Modern Trend of High-Speed SiGe Heterojunction Bipolar Transistors (HBTs)

(*Invited Paper*)

Jae-Sung Rieh
Dept. of Electronics and Computer Engineering, Korea University
1, 5-ga Anam-dong, Seongbuk-gu, Seoul 136-701, Korea
jsrieh@korea.ac.kr

Abstract

SiGe HBT technology has emerged as a strong contender for broadband communication applications owing to its excellent RF characteristics as well as the compatibility with the conventional CMOS technology. In this paper, the performance trends of modern SiGe HBT are reviewed in terms of the operation speed, noise, and reliability-related issues such as breakdown voltages and operation current density. Continuous scaling and structural innovations has led to a remarkable improvement in the operation speed and broadband noise characteristics of SiGe HBTs over the past years. Although reduced breakdown voltages and increased current density result from the aggressive scaling, they do not appear as a major obstacle for continued performance improvement of SiGe HBTs in the foreseeable future.

Introduction

As the required bandwidth for communication systems continues to increase, the demand for high-speed semiconductor technologies to implement such systems also increases. While there are a variety of semiconductor technologies commercially available today, the technologies opted for the broadband communication systems have to meet a certain set of criteria. Although such criteria may change over the time depending on the market situation, high-speed operation, low cost, the compatibility with baseband function blocks will remain as the key requirements for the candidate technologies for the broadband applications. SiGe HBT technology has emerged as a strong contender for such applications as it exhibits comparable operation speed with III-V technologies (such as GaAs or InP technologies), while being compatible with the conventional CMOS technology which enables the integration with baseband function blocks.

It is noteworthy that the recent high-speed performance improvement of CMOS technology is remarkable and it has enabled the adoption of CMOS technology for a wide range of RF applications. However, SiGe HBT technology still benefits from various advantages over CMOS technology such as higher transconductance, larger output impedance, better device matching performance, superior 1/f noise characteristics, more favorable linearity, etc. There certainly exist extra cost for SiGe HBT technology compared to baseline CMOS technology, which arises from additional mask levels required for bipolar-related processes. However, it should be noted that the cost for the highly expensive phase-shift mask(s) required for CMOS technology to exhibit performance comparable to SiGe technology is excessive, too. Depending on the wafer volume for a given set of masks, such extra cost for phase-shift mask(s) can be comparable to, or even greater than, the cost for the additional steps for bipolar processes.

In this paper, a brief review of modern SiGe HBT technologies is presented with a major focus on the operation speed, noise, and reliability related issues such as breakdown voltages and operation current density. A historical overview of SiGe technology will be first introduced to provide a brief background on the topic.

A Brief Historical Overview of SiGe HBT Technologies

A SiGe HBT is basically a Si bipolar transistor. The only difference from the conventional Si BJTs is the small amount of Ge incorporated in the base region, which induces a quasi-electric field in the conduction band by a careful tailoring of the composition profile across the base layer. This leads to a reduction in the base transit time of electrons and a significant improvement in the operation speed of the device. Such advantage of Ge incorporation was already expected when the epitaxial growth of the high-quality SiGe layers became available [1].

The first experimental demonstration of SiGe HBT was reported in 1987 by a group of researchers at IBM [2], which was an extension of the effort to further increase the operation speed of Si bipolar transistors for high-speed digital applications. This report has triggered a series of successful results on the performance improvement of SiGe HBTs in the following years, leading to a device exhibiting cutoff frequency f_T exceeding 100 GHz [3]. This success in the early 1990's was, however, followed by a several years of stagnation, which was apparently caused by the rapid replacement of bipolar digital logic with CMOS logic due to the excessive heat generation in highly integrated bipolar chips for digital applications.

Nevertheless, SiGe HBTs were quick to discover another attractive application field, that is, RF and mixed-signal applications, which favor high device speed yet do not require as high a device density as the digital applications. This has fueled another round of research and development activities on SiGe HBTs technologies, leading to the first commercial BiCMOS technology based on SiGe HBTs in 1996 [4]. This successful commercialization has persuaded most of the major companies which had already had Si-only (as opposed to SiGe) bipolar technologies to enter this new arena and develop their own SiGe BiCMOS technologies. As a result, a wide range of commercial applications from multiple companies around the world became available in early 2000's. Continuing efforts to improve the performance of SiGe HBTs, partly caused by the competition between the companies, has recently resulted in a device which demonstrated a f_T higher than 350 GHz [5, 6], which is comparable to the best results obtained from InP-based HBT technologies.

