A Theory of High-Frequency Distortion in Bipolar Transistors

Mani Vaidyanathan, *Member, IEEE*, Masaya Iwamoto, *Student Member, IEEE*, Lawrence E. Larson, *Fellow, IEEE*, Prasad S. Gudem, *Member, IEEE*, and Peter M. Asbeck, *Fellow, IEEE*

***Abstract*—High-frequency distortion in bipolar transistors is examined by using the charge-control approach of Poon and Narayanan to connect the device's distortion behavior to its "loaded" unity–current–gain frequency ($\hat{\omega}_T$). The resulting expressions for the distortion reveal considerable information on its frequency and bias dependence. Points on the $\hat{\omega}_T$ versus collector current curve yielding optimum distortion performance are identified and interpreted in terms of current cancellation. Both second- and third-order distortion are considered, and the results are validated by both simulation and experiment.**

***Index Terms*—Current cancellation, harmonic distortion, heterojunction and homojunction bipolar transistors, high-frequency distortion, intermodulation distortion, linearity, nonlinear distortion, unity–current–gain frequency, Volterra series.**

I. Introduction

OVER THE years, the problem of high-frequency distortion in bipolar transistors has been extensively studied. Among the earlier works, Narayanan [1], [2] was the first to present a detailed examination of distortion using Volterra series; Chisholm and Nagel [3] and Kuo [4] focused on computer algorithms for calculating distortion in transistor circuits; Poon and Narayanan [5]–[7] combined Volterra series with a charge-control approach; Abraham and Meyer [8] employed a simplified transistor model to suggest design guidelines for low distortion; and many others contributed to the literature, as cited by these authors, and in a comprehensive review paper by Lotsch [9]. More recently, Maas *et al.* [10] attributed the surprisingly good linearity of heterojunction bipolar transistors (HBTs) working at high frequencies to a cancellation of nonlinear currents arising from the dynamic resistance and capacitance of the emitter–base junction. This analysis was followed by a number of studies that were mostly of an empirical nature, leading to various observations on the factors affecting high-frequency distortion; for example, comments were made on the role of current cancellation [11]–[14], the feedback effect of the parasitic base and emitter resistances [15], [16], the impact of the nonlinear, collector–base depletion capacitance [14]–[19], the importance of base–collector transit time [16], [19], and the choice of bias voltage and current [14], [16]–[20]. Studies have also been undertaken to examine the distortion behavior of other important microwave devices; for example, Pedro *et al.* [21]–[23] recently examined distortion in metal–semiconductor field-effect transistors (MESFETs). Despite all of these investigations, a general description of high-frequency distortion, which offers good physical insight and can be applied to a wide variety of devices, is still lacking. This shortcoming stems mainly from the fact that nonlinear distortion is an involved problem, which is not amenable to easy solution; even when expressions for transistor distortion can be found, they are typically very complex, and involve numerous terms that offer little intuition.

In this study, we develop a basic theory of high-frequency distortion in bipolar transistors by employing the charge-control approach suggested by Poon and Narayanan [5]–[7]. Use of the charge-control approach alleviates much of the usual difficulty in analyzing distortion, and leads to powerful expressions that relate the distortion generated by the transistor to its transconductance and unity–current–gain frequency, and to the derivatives of these quantities with respect to base–emitter voltage and collector current, respectively. In particular, the connections between the distortion and unity–current–gain frequency provide substantial information on the frequency and bias dependence of the distortion characteristics, and offer new insight into the cancellation phenomenon described by Maas *et al.* [10]. The result, which we validate using both simulation and experiment, is a useful step toward a general theory of distortion.

In Section II, the transistor model used for the analysis is presented, and the equations needed to combine the charge-control approach with Volterra series are formulated. In Section III, expressions are found for the second-order distortion characteristics, and their predictions are compared with simulation and experiment. In Section IV, expressions are derived for the third-order intermodulation distortion and then applied to practical devices. Section V summarizes the conclusions.

