Noise in SiGe HBTs: Opportunities and Challenges

John D. Cressler

School of Electrical and Computer Engineering, 777 Atlantic Drive, N.W.
Georgia Institute of Technology, Atlanta, GA, 30332-0250 U.S.A.

ABSTRACT

SiGe technology represents a remarkable success story for the microelectronics industry, and possesses the capability to fundamentally reshape the way broadband communications systems are conceived and built in the 21st century. From the first demonstration of a functional SiGe HBT in 1987, until the achievement of the present performance record of 375 GHz peak cutoff frequency, a mere 18 years has elapsed! The SiGe HBT is the first practical bandgap-engineered Si device, and has evolved from simple transistor and circuit demonstrations in a select few research laboratories to robust production in upwards of two-dozen manufacturing facilities around the world in 2005, and commercial products abound across a wide spectrum of commercial applications. This paper reviews the state-of-the-art in SiGe technology, discusses the design and operational principles of SiGe HBTs, and then focuses on the broadband and low-frequency noise characteristics of SiGe HBTs, emphasizing both the opportunities and the challenges which will necessarily be faced with continued device scaling.

Keywords: SiGe HBT, silicon-germanium, heterojunction bipolar transistor, noise, 1/f noise, broadband noise

1. INTRODUCTION

During the past several years silicon-germanium heterojunction bipolar transistor (SiGe HBT) technology has entered the global semiconductor electronics market with a bang, and SiGe HBT technologies are being increasingly deployed by a host of companies in the North America, Europe, and the Far East for a wide variety of communications circuit applications.[1–3] Progress in SiGe HBT device performance has proceeded at a truly dizzying pace. SiGe HBT's offer transistor-level performance metrics which are competitive with the best III-V technologies (GaAs or InP), while maintaining strict fabrication compatibility with low-cost, high yield, Si CMOS foundry processes on large (200 mm) wafers. Thus, SiGe HBT technology effectively marries III-V-like device performance with the compelling economy-of-scale of Si integrated circuit manufacturing.

In 2005, robust (meaning manufacturable, at high yield) 1st generation (50 GHz peak f_T) SiGe HBT technologies exist in upwards of two dozen companies worldwide, 2nd generation (100 GHz peak f_T) SiGe technologies in 8-10 companies, and 3rd generation (200 GHz peak f_T) SiGe technologies in 4-5 companies (Figure 1).[4] The number of SiGe "players" is thus climbing rapidly worldwide. As almost universally practiced today, SiGe technology exists in a BiCMOS implementation (SiGe HBT + Si CMOS), and as an "adder" module to a core deep-submicron CMOS technology, which conveniently facilitates using the SiGe HBT where it offers the most advantage (i.e., in mixed-signal circuits: analog, RF, microwave, mm-wave, and/or very high-speed digital circuits), and using the Si CMOS to its strongest advantage in highly-integrated, lower performance memory and digital circuits. The commercial market for such mixed-signal ICs is exploding, in the ever-expanding quest to build the requisite electronic infrastructure for the emerging 21st century communications revolution.

E-mail: cressler@ece.gatech.edu / Tel: (404) 894-5161 / Fax: (404) 894-4641

Noise in Devices and Circuits III, edited by Alexander A. Balandin, François Danneville,
M. Jamal Deen, Daniel M. Fleetwood, Proc. of SPIE Vol. 5844 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.608773

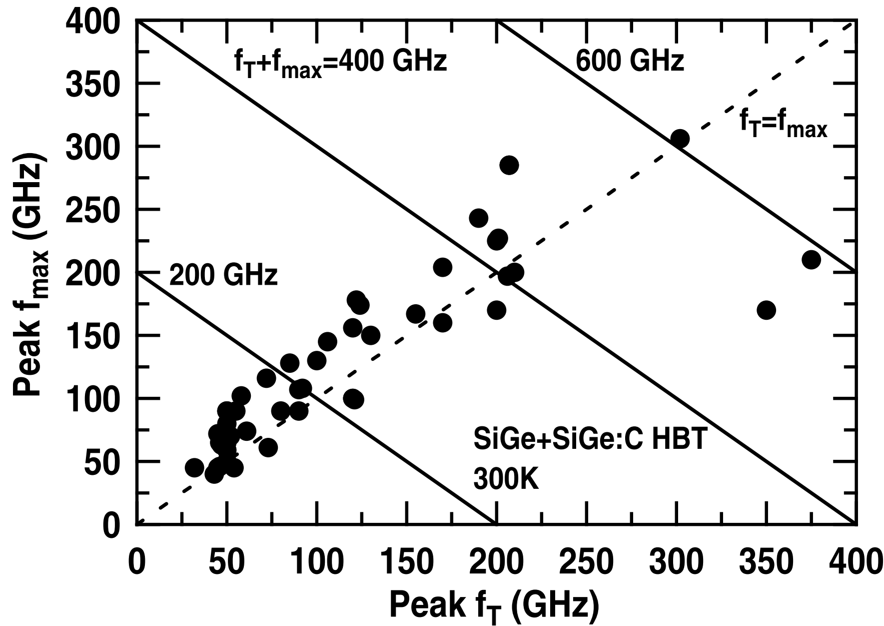

Figure 1. Measured maximum oscillation frequency versus cutoff frequency for a variety of SiGe HBT technology generations.

2. SIGE HBT TECHNOLOGY

While the specific composition of a given SiGe HBT clearly depends on the technology-generation, and obviously the company manufacturing it, a representative state-of-the-art SiGe HBT BiCMOS technology integrates a multiple SiGe HBTs having different breakdown voltages, with conventional silicon CMOS devices. Such SiGe HBT BiCMOS technologies are 100% silicon processing compatible, employ deep- and shallow-trench isolation, a polysilicon emitter contact, and a carbon-doped (to decrease the boron diffusivity), graded epitaxial SiGe base which is thermodynamically stable as deposited, and typically grown by some variant of a chemical vapor deposition process (e.g., UHV/CVD or RPCVD). Such SiGe technologies are designed to be high-yielding and manufacturable, and often achieve transistor chain yields (e.g., 5,000 SiGe HBTs wired in parallel) in excess of 90%. Figure 2 shows a schematic cross-section of a 1st generation SiGe HBT.

A representative 1st generation doping and Ge profile is shown in Figure 3. In this case, the metallurgical base width is about 90 nm (about 65-nm neutral base width under forward-active bias), the metallurgical emitter junction depth is about 35 nm (from the Si surface), and the peak Ge content is about 8% (it is thermodynamically stable). The emitter polysilicon layer is doped to its solid-solubility limit, multiple self-aligned phosphorus implants are used to locally tailor the collector doping profile, and the peak base doping is about $4x10^{18}$ cm^{-3}. The Ge profile is trapezoidal in shape, with substantial grading across the neutral base. This vertical profile design can be considered quite conservative by today's standards, but it nonetheless achieves a peak f_T of 50 GHz (70-GHz peak f_{max}) at a BV_{CEO} of 3.3 V, solidly in the range of a first generation SiGe technology). Cross-company typical profile numbers for 1st generation SiGe technologies are: $W_{b0} = 60 - 90$ nm, $W_e = 20 - 40$ nm, peak Ge = 8–12%.

In 3rd-generation SiGe technology, an improvement in f_T to 200 GHz (and above) is typically realized only through fundamental changes in the physical structure of this 1st generation transistor. Specifically, a reduced thermal cycle "raised extrinsic base" structure is implemented using conventional deep and shallow trench isolation, and an in-situ doped polysilicon emitter. The key features of this more advanced structure include: 1) a removal of any additional extrinsic base ion implantation, which generally produces an undesirable enhanced base dopant diffusion, and 2) the physical relocation of the extrinsic base-collector junction, making reduction of collector-base capacitance easier. The entire thermal cycle of required to build the device is reduced. Those acquainted with Si BJT technologies will recognize the similarity in doping profiles between this SiGe HBT and advanced ion-implanted Si BJTs (just removing the Ge makes it look like a high-speed Si BJT). The key difference between this SiGe HBT and a conventional ion-implanted double-poly Si BJT lies in the base

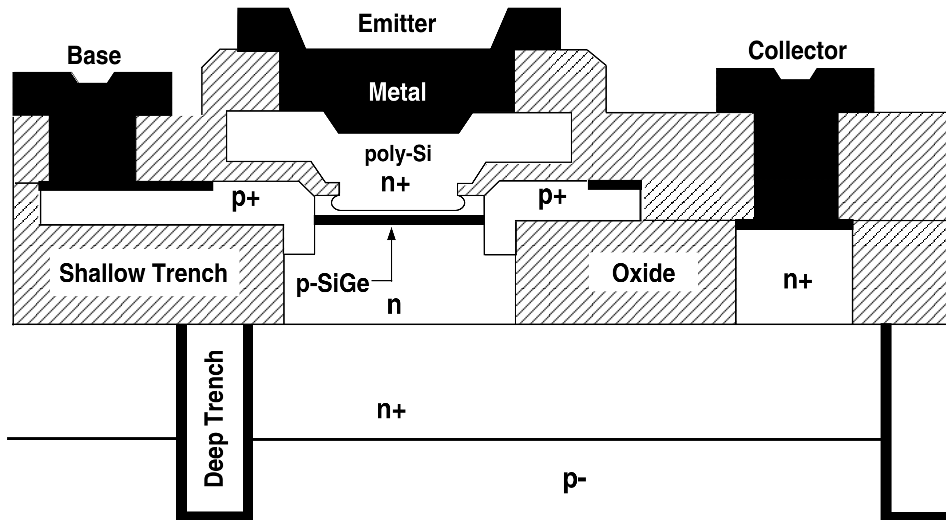

Figure 2. Schematic device cross-section of a first-generation SiGe HBT.