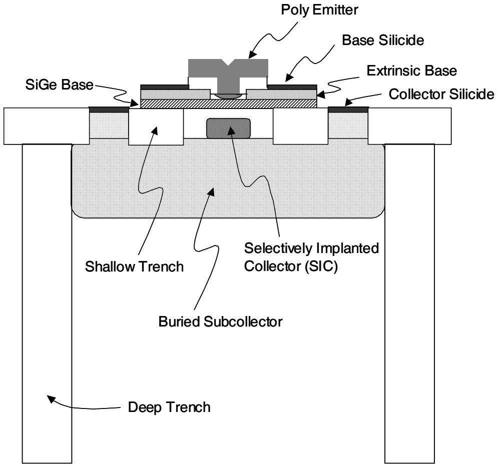

Fig. 1: The schematic cross-section of a typical SiGe HBT.

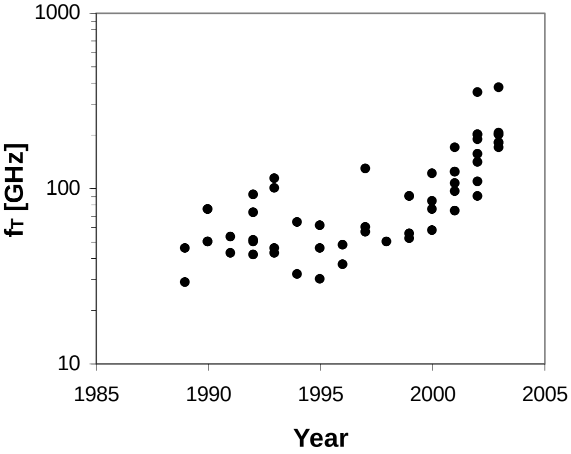

Fig. 2: The trend of the cutoff frequency f_T of SiGe HBTs.

Operation Speed of SiGe HBTs

The enhancement of operation speed of SiGe HBTs has been mostly driven by scaling and structural innovations. Typical approaches for the scaling is base and collector vertical scalings, or a combination of both. Base scaling is achieved by narrowing the boron-doped layer in the base, which results in a reduction in the base transit time. Too narrow base layer, however, may lead to an excessive base resistance and, eventually, a punch-though. Collector scaling is commonly realized by the increase of the doping concentration of selectively implanted collector (SIC), thus reducing the base-collector space charge region and, hence, the transit time across the space charge region. The increased doping concentration near the space-charge region also leads to the increased current density before the onset of the Kirk Effect, which helps to reduce the junction charge times. However, an apparent penalty for the increased collector doping concentration is the lowered breakdown voltage, as will be discussed later in more detail. Structural innovations such as raised extrinsic base [7] or the extension of silicided region toward the device center contribute to the reduction of parasitic resistance/capacitance, which leads to reduced RC delay and improved device operation speed. The cross sectional structure of a typical modern SiGe HBT is illustrated in Fig. 1.

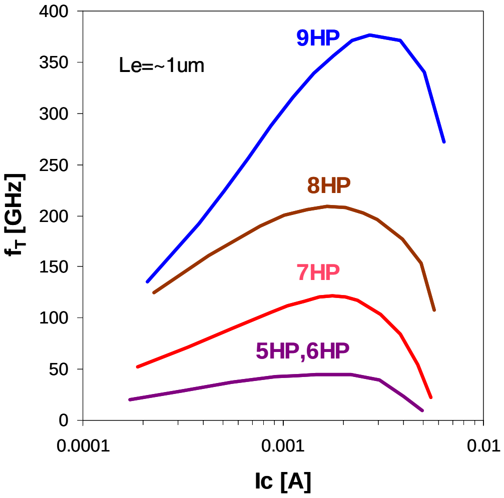

Fig. 3: Evolution of f_T for IBM SiGe HBT technologies.