II. Analytic Approach

A. Model

Fig. 1 shows the transistor model used in the analysis. The elements and their assumed functional dependencies on the total

Manuscript received June 26, 2002. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, by IBM under a University Partnership Program grant, by Agilent Technologies, by the Center for Wireless Communications at the University of California at San Diego, and by the Army Research Office under the Multiuniversity Research Initiative Program "Digital Communications Based on Nonlinear Dynamics and Chaos."

M. Vaidyanathan, M. Iwamoto, L. E. Larson, and P. M. Asbeck are with the Center for Wireless Communications, Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA.

P. S. Gudem was with the West Coast Design Center of Excellence, IBM T. J. Watson Research Center, Encinitas, CA 92024 USA. He is now with Qualcomm Inc., San Diego, CA 92121 USA.

Digital Object Identifier 10.1109/TMTT.2002.807821

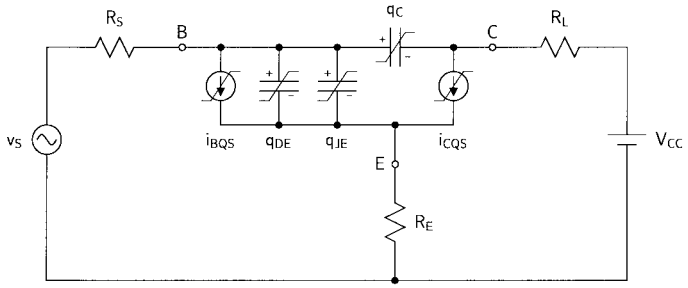


Fig. 1. Large-signal transistor model used in the analysis. The definitions of the elements are given in the text. Nonlinear elements are marked in the usual fashion.

(static plus small-signal) values of the internal base–emitter and base–collector voltages, $v_{\rm BE}$ and $v_{\rm BC}$, are as follows.

1) $V_{\rm CC}$ is the constant supply voltage.
2) v_S is the total source voltage.
3) R_S is the sum of the external source and device base resistances.
4) R_L is the sum of the external load and device collector resistances.
5) R_E is the sum of any external emitter and device emitter resistances.
6) $q_{\rm JE}(v_{\rm BE})$ is the charge associated with the emitter–base depletion capacitance.
7) $q_{\rm DE}(v_{\rm BE})$ is the emitter portion of the stored free charge.
8) $q_C(v_{\rm BE}, v_{\rm BC})$ is the total collector charge, which includes both the collector portion of the stored free charge *and* the charge associated with the collector–base depletion capacitance.
9) $i_{\rm CQS}(v_{\rm BE})$ is the quasi-static collector current.
10) $i_{\rm BQS}(v_{\rm BE})$ is the quasi-static base current.

The circuit in Fig. 1 represents the simplest adequate model with which to examine the problem. Important features of this model are as follows.

- Effects arising from the falloff of the transistor's unity–current–gain frequency at high currents are automatically included, since the charges $q_{\rm DE}(v_{\rm BE})$ and $q_C(v_{\rm BE}, v_{\rm BC})$ can have an arbitrary $v_{\rm BE}$ (or $i_{\rm CQS}$) and $v_{\rm BC}$ dependence.
- The terminations R_S, R_L, and R_E are assumed to be purely resistive. This assumption not only simplifies the analysis, but also leads to useful results that connect (through charge-control relations) the distortion generated by the transistor to its unity–current–gain frequency, the latter being defined under conditions in which R_S, R_L, and R_E take on resistive values determined solely by the device parasitics. In addition, characterizing distortion with resistive terminations offers advantages with respect to evaluating the capabilities of a technology under standardized conditions [20, p. 1530]. From these perspectives, the distortion expressions derived in this study can be viewed as figures-of-merit for device linearity. We have also found these expressions to be useful in the design of real broad-band power amplifiers [24]. However, in specific circuit applications, it should be noted that the distortion performance will depend on the exact nature of the terminations [25], [26].
- A number of other simplifications in the model are necessary to keep the analysis manageable, including the neglect of the Early effect, the neglect of avalanche breakdown, the neglect of self-heating [27], the assumption of a linear base resistance (which is lumped into R_S), and the neglect of collector–substrate capacitance, which is present in Si-based devices. However, each of these should have only secondary impacts, as illustrated, for example, by the work in [20], where numerical results from a model [20, Fig. 1(b)] that is very similar to the one used here were shown to be in good agreement [20, Figs. 4–7] with those from a much more involved model [20, Fig. 1(a)] as well as experiment.