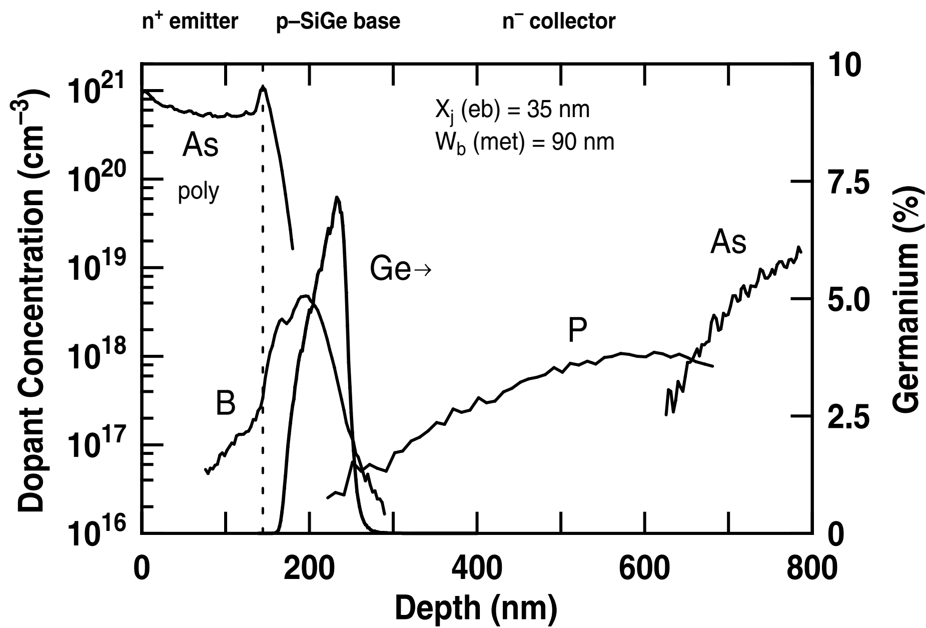

Figure 3. Representative first-generation SiGe HBT doping profile, as measured by SIMS.

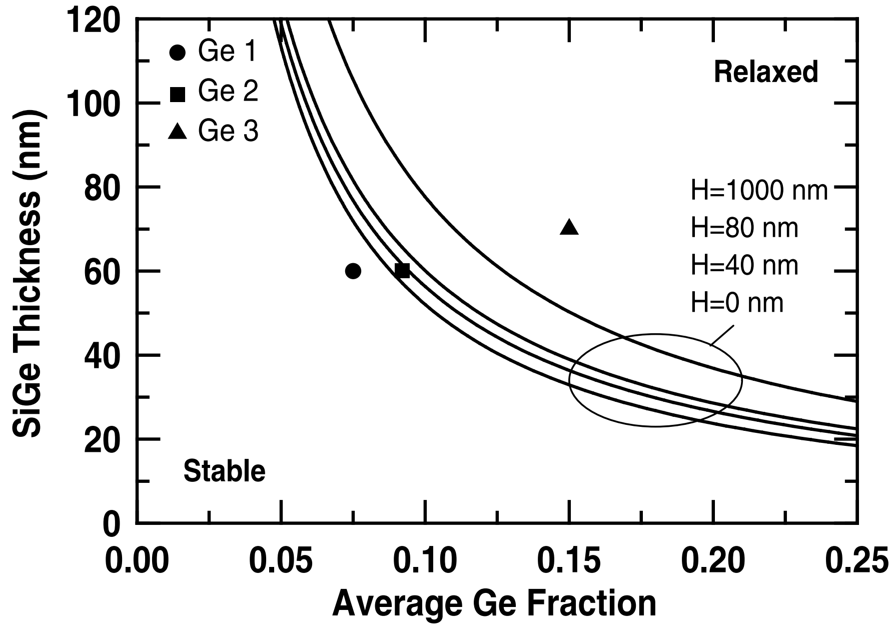

Figure 4. Theoretical stability constraint curves as a function of average Ge fraction and Si "cap" thickness (H). Also shown are several representative SiGe profiles, two of which are thermodynamically stable, and one which is metastable, and hence will relax during fabrication.

profile, which can be much more heavily doped at a given base width using epitaxial growth (leading to much lower base resistance, better dynamic response, and lower broadband noise).

Regardless of the SiGe growth technique used, the structure, the self-alignment scheme employed in fabricating the transistor, or technology generation, strained SiGe films found in today's commercially viable SiGe HBTs all have a similar form. The deposited SiGe film actually consists of a three-layer composite structure: 1) A thin, undoped Si buffer layer, 2) the actual boron-doped SiGe active layer, and 3) a thin, undoped Si cap layer. The thickness of the SiGe-bearing layer is clearly a key variable in SiGe HBT device design. The maximum thickness for obtaining pseudomorphic growth post-fabrication (i.e., after any thermal anneals or ion-implantation steps which might relax an overstable film) is known as the "critical thickness," and the dependence of the SiGe film thickness on average Ge film content is known as the SiGe "stability constraint curve." Recent theoretical approaches[5] which properly account for effects of the Si cap layer on film stability show good agreement between stability calculations and data for real SiGe films used in practical SiGe HBTs. Representative SiGe stability constraint curves is shown in Figure 4.

3. SIGE HBT DEVICE PHYSICS

The introduction of Ge into the base region of a bipolar transistor has two tangible *dc* consequences, and can best be appreciated by viewing an energy band diagram (Figure 5): 1) the potential barrier to injection of electrons from emitter into the base is decreased. Intuitively, this will yield exponentially more electron injection for the same applied V_{BE}, translating into higher collector current and hence higher current gain (β) in the device, provided the base current remains unchanged. Of great practical importance, the introduction of Ge effectively decouples the base doping from the current gain, thereby providing device engineers with much greater design flexibility than in Si BJTs. 2) The presence of a finite Ge content in the CB junction will positively influence the output conductance of the transistor, yielding higher Early voltage (V_A). For an identically-constructed (doping levels, layout, emitter contact, etc.) SiGe HBT and Si BJT, we find:[1]

$$\left.\frac{\beta_{SiGe}}{\beta_{si}}\right|_{V_{BE}} \simeq \left\{ \frac{\gamma\,\eta\,\Delta E_{g,Ge}(grade)/kT\; e^{\Delta E_{g,Ge}(0)/kT}}{1 - e^{-\Delta E_{g,Ge}(grade)/kT}} \right\}, \tag{1}$$

Figure 5. Energy band diagram for a Si BJT and graded-base SiGe HBT, both biased in forward active mode at low-injection.

where we have defined a minority electron diffusivity ratio between SiGe and Si as

$$\eta = \frac{(D_{nb})_{SiGe}}{(D_{nb})_{Si}}, \tag{2}$$

and an "effective density-of-states ratio" between SiGe and Si according to

$$\gamma = \frac{(N_C N_V)_{SiGe}}{(N_C N_V)_{Si}} < 1. \tag{3}$$

The Ge-induced reduction in the base bandgap occurring at the emitter-base edge of the quasi-neutral base is $\Delta E_{g,Ge}(x = 0)$, and

$$\Delta E_{g,Ge}(grade) = \Delta E_{g,Ge}(W_b) - \Delta E_{g,Ge}(0), \tag{4}$$

where W_b is the quasi-neutral base width.

2) The output conductance ratio (as reflected in the Early voltage) between a SiGe HBT and a Si BJT exponentially depends on the amount of bandgap grading across the base divided by kT,

$$\left.\frac{V_{A,SiGe}}{V_{A,Si}}\right|_{V_{BE}} \simeq e^{\Delta E_{g,Ge}(grade)/kT} \left[\frac{1 - e^{-\Delta E_{g,Ge}(grade)/kT}}{\Delta E_{g,Ge}(grade)/kT}\right], \tag{5}$$

Combining eqns. (1) and (2) yields the key analog figure-of-merit (βV_A).

$$\frac{\beta V_{A,SiGe}}{\beta V_{A,Si}} = \widetilde{\gamma}\widetilde{\eta}\, e^{\Delta E_{g,Ge}(0)/kT} e^{\Delta E_{g,Ge}(grade)/kT}. \tag{6}$$

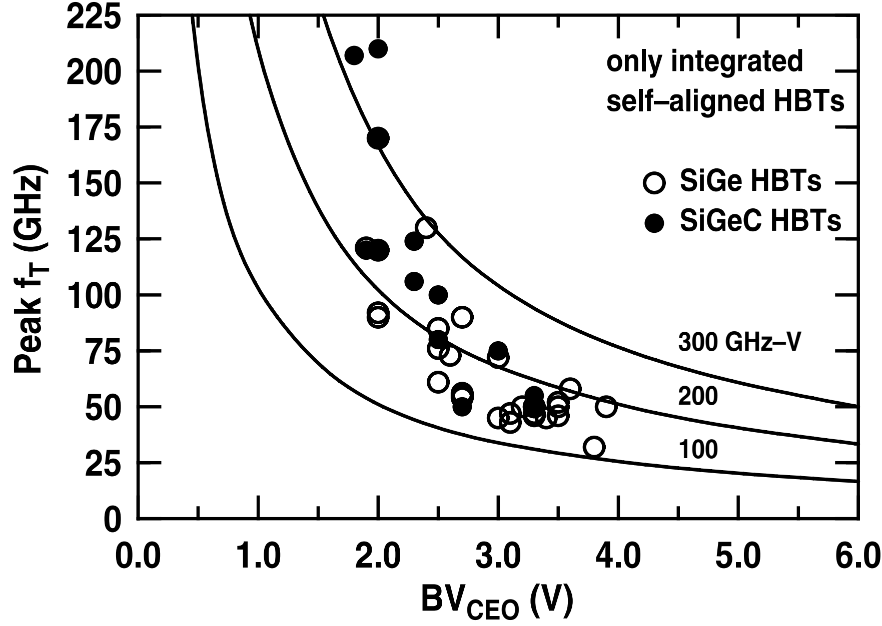

Figure 6. Peak unity gain cutoff frequency as a function of open-base collector-to-emitter breakdown voltage for self-aligned SiGe and SiGeC HBTs.