Based on these approaches, an enormous improvement in the operation speed of SiGe HBTs has been achieved since late 1980's. A trend chart for f_T of SiGe HBTs is provided in Fig. 2, indicating a speed improvement larger than 10× since the first demonstration of the RF performance. A slight saturation during mid-1990's is related to the fading of bipolar logic as mentioned before. The trend chart includes data points for both laboratory-level experimental devices and manufacture-ready production devices. As the impact on the semiconductor industry comes mostly from manufacturable devices, it would be useful to overview the development of production SiGe HBTs which are generally offered as a part of BiCMOS technology. The first commercial SiGe BiCMOS technology was introduced in 1996 by IBM [4]. This 0.5 μm BiCMOS technology offers SiGe HBTs exhibiting f_T and f_{max} (maximum oscillation frequency) of 47 GHz and 65 GHz, respectively. This was followed by 0.25 μm BiCMOS technology with similar HBT performance, but with enhanced CMOS offering [8]. While both of these early technologies were suitable for applications at less than 5 GHz, the subsequent 0.18 μm technology was intended for increased frequency operation such as 40 Gb/s wireline communication applications, with f_T and f_{max} both exceeding 100 GHz [9]. The speed enhancement was primarily achieved through vertical and lateral scaling, resulting in the reduction of transit time and parasitic resistance and capacitance. Continuing scaling and structural innovation led to a 200 GHz technology based on 0.13 μm lithographical node [7]. The latest vertical scaling effort with 0.13 μm node was marked by a record f_T higher than 350 GHz [5, 6]. Presented in Fig. 3 is f_T of each generation plotted against the collector current normalized to unit emitter length.

Broadband Noise of SiGe HBTs

The broadband noise of an active device can be characterized by four noise parameters: minimum noise figure (F_{min}), noise resistance (R_n) and the real and imaginary components of the source impedance match for lowest noise

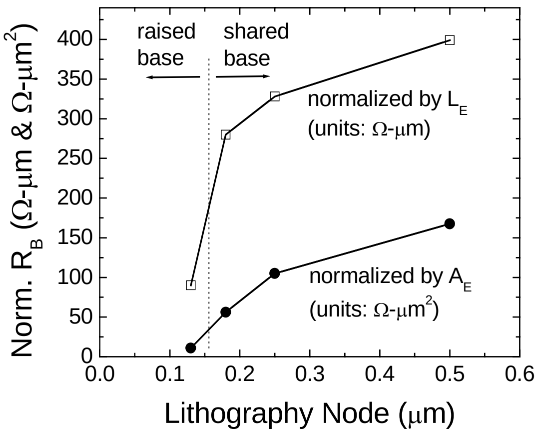

Fig. 4: Base resistance, R_B (normalized by both emitter length and area) vs. lithography node for four generations of IBM SiGe HBT technology [10].

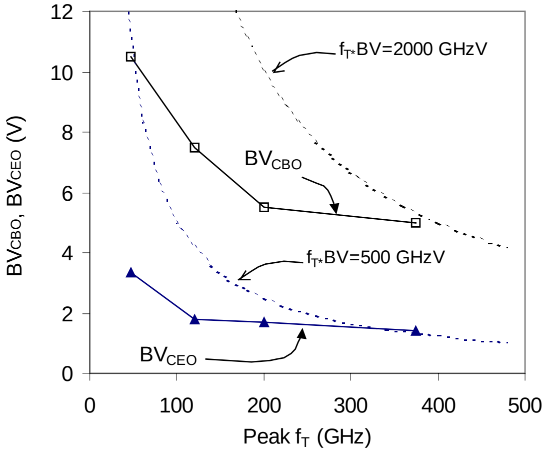

Fig. 6: Trend of BV_{CEO} and BV_{CBO} with increasing f_T. Both breakdown voltages show a signature of saturation. Shown as dotted lines are constant f_T-BV product contours.

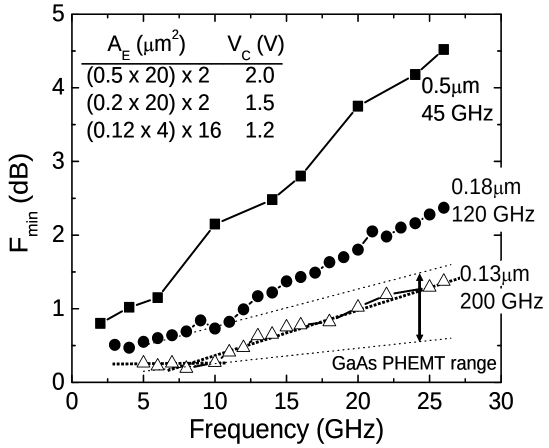

Fig. 5: F_{min} vs. frequency for 3 generations of SiGe BiCMOS technology, with the range of reported GaAs PHEMT values included for reference.