B. Formulation

Using Kirchoff's laws, it is possible to write the following circuit equations:

$$V_{\rm CC} - i_C R_L^\star + v_{\rm BC} - v_{\rm BE} = i_B R_E \tag{1}$$

$$\frac{v_S - v_{\rm BE}}{R_S^\star} - i_C \left(\frac{R_E}{R_S^\star}\right) = i_B \tag{2}$$

where $R_L^\star \equiv R_L + R_E$ and $R_S^\star \equiv R_S + R_E$ are effective load and source resistances, respectively, and where the terminal currents are

$$i_C = i_{\rm CQS} - \frac{dq_C}{dt} \tag{3}$$

and

$$i_B = i_{\rm BQS} + \frac{dq_T}{dt} \tag{4}$$

with q_T being the total charge in the transistor, given by

$$q_T \equiv q_{\rm DE}(v_{\rm BE}) + q_{\rm JE}(v_{\rm BE}) + q_C(v_{\rm BE}, v_{\rm BC}). \tag{5}$$

If the current gain is high and the operating frequency is restricted to a few times below the unity–current–gain frequency, then combining (1)–(4) and retaining only the most important terms, it is easy to obtain the following simplified set of circuit equations:

$$i_{\rm CQS} = \frac{V_{\rm CC} + v_{\rm BC} - v_{\rm BE}}{R_L^\star} \tag{6}$$

$$\frac{dq_T}{dt} = \frac{v_S - v_{\rm BE}}{R_S^\star} - i_{\rm BQS} - i_{\rm CQS}\left(\frac{R_E}{R_S^\star}\right). \tag{7}$$

Since $i_{\rm CQS}$ in Fig. 1 is a known function of $v_{\rm BE}$, the voltage $v_{\rm BC}$ can be eliminated between (5) and (6), and the small-signal base–emitter voltage $v_{\rm be}$, and the small-signal collector current $i_{\rm cqs}$, can each be expanded as a Taylor series in the small-signal charge q_t:

$$v_{\rm be} = \varsigma_1 q_t + \varsigma_2 q_t^2 + \varsigma_3 q_t^3 \tag{8}$$

$$i_{\rm cqs} = \nu_1 q_t + \nu_2 q_t^2 + \nu_3 q_t^3. \tag{9}$$

As shown in Appendix I, the series coefficients in (8) and (9) can be expressed in terms of the transconductance g_m and its

derivatives ($\dot{g}_m$ and $\ddot{g}_m$) with respect to the base–emitter voltage $v_{\rm BE}$, and in terms of the "loaded" unity–current–gain frequency, denoted $\hat{\omega}_T$, and its derivatives ($\hat{\omega}_T'$ and $\hat{\omega}_T''$) with respect to $i_{\rm CQS}$.

The definition of $\hat{\omega}_T$, given by (56) in Appendix I, differs from the usual unity–current–gain frequency of the transistor $\omega_T \equiv 2\pi f_T$, since R_L and R_E in Fig. 1 include *external* loading elements in addition to the transistor's own parasitic collector and emitter resistances. These external elements accentuate the feedback effect of q_C, which is represented in (56) by the $(R_L + R_E)C_{\rm jc}$ time constant. Thus, $\hat{\omega}_T$ and its derivatives contain *more* information on this feedback than ω_T and its derivatives. However, for typical external loads, the distinction is unimportant; the qualitative behavior of both sets of quantities will be similar. This follows because, at low currents, both $\hat{\omega}_T$ and ω_T are primarily determined by the depletion capacitances (as opposed to the feedback through q_C), and at high currents, the peaking and subsequent falloff of both depends primarily on the nonlinear behavior of the stored free charge (arising, for example, from base-widening effects). Therefore, the *qualitative* links that connect $\hat{\omega}_T$ to the distortion behavior also apply to ω_T (see Figs. 3 and 6).