To understand the dynamic response of the SiGe HBT, and the role Ge plays in transistor frequency response, we must first formally relate the changes in the base transit time and emitter transit time to the physical variables of this problem. The base transit time ratio, which typically limits the achievable frequency response ($f_T \propto 1/\tau_b$), is given by

$$\frac{\tau_{b,SiGe}}{\tau_{b,Si}} = \frac{2}{\eta} \frac{kT}{\Delta E_{g,Ge}(grade)} \cdot \left\{ 1 - \frac{kT}{\Delta E_{g,Ge}(grade)} \left[1 - e^{-\Delta E_{g,Ge}(grade)/kT} \right] \right\}. \tag{7}$$

The increase in current gain also has a favorable impact on the f_T of the transistor, since the emitter delay time can be a non-negligible component of the overall emitter-to-collector delay ($\tau_e \propto 1/\beta$), such that

$$\frac{\tau_{e,SiGe}}{\tau_{e,Si}} \simeq \frac{J_{C,Si}}{J_{C,SiGe}} = \frac{1 - e^{-\Delta E_{g,Ge}(grade)/kT}}{\widetilde{\gamma\eta}\, \frac{\Delta E_{g,Ge}(grade)}{kT}\, e^{\Delta E_{g,Ge}(0)/kT}}, \tag{8}$$

Both τ_b and τ_e couple to the overall unity-gain cutoff frequency (f_T) according to

$$f_T = \frac{1}{2\pi} \left[\frac{kT}{qI_C} (C_{te} + C_{tc}) + \tau_b + \tau_e + \frac{W_{CB}}{2v_{sat}} + r_c C_{tc} \right]^{-1}. \tag{9}$$

These Ge-induced improvements in the base and emitter transit times, when coupled to a laterally scaled device structure, translate into significant improvements in the maximum oscillation frequency (f_{max}) of the SiGe HBT as well, which is a more relevant figure-of-merit for most mixed-signal circuit applications. It is important to note that with technology scaling, the average collector current density at which peak f_T (or f_{max}) occurs increases rapidly. For a typical first-generation (50 GHz) SiGe HBT, peak f_T might be at J_C = 1-2 $mA/\mu m^2$, whereas it might be 8-10 $mA/\mu m^2$ for

second-generation (120 GHz) technology, and 15-20 $mA/\mu m^2$ for third-generation (200 GHz) technology. This increase in operating current density is the result of the need to delay the onset of both Kirk effect (base push-out) and high-injection heterojunction barrier effects,[1] which, if unchecked, will quickly limit the maximum performance attainable. Note, however, that this increase in current density is usually obtained at a overall decrease in operating current, since lateral scaling produces smaller emitter geometries, and thus thermal effects (driven by I_C and not J_C) do not necessarily worsen with scaling.

An unfortunate consequence of optimizing SiGe HBTs for higher peak f_T (via higher collector doping) is a strong increase in impact-ionization in the base-collector space-charge-region. Thus, an inherent (and well known) tradeoff exists between peak f_T and BV_{CEO} in SiGe HBT device design, often referred to as the "Johnson limit" (Figure 6).[6] This Johnson limit, however, is actually more accurately described by the (larger) $BV_{CES} f_T$ product (BV_{CES} is about equal to BV_{CBO}) than the traditional $BV_{CEO} f_T$ product, but clearly the ever-decreasing operational voltage limits of scaled SiGe HBTs pose important (and often non-obvious) constraints on the biasing and operation of SiGe HBTs used in mixed-signal circuits. The understanding, for instance, of how much "usable" voltage actually exists in the region between BV_{CEO} and BV_{CES} remain active research topics, particularly when considering the complex interactions in breakdown between impact-ionization, self-heating, and avalanche-induced, current-crowding instabilities (often referred to as "pinch-in" effects), and their corresponding dependence on operational current density within a given circuit topology.

4. NOISE IN SIGE HBTS

One of the most desirable attributes of SiGe HBTs is their low noise capability.[1] To illustrate the importance of noise, consider a wireless transceiver application. The lower limit of dynamic range is set by the system noise, while the upper limit is set by the system linearity. The first stage of the receiver, typically an LNA, must amplify signals as low as -100 dBm, while maintaining adequate signal-to-noise ratio. The amount of noise added by the LNA (as measured by the noise figure) must thus be sufficiently low. The receiver must also be highly linear in order to handle signal levels which fluctuate between 20–40 dB from the mean value as a function of time. The predominant contributor to the LNA noise figure is the RF noise of the transistor. Another key concern in RF circuits is "phase noise." In phase noise, the transistor low-frequency noise is upconverted to RF frequencies through the device nonlinearities, producing phase noise in local oscillator (LO) reference signals. The LO phase noise mixes with the RF signal, thus broadening the signal in the intermediate frequency band and degrading signal integrity. The minimum channel spacing of the receiver is thus limited by the LO phase noise, and ultimately by the low-frequency noise of the transistor. We will now address both broadband and low-frequency noise in SiGe HBTs.

4.1. Broadband Noise

To gain intuitive insight into broadband noise in SiGe HBTs, analytical expressions for the broadband noise parameters (NF_{min}, R_n, Y_{opt} and G_A^{ass}) are obviously desirable. This can be accomplished using Y-parameters. We begin from the power spectral densities of the input noise current (S_{i_n}), the input noise voltage (S_{v_n}), and their cross-correlation ($S_{i_n v_n^*}$), as obtained generally from linear noisy two-port theory,[1] and then relate the transistor Y-parameters to the physical device parameters, using: $g_m = qkT/I_C$ and $C_i = C_{be} + C_{bc}$. Here, C_{be} consists of the emitter-base (EB) depletion capacitance C_{te} and the EB diffusion capacitance $g_m\tau$ ($C_{be} = C_{te} + g_m\tau$), with τ being the device transit time, and C_i is related to f_T through $f_T = g_m/2\pi C_i$. Thus, S_{i_n}, S_{v_n}, and $S_{i_n v_n^*}$ can then be expressed in terms of I_C (or g_m), β, C_i, and r_b to obtain the (approximate) broadband noise parameters: 1) the noise resistance R_n, 2) the optimum source termination admittance $Y_{s,opt}$, 3) the noise figure NF_{min}, and 4) the associated gain G_A^{ass}.

$$R_n = \frac{S_{vn}}{4kT} = r_b + \frac{1}{2g_m} \tag{10}$$

$$Y_{s,opt} = G_{s,opt} + jB_{s,opt} \tag{11}$$

where

$$G_{s,opt} = \sqrt{\frac{g_m}{2R_n}\frac{1}{\beta} + \frac{(\omega C_i)^2}{2g_m R_n}\left(1 - \frac{1}{2g_m R_n}\right)} \tag{12}$$

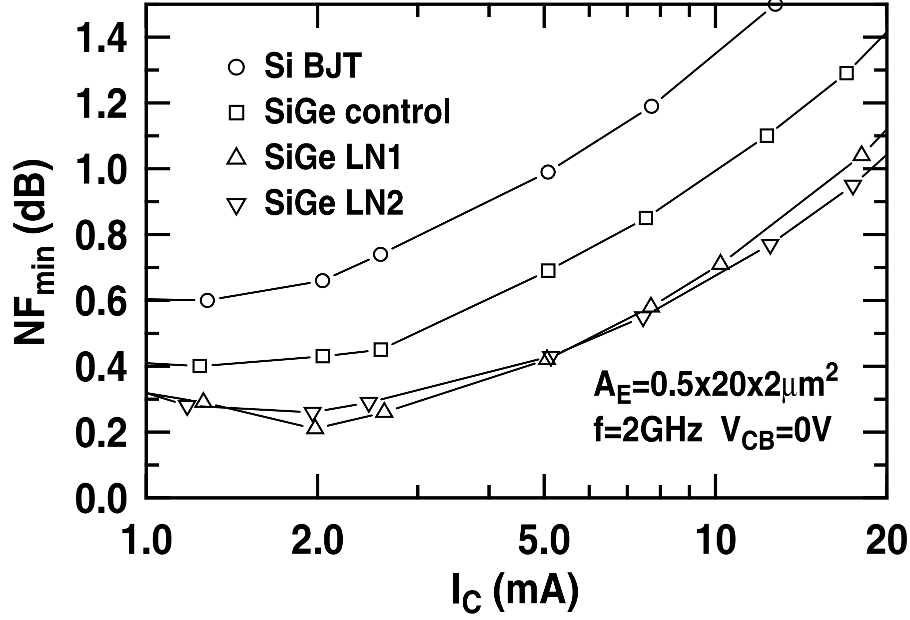

Figure 7. Measured minimum noise figure (NF_{min}) versus collector current at 2 GHz for the fabricated Si BJT, SiGe control, and the two optimized SiGe low noise profiles (LN1 and LN2).

and

$$B_{s,opt} = -\frac{\omega C_i}{2g_m R_n} \tag{13}$$

$$NF_{min} = 1 + \frac{1}{\beta} + \sqrt{\frac{2g_m R_n}{\beta} + \frac{2R_n(\omega C_i)^2}{g_m}\left(1 - \frac{1}{2g_m R_n}\right)} \tag{14}$$

In circuit applications where $g_m r_b \gg 1/2$, (14) can be further simplified to

$$NF_{min} = 1 + \frac{1}{\beta} + \sqrt{2g_m r_b}\sqrt{\frac{1}{\beta} + \left(\frac{f}{f_T}\right)^2} \tag{15}$$

And finally,

$$G_A^{\text{ass}} = \frac{1}{\omega^2 C_{bc} C_i r_b}\sqrt{\frac{g_m r_b + 1/2}{2}\frac{g_m^2}{\beta} + \frac{(\omega C_i)^2}{2} g_m r_b} \tag{16}$$

Equation (14) can be used to identify the frequency and bias current dependence of the highly-circuit-relevant NF_{min}. Observe that NF_{min} monotonically increases with frequency, making low-noise circuit design increasingly difficult as system operating frequencies increase. The current dependence of the noise figure is more complicated. We need to point out that the capacitance term is I_C dependent for bipolar transistors. However, it can be easily shown that at relatively high current, F_{min} increases with I_C monotonically through the g_m term. Equation (14) suggests that NF_{min} decreases (improves) with increasing β, decreasing C_i (transit time), and decreasing r_b. The two terms inside the second square root become equal at $f = f_T/\sqrt{\beta}$, which defines a transition of NF_{min} from a white noise behavior (independent of frequency) to a 10 dB/decade increase as the frequency increases. A smaller r_b in the transistor reduces not only the sensitivity of NF

to deviations from noise matching, but also the minimum noise figure NF_{min}. Equation (15) also indicates that a low r_b is the key to reducing NF_{min} when $f > f_T/\sqrt{\beta}$.