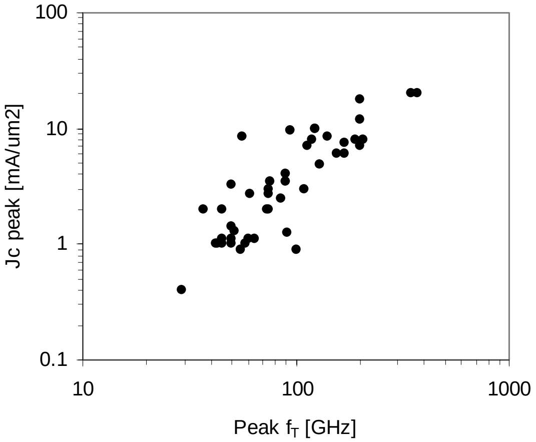

Fig. 7: Trend of J_C - f_T relation for SiGe HBTs.

(Γ_{opt}). Among these parameters, F_{min} is considered as the key parameter as it determines the floor level of the broadband noise. The most influential device parameters that modulate F_{min} of bipolar transistors are the base resistance R_B and f_T. While base vertical scaling is effective for reducing transit time as described before, it inevitably leads to a degradation in R_B as the base sheet resistance is increased. In order to maintain or even improve R_B, lateral scaling needs to accompany the vertical scaling so that the lateral current path along the resistive base layer can be shortened. Additional structural innovation such as raised extrinsic base would help a further reduction of R_B.

Fig. 4 plots R_B (normalized against both emitter length and area) against lithographic node across four generations of IBM SiGe BiCMOS technology [10]. It shows a significant reduction of R_B over the generations, which is achieved primarily from lateral scaling and raised extrinsic base. This decreasing trend of R_B, together with the increasing trend of f_T, leads to a significantly improved broadband noise property as the technology generation evolves. Figure 5 compares F_{min} vs. frequency for three generations of SiGe HBT technology ranging from 0.5 μm down to 0.13 μm. It is apparent from the plot that F_{min} drops with each generation over the entire range of measurement frequency. Also shown as a reference in the figure is the range of reported F_{min} values for GaAs PHEMT devices, which indicates that Si-based technology can now attain the levels of noise performance previously associated with more exotic processes.

Breakdown Voltages and Current Density of SiGe HBTs

As devices are aggressively scaled for the operation speed enhancement, the electric field at the junctions tends to increase, leading to reduced breakdown voltages. The trend of breakdown voltages of SiGe HBTs with increasing device speed is shown in Fig. 6 [10], taken from four generations of IBM SiGe technology. Both BV_{CEO} (open-base C-E breakdown voltage) and BV_{CBO} (open-emitter C-B breakdown voltage) continue to decrease with increasing f_T, as a result of the increased electric field at the base-collector junction. However, they show a signature of saturation as f_T enters the multi-hundred GHz regime. This is an important trend for device design, since it indicates that scaling can be continued

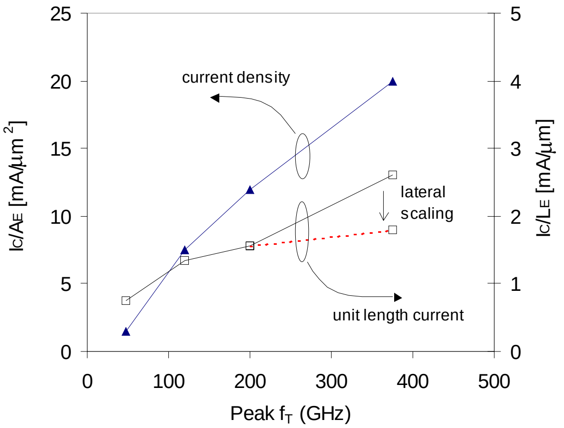

Fig. 8: Trend of collector current density (I_C / A_E) and unit length current (I_C / L_E) with f_T. Unit length current increases more slowly than current density with increasing device speed. The dotted line shows the unit length current predicted with appropriate lateral scaling for f_T = 375GHz point.

for speed enhancement without a significant penalty on breakdown voltages. Also plotted in the figure as dotted lines are the constant f_T – breakdown voltage product contours, which clearly indicates the increase of such product with the generation. It is interesting to note that BV_{CEO} does not serve as the upper limit of voltage allowed across the collector and the emitter, since the open-base configuration, which is used for BV_{CEO} measurement, is the worst case for C-E breakdown voltages and rarely happens in actual circuits. In fact, such upper limit is rather given by BV_{CBO}, which is typically a few times larger than BV_{CEO} as shown in Fig. 6.

