ in Eq. (3), the equivalent lattice constant $a^*(x)$ of uni-biaxially strained base can be written as:

$$a^*(x) = \frac{a(x)}{\left[1 + \frac{1}{2}(s_{11} + s_{12})\sigma_{xx}\right]}. \tag{4}$$

For compressive stress, $\sigma_{xx} < 0$, and conversely, $\sigma_{xx} > 0$. It is clear that $a^*(x) > a(x)$ in Eq. (4), which corresponds to the fact that the uniaxial compressive stress applied in the collector is equivalent to an increase in Ge mole fraction of $Si_{1-x}Ge_x$ base while maintaining the same level of strain as in a biaxially strained base, which leads to a decrease in the band-gap and an increase in the intrinsic carrier concentration of $Si_{1-x}Ge_x$ base and SSi cap layer. f_T is usually considered to represent the characteristic frequency. In the small-signal model, the expressions of f_T and f_{max} are as follows [13]:

$$f_T = \frac{1}{2\pi\left[\tau_F + \frac{kT}{qI_C}(C_{BE} + C_{BC}) + (R_E + R_C)C_{BC}\right]}, \tag{5}$$

and

$$f_{max} = \sqrt{\frac{f_T}{8\pi R_B C_{BC}}}, \tag{6}$$

where kT/qI_C is the BE junction dynamic resistance, I_C is the collector current, C_{BC} and C_{BE} are the BC junction capacitance and BE junction capacitance, and R_E and R_C the emitter resistance and collector resistance, respectively. The forward transit time is approximately expressed as [14]

$$\tau_F \approx \frac{1}{\beta}\left(\frac{W_E}{S_{pE}} + \frac{W_E^2}{2D_{pE}}\right) + \frac{W_B^2}{2D_{nB}}. \tag{7}$$

In order to reflect the effect of stress on the characteristic frequency, this work only considers the Ge mole fraction of the $Si_{1-x}Ge_x$ base as a constant, and uniform doping in the base region. In Eq. (7), W_E and W_B are the emitter region thickness and base region thickness, respectively. β is the current gain, and D_{pE} and D_{nE} are the minority carrier diffusion coefficients in the emitter region and base region, respectively. As shown in Fig. 3, an extremely thin oxide-like (SiO_r) layer is formed at the interface between the polysilicon and SSi cap layer, S_{pE} represents the holes recombination velocity at the interface, and d is the thickness of the oxide-like layer. χ_e and χ_h represent the barrier heights for electrons and holes, respectively. For electrons injected from the emitter into the base, the transport properties are virtually unaffected by the SiO_r layer and are transparent
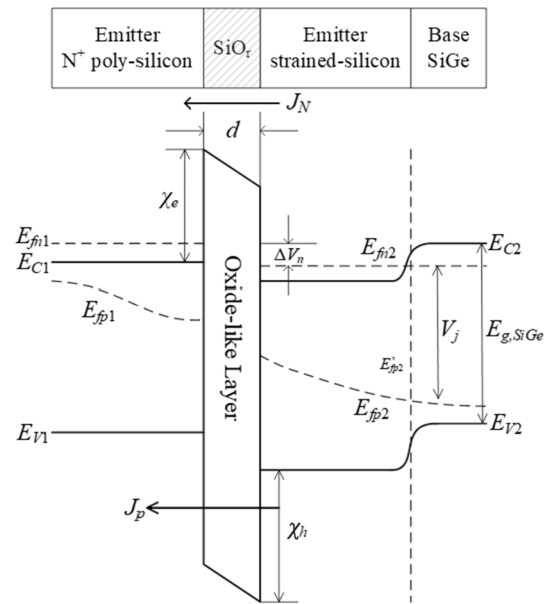

Fig. 3 Energy-band system of poly-Si/oxide-like layer/SSi cap layer

to the electrons due to small χ_e [15]. The holes from the base are injected into the polysilicon layer through the SiO_r layer in the form of quantum tunneling. V_j is the voltage across Si/SiGe heterojunction, and the applied voltage is $V_{BE} = V_j + \Delta V_n$. $\Delta V_p (>> kT/q)$ is the difference between quasi-Fermi energy levels E_{fp1} and E_{fp2}. Using the Boltzmann distribution, the tunneling current can be written as [16]

$$J_P = J_{PT} = \frac{4q\pi m_h^*(kT)^2 P_T}{h^3}\exp\left(-\frac{q\Delta V_p - E_{fp1}}{kT}\right), \tag{8}$$

where h is the Planck constant, m_h^* is the effective mass of holes,

$$P_{T,Si} = \exp(-d\sqrt{\chi_h})$$

is the tunneling probability of holes. The SSi cap layer is very thin and is uniformly doped with the concentration of N_D. χ_h increases $\Delta E_{V,SSi}$ due to the upward shift of the valence band under the action of stress, and the tunneling probability becomes

$$P_{T,SSi} \approx \exp(-d\sqrt{\chi_h + \Delta E_{V,SSi}}). \tag{9}$$

Neglecting the recombination in the cap layer, the tunneling current is rewritten as