The small-signal version of (7) is

$$\frac{dq_t}{dt} = \frac{v_s - v_{\rm be}}{R_S^\star} - i_{\rm bqs} - i_{\rm cqs}\left(\frac{R_E}{R_S^\star}\right). \tag{10}$$

In this equation, $i_{\rm bqs}$ can be expressed as a power series in $i_{\rm cqs}$:

$$i_{\rm bqs} = b_1 i_{\rm cqs} + b_2 i_{\rm cqs}^2 + b_3 i_{\rm cqs}^3. \tag{11}$$

As shown in Appendix I, the coefficients in this expansion may be written in terms of the low-frequency current gain β_0, and its derivatives with respect to $i_{\rm CQS}$. However, we will assume that the relationship between $i_{\rm bqs}$ and $i_{\rm cqs}$ is purely linear, in which case (11) reduces to

$$i_{\rm bqs} \approx \frac{i_{\rm cqs}}{\beta_0} \tag{12}$$

with $b_1 \equiv di_{\rm BQS}/di_{\rm CQS} = \beta_0^{-1}$. While the relationship between the low-frequency base and collector currents *can* be significantly nonlinear, especially at high bias, the resulting effects are less important for the distortion behavior at *high* operating frequencies, unless the value of β_0 is very low. Thus, employing (12), (10) becomes

$$\frac{dq_t}{dt} = \frac{v_s - v_{\rm be}}{R_S^\star} - \frac{i_{\rm cqs}}{\beta_0^\star} \tag{13}$$

where $1/\beta_0^\star \equiv 1/\beta_0 + R_E/R_S^\star$ defines an effective current gain.

Substituting (8) and (9) into (13), and expanding q_t as a Volterra series in v_s

$$q_t = H_1(\omega_a)\circ v_s + H_2(\omega_a,\omega_b)\circ v_s^2 + H_3(\omega_a,\omega_b,\omega_c)\circ v_s^3 \tag{14}$$

where "$\circ$" denotes the Volterra operator [25, p. 549], and where ω_a, ω_b, and ω_c are the parametric input frequencies within the Volterra formalism, one can solve successively for the kernels $H_1(\omega_a)$, $H_2(\omega_a,\omega_b)$, and $H_3(\omega_a,\omega_b,\omega_c)$ in the customary way [28, pp. 178–186]. Writing the output current $i_c \approx i_{\rm cqs}$ as a Volterra series in v_s

$$i_{\rm cqs} = G_1(\omega_a)\circ v_s + G_2(\omega_a,\omega_b)\circ v_s^2 + G_3(\omega_a,\omega_b,\omega_c)\circ v_s^3 \tag{15}$$

and substituting (14) and (15) into (9), one can then obtain the kernels $G_1(\omega_a)$, $G_2(\omega_a,\omega_b)$, and $G_3(\omega_a,\omega_b,\omega_c)$.

III. Second-Order Distortion

A. Kernel

The result for the second-order kernel is

$$G_2(\omega_a,\omega_b) = \frac{1}{2}\left[\frac{\frac{1}{R_S^\star}}{\omega_D + j\omega_a}\right]\left[\frac{\frac{1}{R_S^\star}}{\omega_D + j\omega_b}\right] \times \left\{\frac{\left(\frac{\dot{g}_m}{g_m^3}\right)\hat{\omega}_T^3 + j(\omega_a+\omega_b)\hat{\omega}_T\hat{\omega}_T' R_S^\star}{[\omega_D + j(\omega_a+\omega_b)]R_S^\star}\right\} \tag{16}$$

where ω_D is a critical corner frequency for distortion behavior, given by

$$\omega_D = \hat{\omega}_\beta\left[\frac{r_\pi + R_S + (\beta_0+1)R_E}{R_S + R_E}\right] \tag{17}$$

with $\hat{\omega}_\beta \equiv \hat{\omega}_T/\beta_0$ being the -3-dB frequency for the loaded current gain, and $r_\pi \equiv \beta_0/g_m$ being the intrinsic common-emitter input resistance. Equation (16) will first be used to examine the cancellation theory presented by Maas *et al.* [10], and then it will be compared with experimental and simulation results for the second-order distortion.