The increase of collector doping concentration, which is routinely employed for collector vertical scaling, is usually accompanied by an increase in the operation current density. This is due to the delayed onset of Kirk effect as mentioned earlier. The increasing trend of current density with speed enhancement is illustrated in Fig. 7, which shows data for J_C - f_T relation accumulated over an extensive period. Raised current density may have implications on the electromigration, self-heating, device robustness, etc. In particular, the electromigration issue has surfaced as a practical problem for optimized interconnection wiring of devices as the current level continues to rise. Fortunately, the lateral scaling that often accompanies the vertical scaling helps to alleviate the electromigration effect. In typical wiring schemes, the current capacity of metal feeding lines is given by the current per unit length of the emitter finger (unit length current), rather than the current density itself. It is obvious that the unit length current decreases with decreasing emitter width for a fixed current density (current per emitter area). As the emitter width tends to decrease over the technology generations, the unit length current exhibits much slower increase than the current density, as is clearly shown in Fig. 8 for four generations of IBM SiGe BiCMOS technology [10]. Note from the plot that lateral scaling for f_T = 375 GHz case (shown as dotted line) would lead to a near-flat trend of the unit length current.

Conclusions

The performance trends of modern SiGe HBTs have been overviewed with a focus on the operation speed, noise, breakdown voltages, and operation current density. Continuous scaling and structural innovations have led to a remarkable improvement in the operation speed and broadband noise characteristics of SiGe HBTs over the past years. Although issues related to reduced breakdown voltages and increased current density arise along with such performance improvements, there seems to exist viable ways to suppress or circumvent such issues. Therefore, the performance of SiGe HBTs will continue to follow the current trends without a major show-stopper for the foreseeable future.

References

[1] R. People, “Physics and applications of Ge_xSi_{1-x}/Si strained-layer heterostructures,” *IEEE Journal of Quantum Electronics*, vol. 22, no. 9, pp. 2899-2907, Sept. 1986.

[2] S. S. Iyer *et al.* "Silicon-Germanium base heterojunction bipolar transistors by molecular beam epitaxy," *Techcinal Digest of International Electron Device Meeting*, pp. 874-876, 1987.

[3] E. Crabbe *et al.* "113-GHz f_T graded-base SiGe HBT's," *IEEE Transactions on Electron Devices*, vol. 40, no. 11, pp. 2100-2101, Nov. 1993.

[4] D. Ahlgren *et al.*"Manufacturability demonstration of an integrated SiGe HBT technology for the analog and wireless marketplace," *Technical Digest of the International Electron Device Meeting*, pp. 859-862, 1996.

[5] J.-S. Rieh *et al.* "SiGe HBTs with cut-off frequency of 350 GHz," *Technical Digest of International Electron Devices Meeting*, pp. 771-774, 2002.

[6] J.-S. Rieh *et al.*"Performance and design considerations for high speed SiGe HBTs of f_T/f_{max}=375GHz/210GHz," *International Conference on Indium Phosphide and Related Materials*, pp. 374-377, 2003.

[7] B. Jagannathan *et al.* "Self-aligned SiGe NPN transistors with 285 GHz f_{max} and 207 GHz f_T in a manufacturable technology," *IEEE Electron Device Letters*, vol. 23, no. 5, pp. 258-260, May 2002.

[8] S. A. St. Onge *et al.* "A 0.24 μm SiGe BiCMOS mixed-signal RF production technology featuring a 47 GHz f_T HBT and 0.18 μm L_{eff} CMOS," *Proceedings of Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 117-120, 1999.

[9] A. Joseph *et al.* "A 0.18μm BiCMOS Technology Featuring 120/100 GHz (f_T/f_{max}) HBT and ASIC-compatible CMOS using copper interconnect," *Proceedings of Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 143-146, 2001.

[10] J.-S. Rieh *et al.* "SiGe heterojunction bipolar transistors and circuits toward terahertz communication applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 10, pp. 2390-2408, Oct. 2004.