$$J_P = \frac{qn_{i,SSi}^2 P_{T,SSi}}{N_D}\sqrt{\frac{kT}{2\pi m_h^*}}\exp\left(\frac{qV_j}{kT}\right) = \frac{qn_{i,Si}^2}{G_{E,SSi}}\exp\left(\frac{qV_{BE}}{kT}\right), \tag{10}$$

where $n_{i,Si}$ and $n_{i,SSi}$ are the intrinsic carrier concentrations of relaxed Si and SSi, respectively. The stress applied to the Si cap layer yields the downward shift of the conduction band, ΔV_n is then increased by $\Delta E_{C,SSi}/q$, and the internal junction voltage is decreased by $\Delta V_j = \Delta E_{C,SSi}/q$. For the uniform doping base with a constant of Ge mole fraction, the Gummel number of the emitter region and the collector current density are as follows:

$$G_{E,SSi} = \frac{N_D}{S_{pE}} \frac{n_{i,\,Si}^2}{n_{i,\,SSi}^2} \exp\left(\frac{q\Delta V_n + \Delta E_{C,SSi}}{kT}\right), \quad (11)$$

$$J_C = \frac{q \exp\left(\frac{qV_{BE}}{kT}\right)}{\left[\int_0^{W_B} \frac{N_B}{D_B^*(n_{i,SiGe}^*)^2} dx\right]} = \frac{q(n_{i,SiGe})^2}{G_{B,SSi}} \exp\left(\frac{qV_{BE}}{kT}\right). \quad (12)$$

The recombination velocity is expressed as $S_{pE} = P_{T,SSi}\sqrt{kT/2\pi m_h^*}$. As discussed earlier, the additional stress leads to an increase in the intrinsic carrier concentration, i.e. $n_{i,SiGe}^* > n_{i,SiGe}$. The minority carrier diffusion coefficient also increases in accordance with $D_{pE}^* \propto \mu_{nB} = (1+3x)\mu_{n0}$[17], where μ_{n0} is the electron mobility of relaxed Si. It follows that because the additional stress makes the Ge fraction of $Si_{1-x}Ge_x$ base increase equivalently, leading to an increase in D_B^*. The Gummel number of the base region can be derived directly from Eq. (12) as

$$G_{B,SSi} = \frac{(n_{i,SiGe})^2}{(n_{i,SiGe}^*)^2} \frac{W_B N_B}{D_B^*} \exp\left(\frac{q\Delta V_n + \Delta E_{C,SSi}}{kT}\right). \quad (13)$$

DC current gain is defined as $\beta = G_{E,SSi}/G_{B,SSi}$ [16]; from Eqs. (12) and (13), we then have the following relationship:

$$\beta = \left(\frac{1}{S_{pE}}\right)\left(\frac{n_{i,\,Si}^2}{n_{i,\,SSi}^2}\right)\frac{(n_{i,SiGe}^*)^2}{(n_{i,SiGe})^2}\left(\frac{D_B^* N_D}{W_B N_B}\right). \quad (14)$$

The product of R_C and C_{BC} in Eq. (5) represents the charge and discharge time of BC junction barrier capacitance, which has a significant effect on the frequency. There is a large amount of mobile charge in the BC junction barrier region. The width of the barrier region is related to J_C. If only the effect of V_{BE} on J_C is considered, C_{BC} can be approximated as

$$C_{BC} = \frac{d}{dV_{CB}}\left[\left(qN_C - \frac{J_C}{v_S}\right)W_{BC}\right] \approx \left(qN_C - \frac{J_C}{v_S}\right)\frac{dW_{BC}}{dV_{CB}}, \quad (15)$$

where v_S is the electron saturation velocity in the collector region, W_{BC} is the width of the depletion region of the BC junction, and it is also related to J_c, i.e.

$$W_{BC} = \sqrt{2\varepsilon_{Si}(V_{bi} + V_{CB})/(qN_C - J_C/v_S)}. \quad (16)$$

The uniaxial stress in the collector region will narrow the band-gap of the collector region, hence, the built-in potential V_{bi} in the depletion region of the BC junction decreases and C_{BC} increases, but J_C is also affected by the stress. If the Ge mole fraction y of the stress raiser is very small, then the effect of stress on C_{BC} is negligible [11]. In addition, considering that the stress of embedded $Si_{1-y}Ge_y$ stress raiser increases with increasing values of y, dislocations can be introduced within the collector region. The dislocations in Si and Ge are mainly prismatic ones that can form suspended chains, and act as donors or acceptors, resulting in a compensation effect on the charge concentration. The dislocations can be considered as scattering centers and affect the mobility and resistivity, which will also affect the frequency characteristics under the action of larger stress.