B. Current Cancellation

In [10], the authors report on the effects of current cancellation on the device linearity at high frequencies. To examine this phenomenon more closely, it is necessary to write (16) in an alternative form. In terms of the power-series coefficients used in [10], namely, those in the expansions

$$i_{\rm cqs} = \alpha_0\left(g_1 v_{\rm be} + g_2 v_{\rm be}^2 + g_3 v_{\rm be}^3\right) \tag{18}$$

and

$$q_t = c_1 v_{\rm be} + c_2 v_{\rm be}^2 + c_3 v_{\rm be}^3 \tag{19}$$

where $\alpha_0 = \beta_0/(\beta_0+1)$ is the low-frequency common-base current gain, (16) can be written as follows:

$$G_2(\omega_a,\omega_b) = \frac{1}{2}\left[\frac{\frac{1}{R_S^\star}}{\omega_D + j\omega_a}\right]\left[\frac{\frac{1}{R_S^\star}}{\omega_D + j\omega_b}\right] \times \left\{\left[\frac{2\alpha_0}{c_1^3}\right]\frac{g_2 + j(\omega_a+\omega_b)g_1\left(\frac{g_2 c_1}{g_1} - c_2\right)R_S^\star}{[\omega_D + j(\omega_a+\omega_b)]R_S^\star}\right\}. \tag{20}$$

Of interest in (20) is the term

$$j(\omega_a+\omega_b)\left[\frac{2\alpha_0 g_1}{c_1^3}\right]\left(\frac{g_2 c_1}{g_1} - c_2\right)R_S^\star \tag{21}$$

in the numerator of the last factor. As pointed out in [10], this term embodies a subtraction, or cancellation, of second-order current components arising from the device's nonlinear transconductance and nonlinear stored charge; these currents are proportional to g_2 and c_2, respectively. The effect of this cancellation is to minimize the overall distortion by driving the magnitude of (21) to zero. For example, consider an "ideal" transistor, in which the stored charge is directly proportional to the collector current:

$$q_T = \tau_F i_{\mathrm{CQS}} \tag{22}$$

where τ_F is a *constant* transit time. For such a device, $c_1 = \alpha_0 \tau_F g_1$, $c_2 = \alpha_0 \tau_F g_2$, and the cancellation is perfect, so that (21) vanishes completely. More generally, of course, q_T is given by (5), and the cancellation is imperfect. A visualization of the cancellation mechanism is provided in Appendix II.

A comparison of (16) and (20) allows an important insight to be gained into the cancellation phenomenon. The term in (16) corresponding to (21) is clearly

$$j(\omega_a + \omega_b)\hat{\omega}_T \hat{\omega}_T' R_S^{\star} \tag{23}$$

where $\hat{\omega}_T = \alpha_0 g_1/c_1$ and $\hat{\omega}_T' = (2/c_1^2)(g_2 c_1/g_1 - c_2)$. Therefore, *the extent of the cancellation is indicated by the slope of the $\hat{\omega}_T$ versus collector current curve; the cancellation is perfect only where the slope is zero*. In an ideal transistor, where q_T is given by (22), $\hat{\omega}_T = 1/\tau_F$ is a constant, and $\hat{\omega}_T' = 0$ for *all* values of the collector current; in this case, the cancellation is always perfect. More generally, in a real device, q_T is given by (5), and the cancellation is imperfect, *unless* the device is biased at the peak of its $\hat{\omega}_T$ curve (where $\hat{\omega}_T' = 0$). Intuitively, it might then be expected that a real device's distortion performance becomes optimal when it is operated at this point. As will be demonstrated shortly, this is indeed true, but only for the distortion at certain mixing frequencies, and only when the fundamental frequency is sufficiently high.