These analytic expressions can be used to optimize SiGe HBTs explicitly for broadband noise.[7,8] To improve SiGe HBT noise performance, we must reduce either input noise voltage generator (S_v) or the input noise current generator (S_i), or ideally both. According to the above analysis, the base resistance r_b needs to be reduced. For meaningful comparisons between devices, the r_b needs to be normalized by the emitter length when devices with differing emitter lengths are used. For similar reasons, noise figure comparisons should be made at the same current density or the same V_{BE}. We emphasize that (contrary to popular opinion) while a simple increase of the emitter length indeed reduces the r_b of the device, it does not improve the noise capability of the transistor, because all of the capacitances increase by the same factor. However, the emitter length (or the number of unit cells) can be optimized to simplify noise matching. At a given SiGe technology generation (node), the minimum emitter width is determined by the minimum feature size. The base sheet resistance is determined by the amount of boron dopants that can be kept in place during processing, and is obviously limited by the overall thermal cycle. Therefore, r_b and S_v are basically fixed. That is, the input noise voltage can only be reduced by lateral and vertical scaling, and the reduction of process thermal cycle, the addition of carbon doping, etc. There is no room for further reduction of S_v at a given technology generation.

The input noise current S_i, however, can be reduced by increasing β (to reduce I_B) and increasing f_T (to increase h_{21}). In particular, at relatively high currents where the RF power gain is large, $2qI_B$ dominates S_i in these devices. Therefore, significant noise improvement can only be achieved through an increase of β at relatively high bias currents. As we will show below, a high β is also desired in order to reduce the $1/f$ noise corner frequency and hence the phase noise in amplifiers and oscillators. The underlying approach to improving broadband noise performance at a given SiGe technology node is clear: the SiGe profiles must to be optimized for higher β and f_T at a given bias current under the fundamental constraint of maintaining overall SiGe film stability. This is an optimization problem ideally suited for bandgap engineering. This optimization approach has been successfully employed in Ge profile optimization for noise in first-generation SiGe HBTs, the results of which are shown in Figure 7. (Here, the LN1 and LN2 profiles are explicitly optimized for minimum noise using custom-designed Ge profiles with a peak Ge content of 14% and 18%, respectively, and compared to a SiGe control with 10% peak Ge and a Si BJT.[1])

4.2. Impact of Technology Scaling on Broadband Noise

Clearly, the achievable broadband noise performance of a given SiGe technology node represents a complex set of device design tradeoffs. Some general trends, however, with technology scaling can be observed.[9] 1) First-generation SiGe technology (50 GHz) can routinely achieve NF_{min} values below 1 dB out to 5 GHz operating frequency and below about 2 dB to 10 GHz. This level of performance is suitable for wireless transceiver front-ends for cost-sensitive, high-volume applications including GSM, CDMA, and 802.11b. The current commercial availability of SiGe parts for these markets offers ample evidence of this assertion. 2) Second-generation SiGe technology (100-120 GHz) can achieve NF_{min} values below 1 dB to about 12 GHz and below 2 dB to about 23 GHz. 3) Third-generation SiGe technology (200 GHz), through a combination of aggressive vertical (increased β and f_T) and lateral scaling (decreased R_B) achieves NF_{min} values below 0.4 dB to 12 GHz and 1.3 dB at 26.5 GHz (Figure 8).[9] This level of transistor broadband noise performance in a highly-integrated, fully-Si-compatible process technology falls within the performance range of GaAs pHEMTs currently in the commercial market. By suitable extrapolation of existing noise data, one can also easily demonstrate that third-generation SiGe HBTs possess sufficiently low noise to enable potentially paradigm-shifting IC applications such as 60 GHz WLAN and 77/94 GHz automotive radar, application regimes which were once believed to be the exclusive domain of III-V technologies. SiGe technology is particularly appealing for such markets due its natural ability to monolithically integrate devices (both HBT and CMOS – e.g., for DSP), passives, transmission lines, and possibly even the antennae on the same die/package, greatly reducing system form factor and cost.

This said, transistor performance obviously doesn't translate directly into identical performance at the circuit level. For example, LNA performance can be limited by the properties of the monolithic inductors which, due to their finite Q, may not always be able to achieve a noise-optimal match and which contribute significant noise themselves. The excellent matching characteristics of the SiGe HBT, however, do enable robust LNA design compared with most FET-based designs. Recently-reported SiGe HBT LNAs designed for the 60 GHz ISM band for WLAN have demonstrated 3.8 dB noise figure, with a corresponding gain of 14.5 dB (including the impact of the bondpads), records for Si-based technologies and more than adequate for system insertion (Figure 9).[10] This level of circuit performance in noise-critical applications falls within

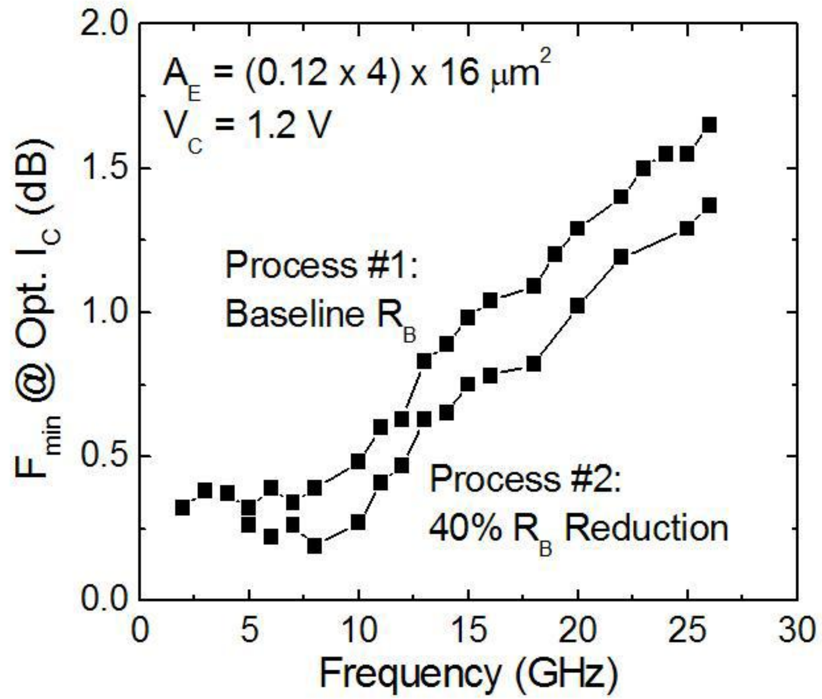

Figure 8. Measured minimum noise figure versus frequency for 2 different process options of a third-generation (200 GHz) SiGe HBT technology.[2]

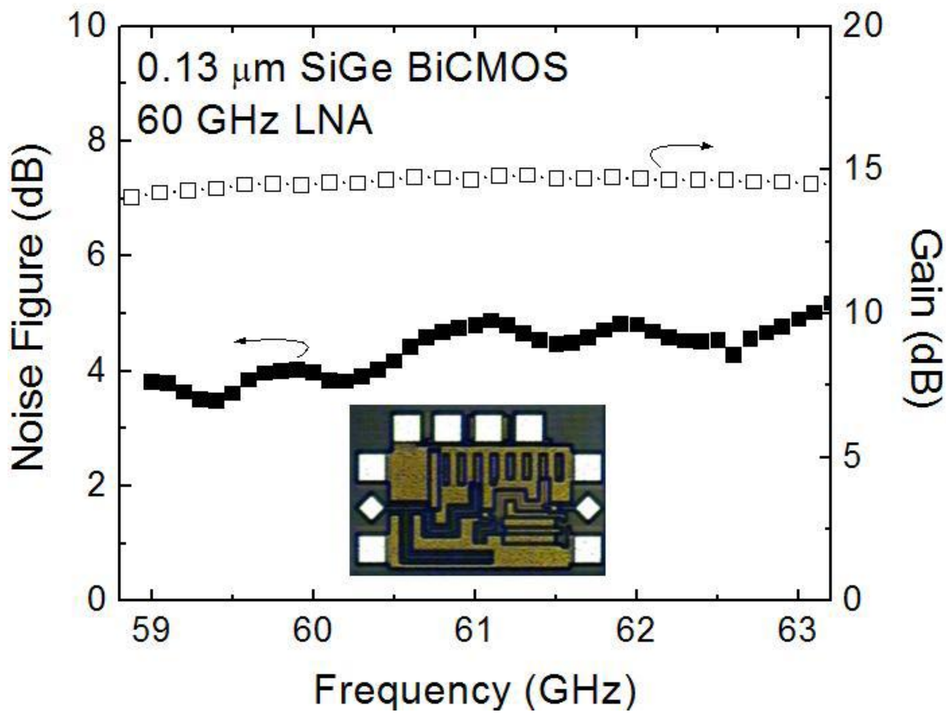

Figure 9. Measured 60GHz LNA performance for a third-generation (200 GHz) SiGe HBT technology.[2]