4 Results and analysis

The physical models used for simulation are the concentration-dependent mobility model, the Auger recombination model, the stress model, the parallel electric field-dependent model, the band-gap narrowing model, the energy balance transport model, the Shockley–Read–Hall recombination model, and the Fermi–Dirac statistical model. After these models have been deployed, numerical methods such as Newton iteration and Gummel iteration are used to calculate the parameters of each grid to obtain the characteristics. In this section, the proposed structure is simulated using the ATHENA module of the SILVACO TCAD tools according to the process steps in Fig. 2, and the device structure file generated by the ATHENA module is imported into the ATLAS module for simulation. Based on the device architecture shown in Fig. 1, the parameters for simulation are listed in Table 1.

Figure 4 shows the variation of the additional uniaxial stress in the collector region along the transport direction (longitudinal) for different Ge mole fractions of the stress raiser. It is clear that the stress raiser can effectively introduce stress within the collector region and that the generated stress increases significantly as the Ge fraction of the stress raiser increases.

The variation of the current gain with different stress raiser Ge mole fractions is represented in Fig. 5. When y is between 0 and 5%, the current gain decreases with increasing stress, after which it increases with increasing y values, but the increase is slight, which can be explained according to Eq. (14): The $Si_{1-y}Ge_y$ stress raiser with Ge fraction between 0 and 5% produces very small stress (< 0.1 GPa), and with such small uniaxial compressive stress, the increase in the equivalent Ge mole fraction of the $Si_{1-x}Ge_x$

Table 1 Structural and process parameters for simulations

Region	Material	Thickness/nm	Width/nm	Doping/cm^{-3}	Ge fraction/%
N$^+$ emitter	Polysilicon	100	80	3×10^{19}	0
Cap layer	Strain Si	20	80	2.5×10^{18}	0
P$^+$ base	Strain Si$_{1-x}$Ge$_x$	30	80	2.5×10^{18}	25
Stress raiser	Strain Si$_{1-y}$Ge$_y$	150	92	0	0–30
N$^-$ collector	Strain Si/relaxed Si	300	280	1×10^{17}	0
N$^+$ collector	Relaxed Si	200	340	1×10^{20}	0
P$^+$ substrate	Relaxed Si	100	340	1×10^{15}	0

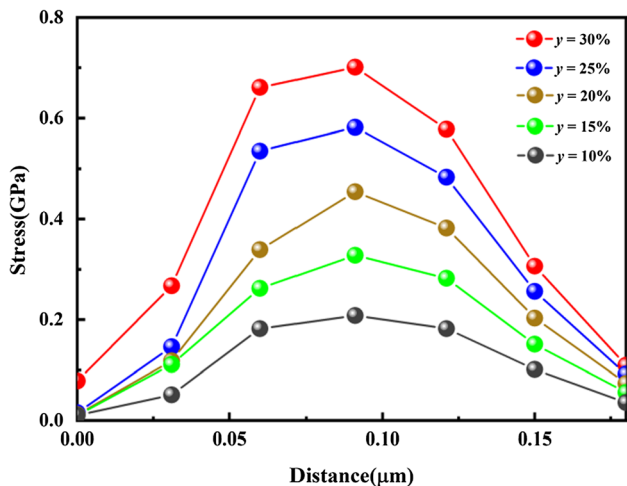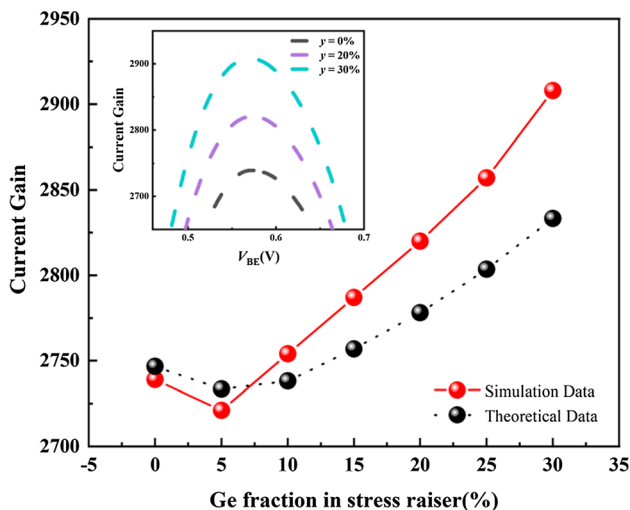**Fig. 4** Simulation result of longitudinal stress distribution in collector region

Fig. 5 Variation of current gain with Ge mole fraction in stress raiser

base is negligible, $\Delta E_{g,SSi}$ and $\Delta E_{V,SSi}$ are also insignificant. Since χ_h is usually about 1 eV [15], the tunneling probability P_T in Eq. (8) has little effect, which approximately yields $\beta \propto \sqrt{m_h^*}$. The theoretical calculations show that the

average effective mass of valence band decreases significantly for stress in the range 0 to 0.2 GPa, thus resulting in a slight decrease in β. When the stress is greater than 0.2 GPa, the magnitude of the band-gap shift increases with enhancing stress. As a result, the accumulation of holes, at the interface between the SSi cap layer and SiO$_r$ layer, is significantly diminished, then χ_h plays an important role in sustainably increasing β. In general, the first and second terms to the right of the equal sign in Eq. (14) decrease as stress increases, while the remaining terms increase as stress increases. When the Ge mole fraction y is small, term 1 causes β to decrease; on the contrary, terms 3 and 4 become the main factors that make β increase.

It is necessary to point that there are many factors affecting the characteristic frequency. The previous theoretical analysis is to provide relevant physical basis for analyzing the influence of applied uniaxial stress on the frequency performance and explain the physical mechanism behind in combination with the simulation results, rather than establish an accurate physical model. The distribution of the applied stress σ_{xx} along the x-direction is not uniform. By fitting the curve of Fig. 4 to $\sigma_{xx}(x)$, the average stress magnitude over the transverse collector length, in the range of 0 to L, can be calculated as

$$<\sigma_{xx}> = \frac{1}{L} \int_0^L \sigma_{xx}(x)dx. \tag{17}$$

By substituting the calculation results of Eq. (17) into Eq. (4), the Ge mole fraction of the uni-biaxially strained base region under additional uniaxial stress conditions can be obtained according to the relationship between the Si$_{1-x}$Ge$_x$ lattice constant and the Ge mole fraction x. The intrinsic carrier concentration of the base region and the diffusion coefficient can then be calculated, and the intrinsic carrier concentration of the SSi cap layer in the emitter region can likewise be obtained [18]. A comparison of the calculated results from Eq. (14) with the simulation results is also exhibited in Fig. 5. It can be found that the theoretical calculation results are basically consistent with the simulation results, which can illustrate the accuracy of the

previous theoretical derivation. The reasons for the difference between the two are: (1) the thickness of the cap layer is so thin that, for the convenience of the analysis, it is approximated that no minority carriers are recombined in this layer and the effect of the thickness of the layer is ignored, (2) the interface states and recombination rate in the oxide-like layer are determined by the specific surface cleaning process, and more detailed physical modeling is dependent on the specific fabrication process, (3) to establish a more complete and accurate physical model of current gain, we also need to solve more complex device equations to calculate the distribution of minority carriers and various microscopic currents

in the emitter region, which are beyond the scope of this work, and (4) the band-gap narrowing (BGN) effect has been ignored in the previous derivation for convenience, while the BGN model has been employed in the TCAD simulation.

Figure 6 gives the variation of the collector current J_C with the BE junction voltage V_{BE} and the Ge mole fraction y. It can be seen that the uniaxial additional stress does increase J_C. According to Eq. (12), it is mainly because of the increase in minority carriers diffusion coefficient and the equivalent Ge mole fraction x in the base region caused by the stress raiser, but there is no change in the order of magnitude. This is due to the fact that the effective voltage V_j falling on the Si/SiGe heterojunction decreases as $\Delta E_{C,SSi}$ increases. J_C is generally enhanced by increasing y. In the collector region, the recombination rate of Si/SiGe interface caused by edge dislocation is enhanced as y increases; hence, the variation of J_C with y is not significant, which is consistent with the experimental conclusion reported in [11]. Moreover, the theoretical results of J_C, also exhibited in Fig. 6, are closer to the simulation results, which is because the electron current injected into the base region is little affected by the barrier height and surface states of the oxide-like layer [15], but mainly depends on the band shift.