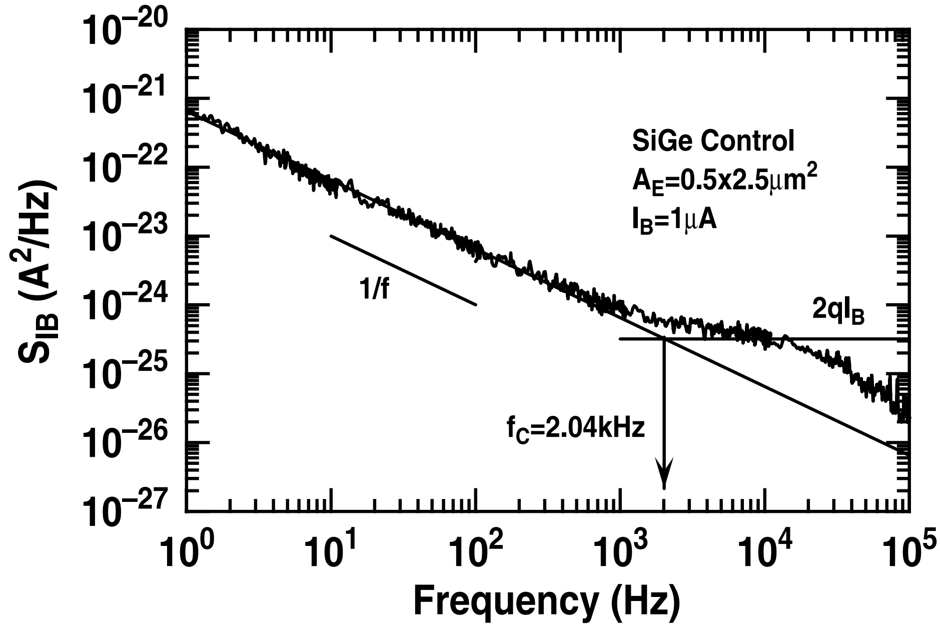

Figure 10. A typical low-frequency noise spectrum of a first generation SiGe HBT ($A_E = 0.5 \times 2.5\ \mu$m^2, and $I_B = 1\ \mu$A).

0.7 dB of the device data presented in Figure 8, and bodes well for the future of SiGe technology for such mm-wave IC applications.

4.3. Low-Frequency Noise

One of the advantages of SiGe HBTs over both III-V HBTs and CMOS is their superior low $1/f$ noise at low frequencies.[11] One might wonder why low-frequency noise (< 10 kHz) is so important for circuits operating at (multi-GHz) RF frequencies? First, low-frequency noise is upconverted to RF frequencies through the nonlinear $I - V$ and $C - V$ relationships inherent in the transistor to produce transistor phase noise (parasitic sidebands on the carrier frequency that fundamentally limit spectral purity of the system). Second, low-frequency noise is clearly important for emerging wireless receivers utilizing zero intermediate frequency (IF) architectures. Consider, for instance, applying an RF signal, f_{RF}, and a low-frequency signal, f_{LF}, to the base of a SiGe HBT. Spectral components of frequencies $f_{RF} \pm f_{LF}$ are generated in the collector current because of the strongly nonlinear $I_C - V_{BE}$ relation. The low-frequency signal is thus upconverted to RF by frequency-mixing through the nonlinear circuit elements in the transistor. Similarly, the low-frequency $1/f$ noise in a RF transistor amplifier is up-converted to RF frequencies by mixing with the incoming RF signal. This produces both amplitude and phase noise at the output, thus degrading spectral purity. In a heterodyne RF receiver, for example, the local oscillator (LO) phase noise results in a broadening of the down-converted signals at the intermediate frequency (IF), thus limiting the minimum channel spacing between adjacent channels. For a 900-MHz transceiver, for instance, a 30-kHz channel spacing typically requires less than -100 dBc/Hz phase noise at 100-kHz offset from the carrier.

Figure 10 shows a typical low-frequency base current noise spectrum (S_{I_B}) for a first-generation SiGe HBT. The noise spectrum shows a clear $1/f$ component as well as the $2qI_B$ shot noise level. The corner frequency f_C is determined from the intercept of the $1/f$ component and the $2qI_B$ shot noise level. The roll-off seen above 10 kHz is due to the bandwidth limitation of the preamp, and at higher I_B values, the $2qI_B$ shot noise level cannot be directly observed for this reason.

For a given device geometry, S_{I_B} is a function of I_B and is modeled by

$$S_{I_B} = K_F \frac{I_B^{\alpha}}{f}, \tag{17}$$

where K_F and α correspond to the KF and AF model parameters used in SPICE. The α value provides information on the physical origins of the $1/f$ noise. First-order theory predicts $\alpha = 1$ for carrier mobility fluctuations, and $\alpha = 2$ for

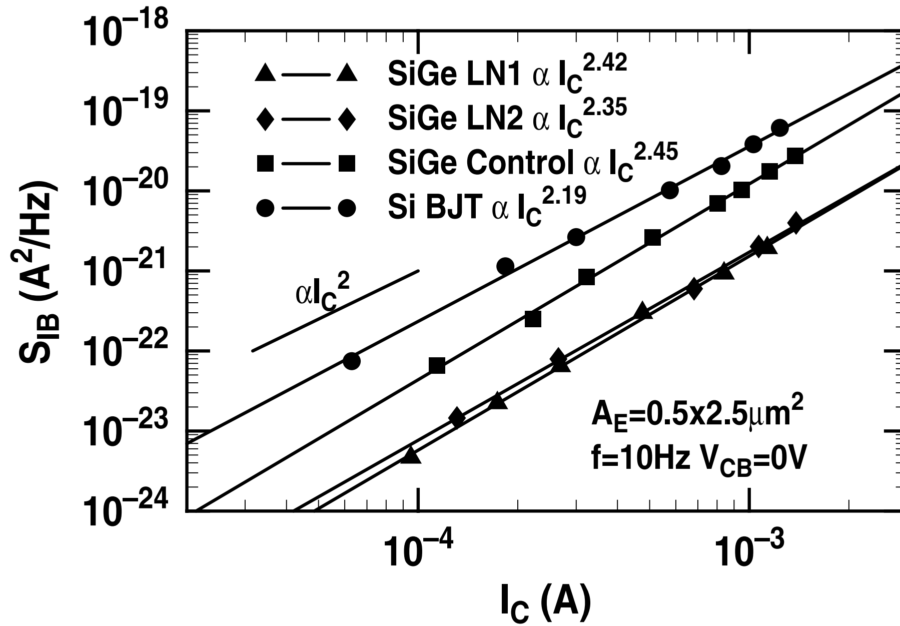

Figure 11. Measured S_{I_B} at 10 Hz as a function of I_C for the Si BJT, the SiGe control, and the two low-noise SiGe HBTs.

carrier number fluctuations.[12–17] The α for typical SiGe HBTs is close to 2, and varies only slightly with SiGe profile and collector doping profile (2 ± 0.2).

The observed $S_{I_B} - I_B$ dependence for the SiGe HBTs is approximately the same as for comparably fabricated Si BJTs, and independent of the SiGe profile. At a given V_{BE}, the I_B is the same for the SiGe HBTs and the Si BJT because of the identical emitter structure. Assuming that the $1/f$ noise is solely a function of the number of minority carriers injected into the emitter, we intuitively expect the same $1/f$ noise at a given V_{BE} (I_B). The fact that SiGe HBTs have the same S_{I_B} as a comparably constructed Si BJT for a given I_B is clearly good news for RF circuit designers. If we compare S_{I_B} at the same I_C (which is most relevant to circuit design), however, S_{I_B} is significantly *lower* (better) in SiGe HBTs than in Si BJTs, because of the lower I_B (higher β) found in SiGe HBTs, all else being equal (Figure 11).

Figure 12 compares the residual phase noise measured on single transistor amplifiers with different SiGe profile designs. The measured results confirm our theoretical expectations. The (14% peak Ge) LN1 SiGe HBT shows the lowest residual phase noise. The carrier frequency in these measurements is 10 GHz, the input power is 0 dBm, and both the source and load were terminated at 50 Ω. The residual phase noise level in these SiGe HBTs is excellent compared to competing technologies (of any type), with values as low as -165 dBrad/Hz at 10 kHz offset from the carrier.

The $1/f$ noise amplitude, as measured by the K_F factor, scales inversely with the total number of carriers in the noise-generating elements, according to the Hooge's theory. The $1/f$ noise generated by sources in the EB spacer oxide at the device periphery is inversely proportional to the emitter perimeter $P_E = W_E + L_E$, while the $1/f$ noise generated by sources located at the intrinsic EB interface (i.e., the emitter polysilicon-silicon interface) across the emitter window is inversely proportional to the emitter area $A_E = W_E \times L_E$. The K_F factor is often examined as a function of the emitter area, the emitter perimeter, or the perimeter-to-area ratio as a means of locating the contributing $1/f$ noise sources.[14–17] For instance, for fixed frequency, the combination of a $1/A_E$ dependence with an $I_B{}^2$ bias dependence for S_{I_B} is consistent with a uniform areal distribution of noise-generating traps across the emitter region. In practice, caution must be exercised in interpreting P_E or A_E scaling data, because test devices are often designed with the emitter width equal to the minimum feature size, and with an emitter length much larger than the emitter width. As a result, such data tend to scale with the emitter perimeter and area in a similar manner, making interpretation difficult. A wide distribution of device sizes and P_E/A_E ratios thus needs to be used when designing test structures for noise scaling studies in order to make a clear distinction between P_E and A_E scaling in SiGe HBTs. For all of the SiGe HBTs described here, the $1/f$ noise K_F factor

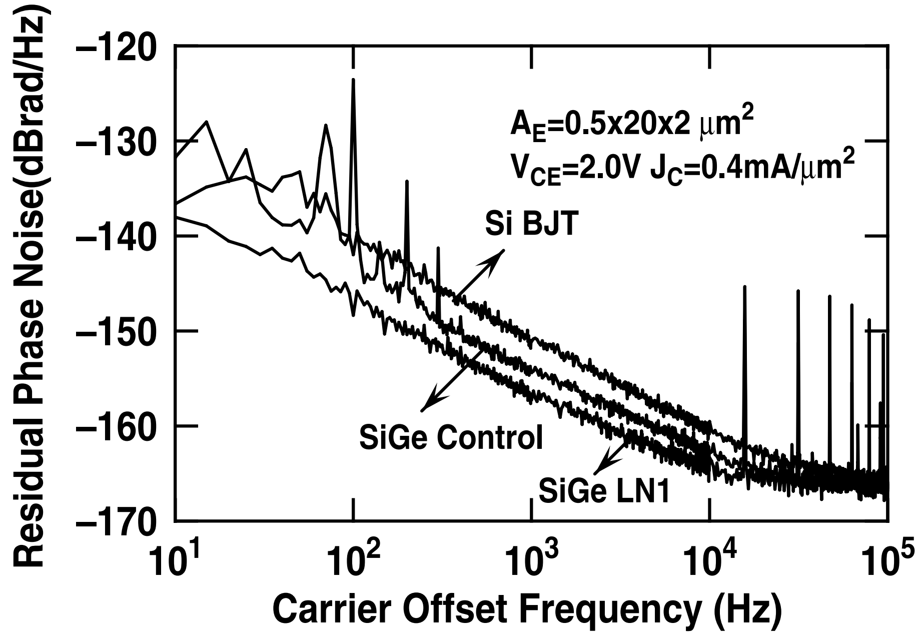

Figure 12. Measured residual phase noise spectra for the Si BJT, the SiGe control, and the LN1 low-noise SiGe HBT (I_C = 8 mA, A_E = 0.5 × 40 μm^2, P_{in}=0 dBm, and the carrier frequency is 10 GHz). Both the input and the output are terminated at 50 Ω.

is inversely proportional to A_E. Equation (17) can thus be rewritten as

$$S_{I_B} = \frac{K}{A_E}\frac{I_B^2}{f} = \frac{K}{\beta^2}\frac{1}{A_E}\frac{I_C^2}{f}, \tag{18}$$

where K is a factor independent of the emitter area and is defined as $K = K_F \times A_E$, where $\alpha = 2$ is assumed. Equation (18) is written as a function of I_C to facilitate technology comparisons for RFIC circuit design, for reasons discussed above. Because the K factor for low-frequency noise is approximately independent of base profile design, a higher β SiGe HBT has a lower S_{I_B}, and hence generates lower phase noise when used in RF amplifiers and oscillators. For a given operating current, a larger device can clearly be used to reduce S_{I_B}. This tactic, however, reduces f_T because of the lower J_C. The maximum device size one can use is usually limited by this f_T requirement. Optimum transistor sizing is thus important not only for reducing NF_{min}, but also for reducing phase noise.

Traditionally, $1/f$ noise performance is characterized by the corner frequency (f_C) figure-of-merit, defined to be the frequency at which the $1/f$ noise equals the shot noise level $2qI_B$. Equating (18) with $2qI_B$ leads to

$$f_C = \frac{KI_B}{2qA_E} = \frac{KJ_C}{2q\beta}, \tag{19}$$

where J_C is the collector current density, and β is the *dc* β. Equation (19) suggests that f_C is proportional to J_C and K, and inversely proportional to β. As expected, f_C is the lowest in the two low-noise SiGe HBTs, LN1 and LN2, and highest for the Si BJT. The modeling results calculated using (19) fit the measured data well.

The corner frequency alone, however, does not take into account transistor frequency response, and is thus not suitable for adequately assessing transistor capability for applications such as high-frequency oscillators. For instance, Si BJTs typically also have low (good) f_C, but do not have sufficient gain to sustain oscillation at RF and microwave frequencies because of their limited f_T. GaAs HBTs, on the other hand, have high f_T, but typically also have high f_C and hence generate larger phase noise when used in oscillators. SiGe HBTs provide f_T comparable to GaAs HBTs, and lower f_C than Si BJTs, making them a very attractive choice for ultra-low phase noise oscillators. A better figure-of-merit to gauge transistor $1/f$ noise performance for oscillator applications is the f_C/f_T ratio,[18] since it also takes into account transistor frequency response via f_T.

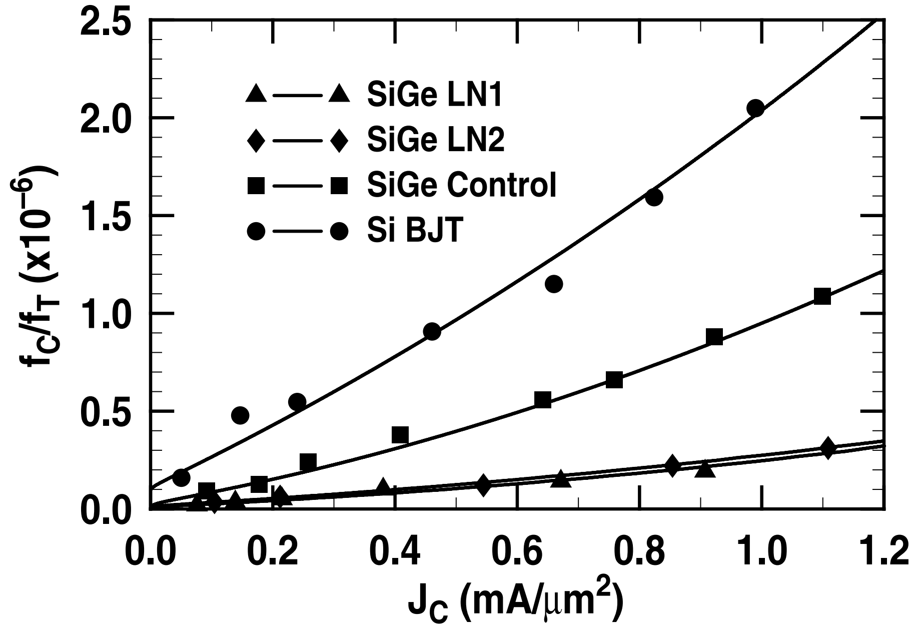

Figure 13. Measured and modeled f_C/f_T ratio as a function of J_C for the standard breakdown voltage Si BJT, the SiGe control, and the two low-noise SiGe HBTs.

The cutoff frequency f_T is related to J_C by

$$\frac{1}{2\pi f_T} \approx \tau_f + \frac{1}{g_m} C_t = \tau_f + \frac{kT}{qJ_C} C_t, \quad (20)$$

where τ_f is the forward transit time, $g_m = qJ_C/kT$ is the transconductance per unit area, and C_t is the total junction depletion capacitance per unit area. Prior to f_T roll-off at high J_C, τ_f and C_t are nearly constant. The f_C/f_T ratio can thus be obtained by combining (19) and (20)

$$\frac{f_C}{f_T} = K \frac{\pi}{q} \frac{J_C}{\beta} \left(\tau_f + kT \frac{C_t}{qJ_C} \right) = \frac{K\pi}{\beta q} \left(\tau_f J_C + kT/qC_t \right). \quad (21)$$

This model thus suggests a *linear* increase of the f_C/f_T ratio with operating collector current density J_C, provided that β and τ_f are constants.[1] This is in contrast to the prediction of a J_C independent f_C/f_T ratio,[18] which assumed $\alpha = 1$. At larger values of J_C where f_T is high, $\tau_f J_C \gg kT/qC_t$, and $f_C/f_T \approx K\pi\tau_f J_C/\beta q$. The f_C/f_T ratio is thus determined by the $K\tau_f/\beta$ term at higher J_C values. A smaller τ_f, a higher β, and a smaller K factor are desired in order to reduce (improve) f_C/f_T. A smaller f_C/f_T indicates better phase noise performance at high frequencies. Figure 13 shows the measured and modeled f_C/f_T-J_C dependence for a first generation SiGe HBT. The agreement between data and model is quite good. The two low-noise SiGe HBTs show the best (lowest) f_C/f_T because of highest f_T and the lowest f_C, as expected.