Under the small-signal condition, once the common emitter current gain H_{21} and unidirectional transmission power gain U are reduced to 0 dB, the characteristic frequency at a certain collector current can be obtained according to the definitions of f_T and f_{max}. For example, the variation of these two parameters with frequency f is simulated at different V_{BE} conditions for $y = 0.1$, as shown in Fig. 7. By extracting the data from the curves, the frequency characteristic curves for different y can be obtained, as shown in Figs. 8 and 9. It is
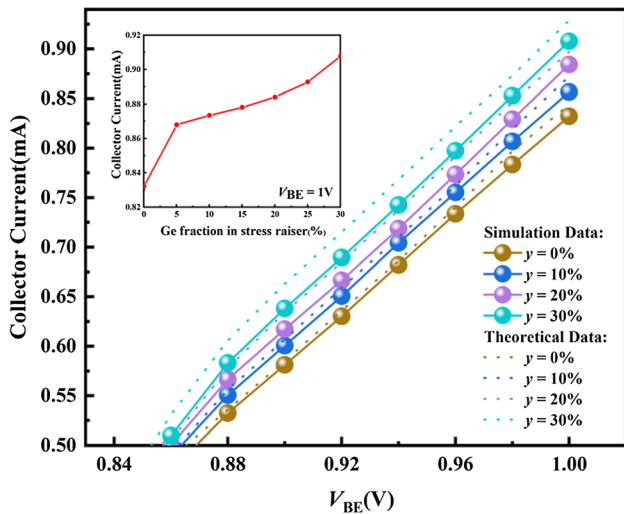
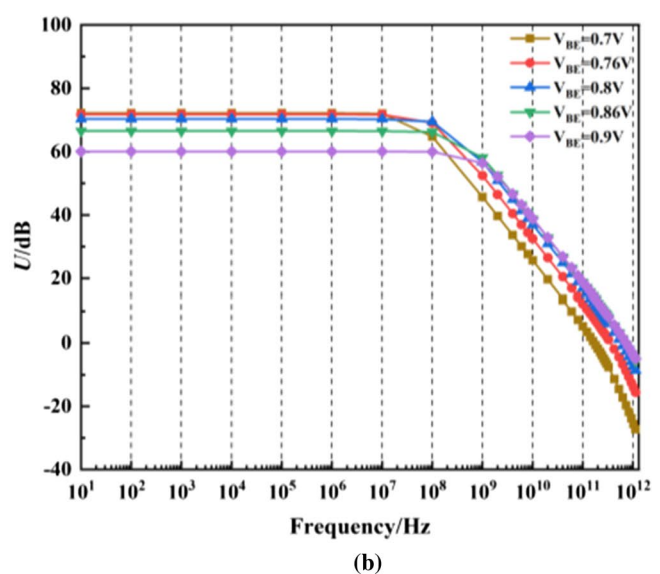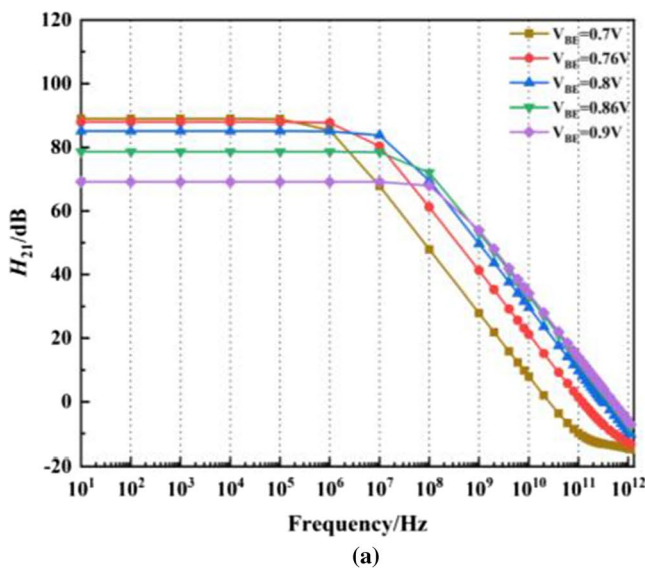**Fig. 6** Variation of collector current with Ge mole fraction in stress raiser

Fig. 7 **a** Variation of small-signal current gain H_{21}, and **b** variation of power gain U, with frequency in unidirectional transmission
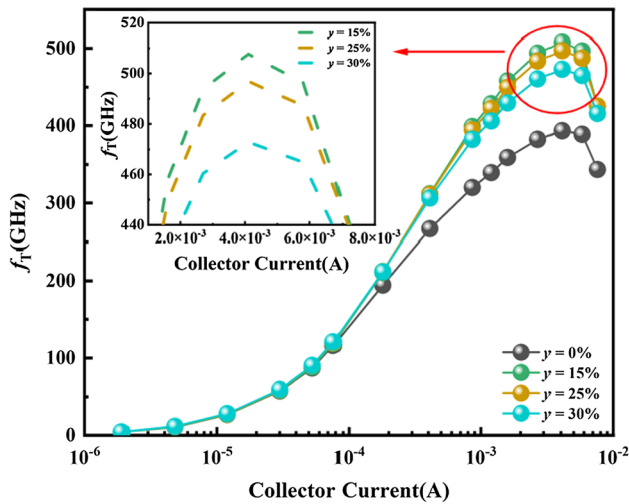
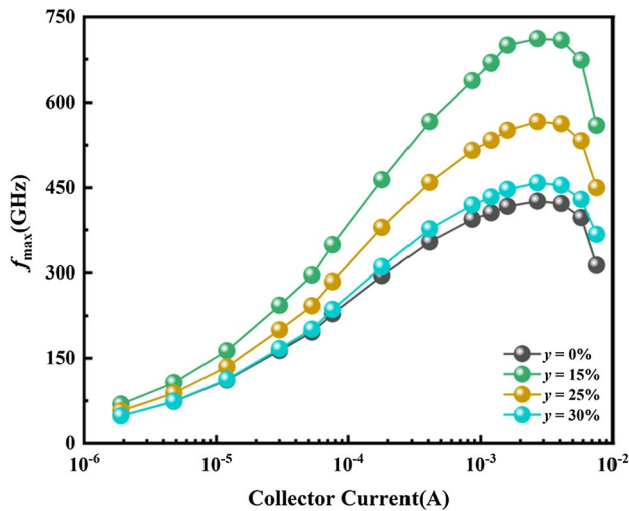**Fig. 8** Cut-off frequency response with different Ge mole fractions in stress raiser

Fig. 9 Maximum oscillation frequency response with different Ge mole fraction in stress raiser

clear that f_T and f_{max} are significantly increased compared to the case without the stress raiser ($y = 0$), by about 29.1% and 71.5%, respectively. The peak values of f_T and f_{max} reached 507.7 GHz and 730.7 GHz for $y = 0.15$ and $y = 0.1$, respectively, entering the "half-THz" frequency band. According to the earlier analysis, the favorable factors leading to an increase in the characteristic frequency as additional stress increases are the increase in current gain, collector current, and interfacial recombination rate. In the base region, the additional uniaxial stress equivalently increases the Ge mole fraction x, resulting an increase in minority carrier diffusion coefficient and a simultaneous decrease in the resistance of the working base region, and in the emitter and collector

regions, the stress increases the carrier mobility and reduces the series resistances.

In Fig. 8, it should be mentioned that f_T increases as y increases from 0 to 0.15 but decreases when y is between 0.15 and 0.3, which is mainly due to the influence of the edge dislocations in the collector region, as described above, on the product $R_C \times C_{BC}$ in Eq. (5). On the one hand, the barrier capacitance C_{BC} is increased because part of mobile charge (J_C/qv_s) is counteracted by dislocations. On the other hand, the distorted lattices become the scattering centers. In the N$^-$-doped collector region, the dislocations as the acceptor centers have anisotropic scattering effect on the carrier, reduces the electron mobility and increases the collector resistance. Therefore, the product $R_C \times C_{BC}$ in Eq. (5) is increases, resulting in a decrease in cut-off frequency.

Figure 9 depicts the variation of the maximum oscillation frequency with collector current for different Ge fraction y. From Eq. (6), the trend is similar to that of the cut-off frequency. At higher strain levels, C_{BC} increases, but f_T and f_{max} increase and then decrease as y increases. Note the change in f_{max} is greater than that in f_T. In addition to the influence of C_{BC}, the base resistance R_B is also another factor. Qualitatively, this is most likely due to the thin base layer, where the dislocations at the Si/Si$_{1-x}$Ge$_x$ interface act as a deep energy level, enhancing the recombination in the base region and reducing the concentration holes. The dislocations scattering mentioned above also has a negative effect on the mobility of holes. The impurity compensation and dislocations scattering effect lead to the increase in R_B, which is greater than that of f_T, thus worsening the f_{max}.

For more visual analysis of the frequency performance, the maximum variation of f_T and f_{max} for different Ge mole fractions of the stress raiser is provided in Fig. 10. Obviously, the Ge fractions corresponding to the maximum of the two are different. When measuring the comprehensive performance of transistors, in fact, the product of some
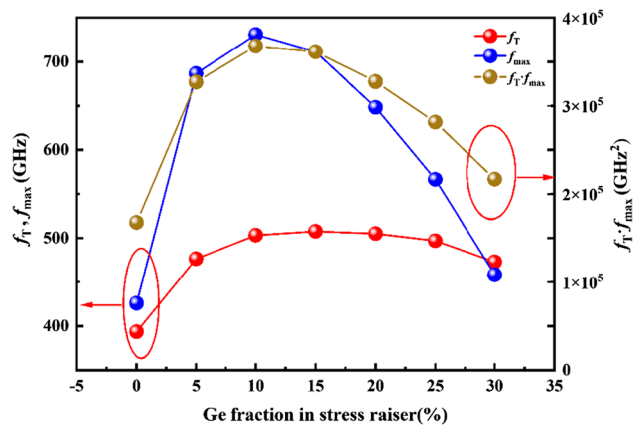

Fig. 10 Defined frequency figure-of-merit with different Ge mole fractions in stress raiser

specific parameters is defined as the figure-of-merit (FoM) considering the correlation or restriction relationship of the parameters. The product $f_T \times f_{max}$ is then used here to define the frequency FoM of the proposed SiGe HBT, so as to determine the fraction in the stress raiser when the best frequency performance is achieved. Also, it is clear that the product $f_T \times f_{max}$ reaches its maximum value at $y = 0.1$, which represents the best frequency characteristics.

As mentioned earlier, the SiGe HBT structure designed in this paper is similar to that of a 90-nm SSi pMOS transistor. In the pMOS transistors, since SiGe stress raiser is

embedded in the source and drain, transverse compressive stress is then induced in the channel, improving the mobility of carriers and enhancing the channel current. Spontaneously, a real-time application of this work is the integration with 90-nm SSi CMOS transistors to form a BiCMOS inverter, which includes one pMOS transistor, two SiGe HBTs and three nMOS transistors, as given in Fig. 11. We now use ATHENA module in TCAD tools to construct the BiCMOS structure, the device-circuit hybrid simulation is carried out by using ATLAS module, and the influence of Ge mole fraction y on the delay time of the proposed SiGe HBT BiCMOS inverter is then explored. Figure 12a shows that the logic function of the inverter is correct, and Fig. 12b shows that the delay time in the falling edge of output signal decreases with the increase in Ge mole fraction of stress raiser used in the proposed SiGe HBT.