These results confirm that SiGe profiles optimized for high β and high f_T have better phase noise performance for the same operating frequency. To achieve the same RF gain, a higher f_T transistor can operate at a lower J_C, thus reducing f_C/f_T, which further improves (lowers) f_C. As can be seen from equation (21), the τ_f/β ratio can be used as a figure-of-merit for SiGe profile optimization, because f_C/f_T is proportional to $K\tau_f/\beta$ according to equation (21). The K factor is primarily determined by the emitter structure, and is independent of the SiGe profile used in the base as well the collector

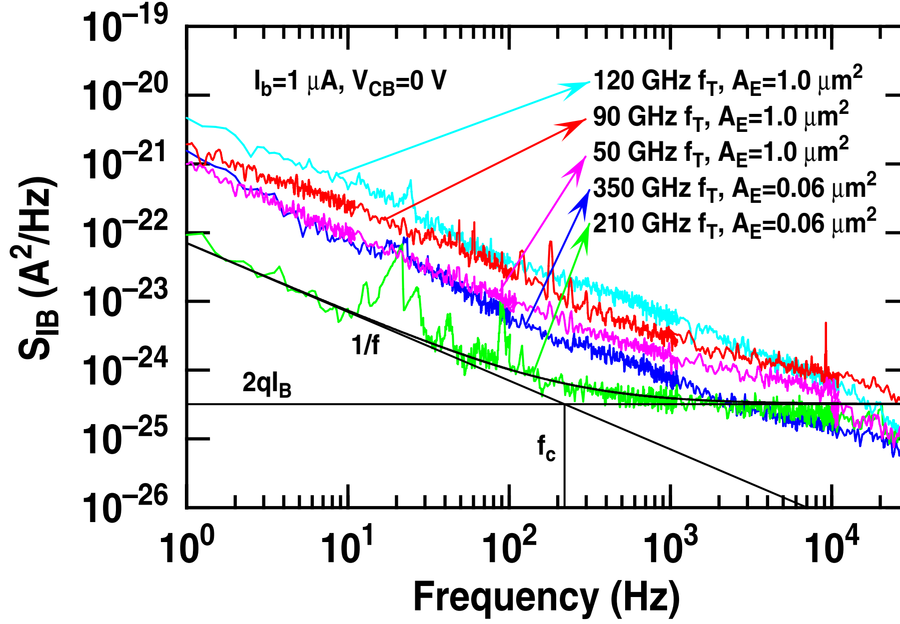

Figure 14. Best case noise spectra for several generations of SiGe HBTs.

doping profile, as evidenced by the experimental data. A SiGe profile producing the lowest τ_f/β ratio leads to the best f_C/f_T ratio, and hence should have the best phase noise performance at RF/microwave frequencies.

4.4. Small-Size Effects

Transistors are aggressively scaled to improve their performance and integration level. One design issue associated with geometrical scaling is that the low-frequency noise often shows a different frequency dependence for each individual device, and this can potentially adversely affect both circuit performance and design. This noise variation with size has been observed in MOSFETs, JFETs, and BJTs in small-sized devices.[19−21] The fundamental noise mechanism inside transistors is generally regarded as a superposition of individual trapping/detrapping processes due to the presence of G/R centers in the device. Each G/R center contributes a Lorentzian-type ($1/f^2$) noise signature, and given a sufficient number of traps, these Lorentzian processes combine to produce the observed $1/f$ noise behavior. At sufficiently small device size, however, the total number of traps is small enough that non-$1/f$ behavior, and large statistical variations, can be observed.[21] Figure 14 compares best-case noise spectra for several SiGe HBT generations. Clearly, SiGe HBTs are capable of extremely low levels of low-frequency noise. A record-low 1/f noise corner frequency of 220 Hz ($I_B = 1$ μA) in a 0.12×0.50 μm^2 SiGe HBT with a peak f_T of 210 GHz and NF_{min} of less than 0.5 dB at 10 GHz is achieved.

Cross-generational noise magnitude as a function of bias current (at 10 Hz) for the devices exhibiting 1/f behavior show a classical I_B^2 dependence, which indicates that the noise mechanism is consistent with number fluctuation theory. It has been recently reported that low-frequency noise in small-sized SiGe HBTs shows large device-to-device variations as the technology scales, an issue of potential concern for both compact modeling and for certain circuit applications.[23,24] The statistical noise variation between samples of the same geometry was quantified using a variation coefficient (δ), as defined by the standard deviation formula:[21]

$$\delta = \frac{1}{S_{I_{B,avg}}} \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(S_{I_{B,i}} - S_{I_{B,avg}} \right)^2} \tag{22}$$

where $S_{I_{B,avg}} = N^{-1} \sum_{i=1}^{N} S_{I_{B,i}}$ is the average noise spectrum, i indicates the ith sample, and N is the total number of samples.

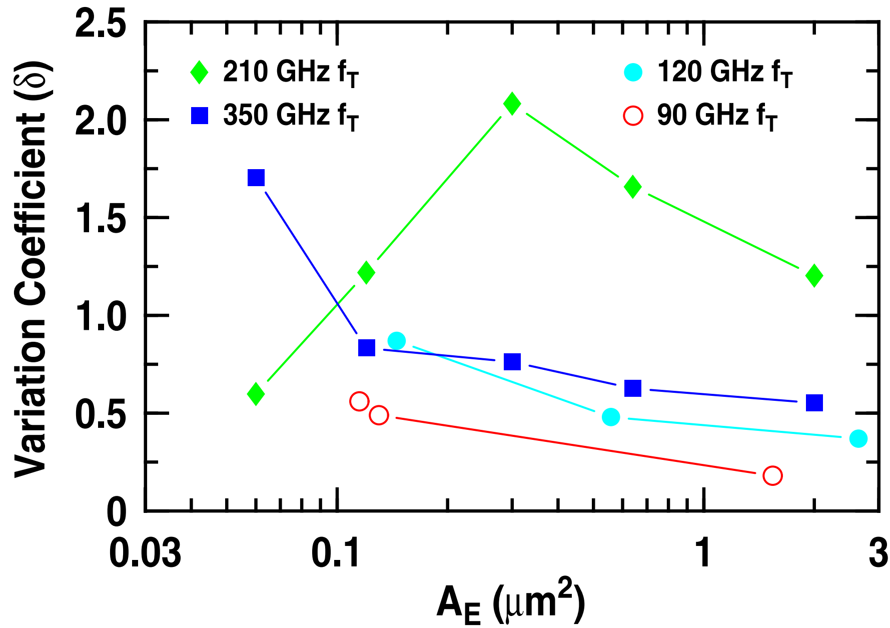

Figure 15. Noise variation versus emitter area for several generations of SiGe HBTs.

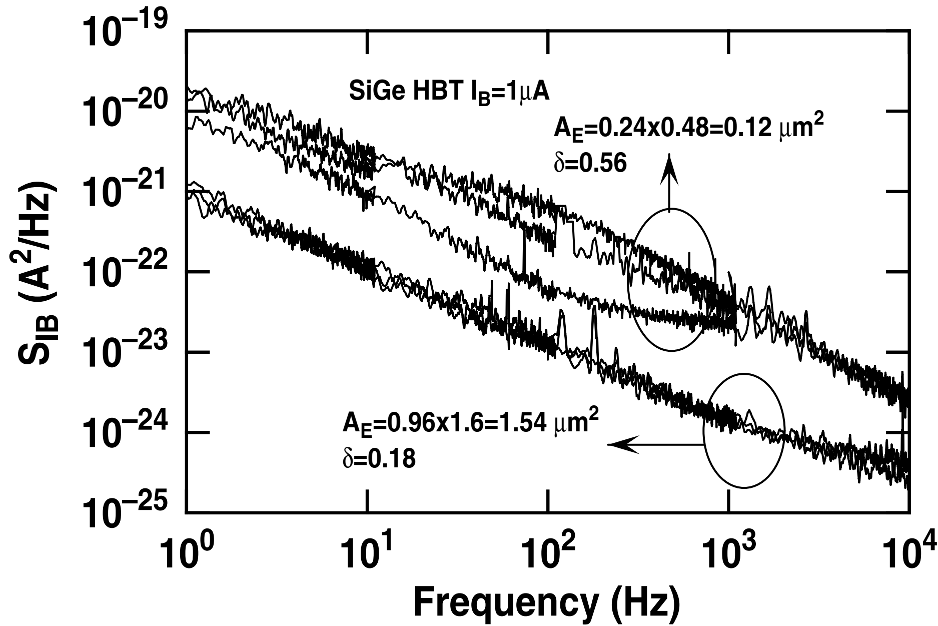

Figure 16. Low-frequency noise spectra in three samples with emitter area of A_E=0.24 × 0.48μm^2 and A_E=0.96 × 1.6μm^2, at I_B=1μA. A noise variation of δ=0.56 and δ=0.18 in eight measured samples was observed, respectively.

Noise variation data is shown in Figure 15 (the variation is negligible in the 50 GHz, first-generation devices), and explicit device-to-device statistical scatter for eight 0.12×0.5 μm^2 350 GHz and 210 GHz SiGe HBTs is shown in Figure 16.[24] Interestingly, observe that the noise variation in the third-generation 210 GHz technology shows anomalous scaling behavior below about 0.2-0.3 μm^2 emitter geometry, below which the noise variation rapidly decreases.