As mentioned in Sect. 1, the additional stresses are currently introduced in a variety of ways, and for the purpose of summary and comparison, Table 2 shows the effect of different stress types and different stress raisers on frequency performance in some previous works with similar device structures. The global additional stress used in [12] has the most obvious effect on the improvement of cut-off frequency, because the stress is as high as—1GPa, and it can be predicted that the greater the stress is, the more significant the enhancement may be, which is consistent with the simulated results of this work, but it is not necessarily that the greater the stress is, the better. The reasons are also analyzed before. The additional stresses in the other two references are actually fixed, which depend on the device process parameters. It should be pointed that, however, it is really difficult to estimate which has a better impact on the
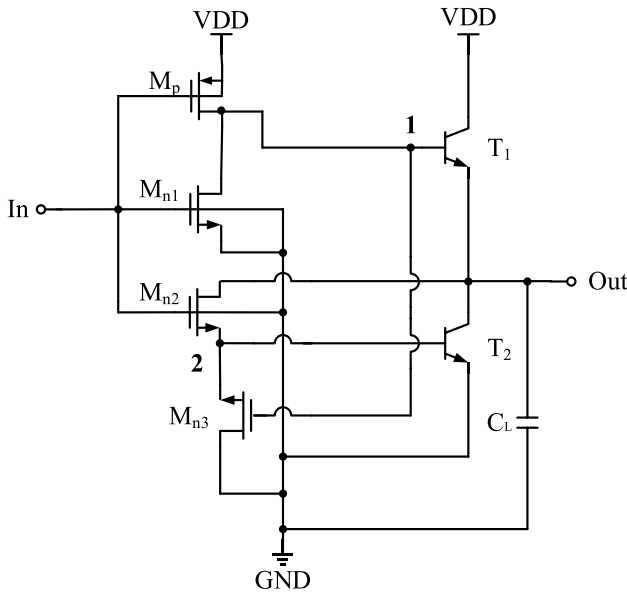

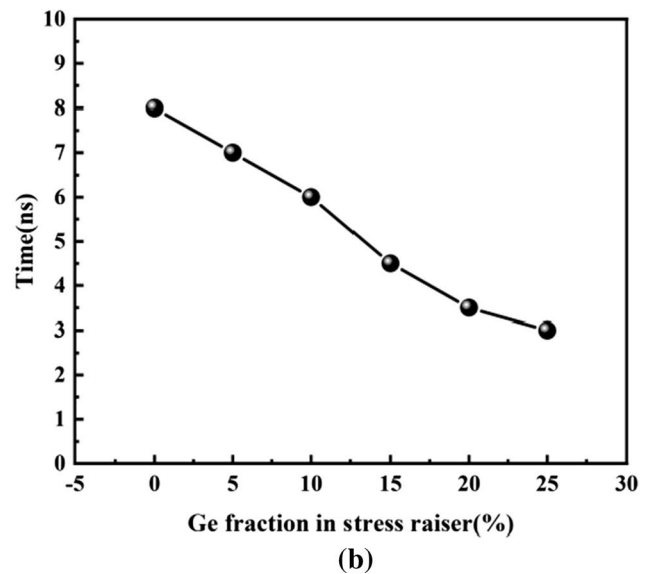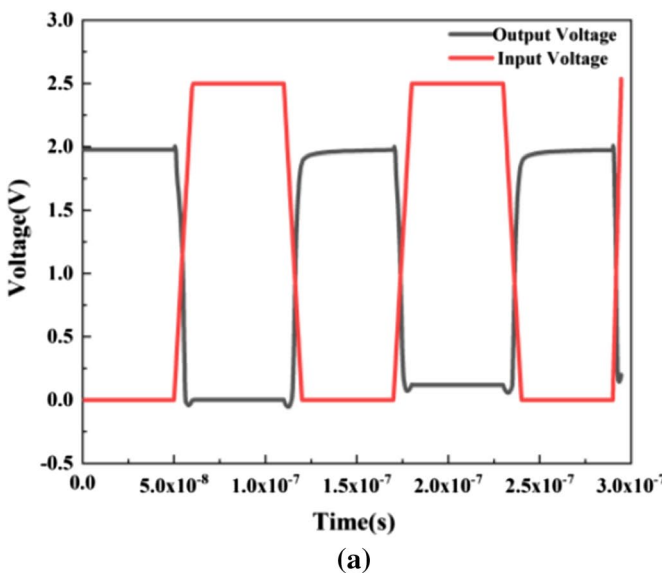Fig. 11 Circuit of BiCMOS inverter used in TCAD simulation

Fig. 12 a Timing sequence response and b delay time of proposed SiGe HBT BiCMOS inverter

Table 2 Improvements of f_T and f_{max} with different stress types and different stress raisers

Reference	Stress raiser	Improvement of f_T	Improvement of f_{max}	Stress action pattern
[9]	Stacked metal layer	21%	12%	Local
[12]	N/A	30%	N/A	Global
[13]	SiO$_2$ STI	14%	9%	Local
This work	Si$_{1-y}$Ge$_y$ layer	29.1%	71.5%	Local

frequency performance because the device structures and process parameters even simulation methods are quite different. From these data available, it seems that embedding a stress raiser inside gives better frequency performance, as the stress acts over a larger scope, and it is believed that, by further improving the manufacturing process, we can obtain even better frequency performance.

5 Conclusion

In this work, a device structure for a scaled SiGe HBT with embedded Si$_{1-y}$Ge$_y$ stress raiser in the collector region is designed. With the additional stress, the theoretical analysis shows that the improvement of the frequency characteristics is mainly due to the change of the strained Si/SiGe band structure, as well as the improvement of the physical parameters such as the current gain, the interface recombination rate and the $R_C \times C_{BC}$ product, while the effect of the dislocation scattering caused by the stress on the relevant physical parameters is also taken into account. The simulation results show that the additional stresses can significantly improve the frequency performance, and the Ge mole fraction in the stress raiser can be flexibly adjusted to achieve the enhancement of the characteristic frequency with different degrees. In particular, the proposed frequency FoM reaches a maximum at a Ge fraction of 0.1, which is similar to that of the SiGe source/drain in an SSi pMOS transistor fabricated with conventional 90-nm technology. The device structure proposed in this work has potential compatibility with Si-base CMOS technology, resulting in a BiCMOS device structure with a smaller size or layout size and better performance. It should be pointed that different additional stress introduction methods have been mentioned in this paper; however, it is really difficult to estimate which has a better impact on the frequency characteristics since the device structures and process parameters with these methods are quite different. It is believed that, by further improving the manufacturing process or device structures of, we can obtain even better frequency performance.

Data availability Data not available at this time as the data also forms part of an ongoing study.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

1. Sengupta, K., Nagatsuma, T., Mittleman, D.M.: Terahertz integrated electronic and hybrid electronic–photonic systems. Nat. Electr. **1**(12), 622–635 (2018)
2. Chantre, A., Chevalier, P., Lacave, T., et al.: Pushing conventional SiGe HBT technology towards" Dotfive" terahertz. The 5th European Microwave Integrated Circuits Conference. 21–24 (2010)
3. www.dotseven.eu.
4. Cressler, J. D.: A retrospective on the SiGe HBT: What we do know, what we don't know, and what we would like to know better. 2013 IEEE 13th topical meeting on silicon monolithic integrated circuits in RF systems. 81–83 (2013)
5. Marani, R., Perri, A.G.: Silicon-Germanium Hbt technology and applications: a review. i-Manager's J. Electr. Eng. **8**(1), 35 (2017)
6. Arturo, S.H., You, S.Z., Van, H.S., et al.: Half-terahertz silicon/germanium heterojunction bipolar technologies: a TCAD based device architecture exploration. Solid-State Electr. **65**, 72–80 (2011)
7. Rücker, H., Heinemann, B.: High-performance SiGe HBTs for next generation BiCMOS technology. Semicond. Sci. Technol. **33**(11), 114003 (2018)
8. Baykan, M.O., Thompson, S.E., Nishida, T.: Strain effects on three-dimensional, two-dimensional, and one-dimensional silicon logic devices: predicting the future of strained silicon. J. Appl. Phys. **108**(9), 093716 (2010)
9. Candele, E., Chevalier, P., Avenier, G., et al.: Impact of BEOL stress on BiCMOS9MW HBTs. IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM). 223–226 (2013)
10. Persson, S., Fjer, M., Escobedo, C.E., et al.: Strained-Silicon Heterojunction Bipolar Transistor. IEEE Trans. Electr. **57**(6), 1243–1252 (2010)
11. Al-Sa'di, M., Fregonese, S., Maneux, C., et al.: TCAD modeling of NPN-SI-BJT electrical performance improvement through SiGe extrinsic stress layer. Mater. Sci. Semiconduc. Process. **13**(5–6), 344–348 (2010)

12. Dinh, T.V., Hong, S.M., Jungemann, C.: Improving the high-frequency performance of SiGe HBTs by a global additional uniaxial stress. Solid-State Electr. **60**(1), 58–64 (2011)

13. Al-Sa'di, M., Fregonese, S., Maneux, C., et al.: Modeling of a novel NPN-SiGe-HBT device structure using strain engineering technology in the collector region for enhanced electrical performance. 2010 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM). 216–219 (2010)

14. Cressler, J. D., Niu, G. F.: Silicon-Germanium Heterojunction Bipolar Transistors. Artech House. (2002)

15. Jin, H.Y., Zhang, L.C., Gao, Y.Z., et al.: An equivalent heterojunction-like model for polysilicon emitter bipolar transistor. Solid-State Electr. **47**(10), 1719–1727 (2003)

16. De Graaff, H.C., De Groot, J.G.: The SIS tunnel emitter: a theory for emitters with thin interface layers. IEEE Trans. Electr. Devices **26**(11), 1771–1776 (1979)

17. Chang, S.T., Liu, C.W., Lu, S.C.: Base transit time of graded-base Si/SiGe HBTs considering recombination lifetime and velocity saturation. Solid-State Electron. **48**(2), 207–215 (2004)

18. Dhar, S., Ungersbock, E., Kosina, H., Grasser, T., Selberherr, S.: Electron mobility model for <110> stressed silicon including strain-dependent mass. IEEE Trans. Nanotechnol. **6**(1), 97–100 (2007)