To better understand these noise variations, we have applied simple modeling techniques to noise variation data in

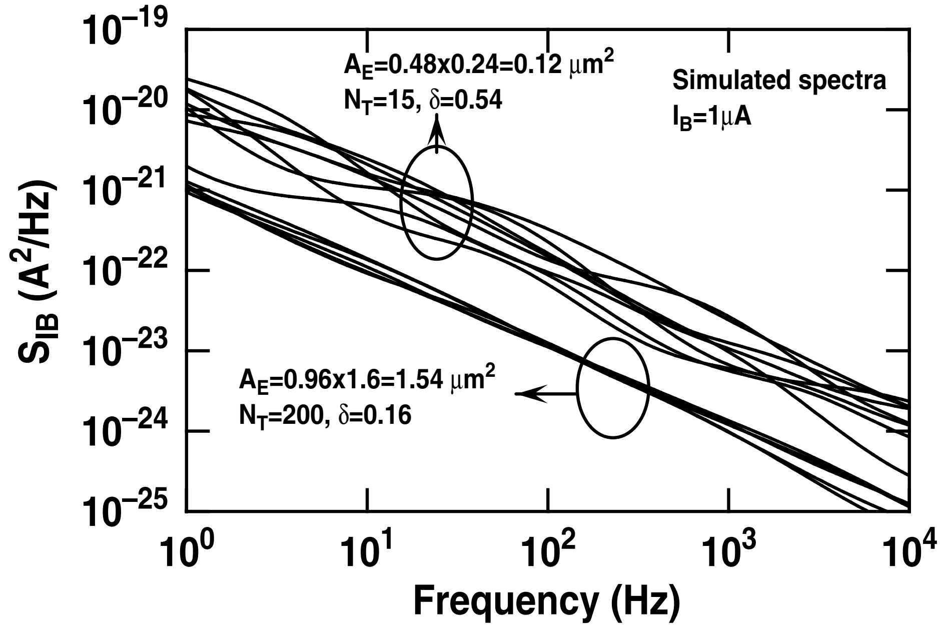

Figure 17. Simulated noise spectra in SiGe HBTs at I_B=1μA. The trap number is N_T=15 for devices with A_E=0.24 × 0.48μm^2, and the noise variation is δ=0.54 for the eight simulated spectra. The trap number is N_T=200 for devices with A_E=0.96 × 1.6μm^2, and the noise variation is δ=0.16 for eight simulated spectra.

second-generation SiGe HBTs.[23] The fundamental noise generating mechanism is generally accepted as a carrier number fluctuation for the moderate and high bias range in modern Si BJTs and SiGe HBTs.[22] An explanation for such carrier number fluctuations is the trapping/detrapping of carriers by traps at the interfaces or in the oxide layers. The noise of each individual trapping/detrapping process exhibits a Lorentzian spectrum ($1/f^2$). The physical origin of excess low-frequency noise in JFETs and MOSFETs is usually attributed to variations in channel carrier density due to fluctuations in occupancy of the G/R centers. The total $1/f$ noise is due to the superposition of multiple G/R noises.[19] Recently, the $1/f$ noise in polysilicon BJTs was also expressed as the superposition of Lorentzian noise signatures,[21] and thus it is desirable to express the noise spectrum in SiGe HBTs here as a summation of Lorentzian spectra. In addition, we observed that the noise spectrum is proportional to I_B^2 and $1/A_E$ in the large devices, and shows a deviation from $1/f$ in the small devices. Consequently, an empirical expression of the low-frequency noise spectrum can be written as[21]

$$S_{I_B} = \sum_{i}^{N_T} A \frac{J_B^2 \tau_i}{1 + (2\pi f \tau_i)^2} \tag{23}$$

where A is a constant, τ_i is the characteristic time constant of the ith trap and has to be distributed as $1/\tau$ to produce the $1/f$ spectrum, and N_T is the total number of traps in the device and proportional to A_E. When N_T is large enough, which is the case in the large device, eq. (23) will exhibit a $1/f$ spectrum. When N_T is small enough, however, corresponding to the small device case here, the spectrum modeled by this equation will show a deviation from $1/f$. Fig. 17 shows the calculated pre-stress noise spectra in the small and large devices. Two thousand characteristic time constants were generated over $1/(2\pi\times10^7)$ to $1/(2\pi\times10^{-3})$ with a $1/\tau$ distribution in the calculations. Fifteen characteristic time constants were randomly selected for every calculation in the small device and two hundred were selected for the large device (A=$1.2\times10^{-25} cm^4$). Eight individual calculations (to mimic the eight measurements) were performed for both the small and large devices, and the noise variation of eight calculated spectra is 0.54 for small device and 0.16 for large device, consistent with the data.

Given the importance of low-frequency noise in RF and microwave circuit applications, careful measurement and modeling of low-frequency noise is clearly important. Advanced SiGe HBTs with aggressively scaled emitter geometries have a size dependent low-frequency noise variation. This not only complicates accurate device compact modeling, but

is of potential concern for phase and jitter noise in circuits utilizing minimum geometry devices (e.g., high speed digital circuits). This size variation is believed to be fundamental to scaled bipolar technologies, SiGe or otherwise, and thus warrants careful attention.

5. SUMMARY

SiGe HBT technology represents a remarkable success story for the microelectronics industry, and is being rapidly deployed by a host of companies globally to a wide variety of mixed-signal IC applications. I have reviewed the state-of-the-art in SiGe technology, and discussed the design and operational principles of SiGe HBTs. The low-noise capability of SiGe HBTs, both broadband and at low-frequency, is a compellingly attractive feature of SiGe technology, and is expected to propel SiGe ICs into application venues traditionally reserved for III-V technologies.

6. ACKNOWLEDGEMENT

I am grateful to my current and former students (especially Zhenrong Jin, Jarle Johansen, Enhai Zhao, and Lakshmi Vempati), my colleague and friend, Guofu Niu, as well as Alvin Joseph, David Harame, Greg Freeman, David Greenberg, Dave Ahlgren, Bernie Meyerson, and the IBM SiGe team for their contributions and support. This work was funded by the SRC, NASA, DARPA, IBM, and the Georgia Electronic Design Center at Georgia Tech.

REFERENCES

1. J.D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*, Artech House, Boston, MA 2003.
2. J.D. Cressler, Editor, *The Silicon-Heterostructure Handbook: Materials, Fabrication, Devices, Circuits, and Applications of SiGe and Si Strained-Layer Epitaxy,* CRC Press, New York, NY 2005.
3. Portions in this paper will appear in: J.D. Cressler, "On the Potential of SiGe HBTs for Extreme Environment Electronics," *Proceedings of the IEEE*, in press.
4. A.J. Joseph *et al.*, "0.13 μm 210 GHz f_T SiGe HBTs - expanding the horizons of SiGe BiCMOS," *Technical Digest of the IEEE International Solid-State Circuits Conference*, pp. 180-182, 2002.
5. A. Fischer *et al.*, "An equilibrium model for buried SiGe strained layers," *Solid-State Electronics*, vol. 44, pp. 869-873, 2000.
6. E.O. Johnson, "Physical limitations on frequency and power parameters of transistors," *RCA Review*, pp. 163-177, 1965.
7. G. Niu *et al.*, "Noise parameter optimization of UHV/CVD SiGe HBT's for RF and microwave applications," *IEEE Trans. Elect. Dev.*, vol. 46, pp. 1347-1354, 1999.
8. G. Niu *et al.*, "A unified approach to RF and microwave noise parameter modeling in bipolar transistors," *IEEE Trans. Elect. Dev.*, vol. 48, pp. 2568-2574, 2001.
9. D. Greenberg, "Broadband Noise," from *The Silicon-Heterostructure Handbook: Materials, Fabrication, Devices, Circuits, and Applications of SiGe and Si Strained-Layer Epitaxy,* J.D. Cressler, Editor, CRC Press, New York, NY 2005.
10. S. Reynolds *et al.*, "60 GHz receiver circuits in SiGe bipolar technology," *Proc. IEEE Int. Solid-State Circuits Conf.*, pp. 442-443,2004
11. L. S. Vempati *et al.*, "Low-frequency noise in UHV/CVD epitaxial Si and SiGe bipolar transistors," *IEEE J. Solid State Circuits,* vol. 31, no. 10, pp. 1458-1467, 1996.
12. L.K.J. Vandamme, "Noise as a diagnostic tool for quality and reliability of electronic devices," *IEEE Trans. Elect. Dev.*, vol. 41, pp. 2174-2187, 1994.
13. A. Mounib *et al.*, "Low-frequency noise sources in polysilicon emitter BJT's: influence of hot-electron-induced degradation and post-stress recovery," *IEEE Trans. Elect. Dev.*, vol. 42, pp. 1647-1652, 1995.
14. M.J. Deen *et al.*, "Low frequency noise in polysilicon-emitter bipolar junction transistors," *J. Appl. Phys.*, vol. 77, pp. 6278-6285, 1995.
15. M. Koolen and J. C. J. Aerts, "The influence of non-ideal base current on 1/f noise behaviour of bipolar transistors," *Proc. IEEE Bipolar/BiCMOS Circ. Tech. Meeting*, pp. 232-235, 1990.
16. H. A. W. Markus and T. G. M. Kleinpenning, "Low-frequency noise in polysilicon emitter bipolar transistors," *IEEE Trans. Elect. Dev.*, vol. 42, pp. 720-727, 1995.

17. M.J. Deen, S. L. Rumyantsev, and M. Schroter, "On the origin of 1/f noise in polysilicon emitter bipolar transistors," *J. Appl. Phys.*, vol. 85, pp. 1192-1195, 1999.
18. L.K.J. Vandamme and G. Trefan, "Review of low-frequency noise in bipolar transistors over the last decade," *Proc. IEEE Bipolar/BiCMOS Circ. Tech. Meeting*, pp. 68-73, 2001.
19. M.J. Uren *et al.*, "$1/f$ and random telegraph noise in silicon MOSFETs," *Appl. Phys. Lett.,* vol. 47, pp. 1195-1197, Dec., 1985.
20. K. Kandiah, "Random telegraph signal currents and low-frequency noise in JFETs," *IEEE Trans. Electron Devices,* vol. 41, no. 11, pp. 2006-2015, 1994.
21. M. Sanden *et al.*, "A new model for the low-frequency noise and the noise level variation in polysilicon emitter BJTs," *IEEE Trans. Electron Devices,* vol. 49, no. 3, pp. 514-520, 2002.
22. P. Llinares *et al.*, "Dimension scaling of 1/f noise in the base current of quasiself-aligned polysilicon emitter bipolar junction transistors," *J. Appl. Phys.,* vol. 82, pp. 2671-2675, 1997.
23. Z. Jin *et al.*, "Impact of geometrical scaling on low-frequency noise in SiGe HBTs," *IEEE Trans. Electron Devices*, vol. 50, pp. 676-682, 2003.
24. J. Johansen *et al.* "On the scaling limits of low-frequency noise in SiGe HBTs," *Solid-State Electronics*, pp. 1897-1900, 2004.