C. High-Frequency Distortion

1) Simulation: In order to validate (16) for the second-order kernel, we compared its predictions with results from numerical simulation. The simulations were performed with a commercial Volterra solver,[1] into which we implemented a nonlinear small-signal equivalent circuit for the transistor, based on the SPICE Gummel-Poon model [29] under low-level injection. Key SPICE parameter values for the chosen device, which was a representative IBM Si/SiGe HBT [30], [31] working below its peak f_T, are given in Table I. At the chosen operating point, the bias values of collector current and external collector–emitter voltage were 1.2 mA and 1.8 V, respectively, and the f_T was 27 GHz. The external source and load were set to 200 and 50 Ω, respectively.

Fig. 2 shows the magnitude and phase of the output distortion currents at the mixing frequencies $2\omega_1$ and $\omega_1 - \omega_2$, found for a two-tone input at frequencies ω_1 and ω_2, and plotted versus the fundamental frequency $f_1 \equiv \omega_1/2\pi$. The tone spacing was set to 1 MHz, and the source amplitude s corresponded to -30 dBm of available power. Values from simulation are presented along with those from the expressions

$$\tilde{i}_{\mathrm{cqs},2\omega_1} \equiv \frac{1}{2} s^2 G_2(\omega_1, \omega_1) \tag{24}$$

and

$$\tilde{i}_{\mathrm{cqs},\omega_1 - \omega_2} \equiv s^2 G_2(\omega_1, -\omega_2). \tag{25}$$

As shown in Fig. 2, there is good agreement in both magnitude and phase, for both the $2\omega_1$ and $\omega_1 - \omega_2$ outputs. The discrepancy in the phase of the $2\omega_1$ output at high fundamental frequencies (on the order of f_T) occurs because of the neglected terms in (1) and (2). Improved agreement in the phase was obtained when the excess-phase parameter (PTF in SPICE) was set to zero in the simulations; alternatively, it should be possible to get better agreement by simply multiplying (16) by an excess-phase factor, such as one of the form $\exp[-j(\omega_a + \omega_b)\tau]$, where τ is the excess-phase time delay.

2) Output Distortion at $2\omega_1$: Equation (16) can be exploited to gain a deeper understanding of the transistor's high-frequency distortion characteristics. For convenience in this and subsequent discussions, we will assume a two-tone input with tone frequencies at $\omega \equiv \omega_1 \approx \omega_2$ and a tone spacing of $\Delta\omega \equiv \omega_1 - \omega_2$, and then express the results in terms of ω and $\Delta\omega$.

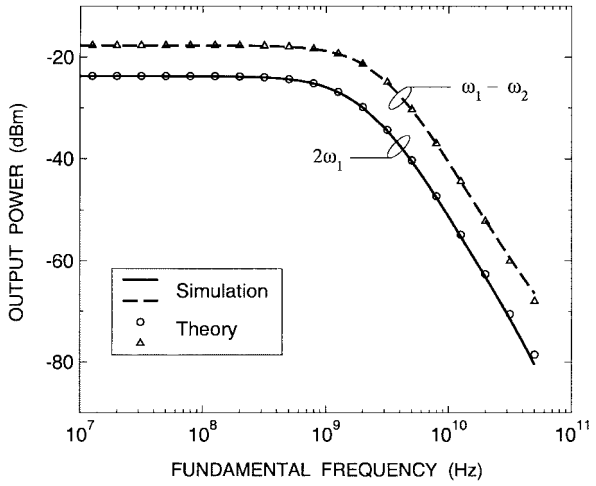
Consider the magnitude of the output distortion at a mixing frequency of $2\omega_1$, given by (24). In the context of the transistor's distortion behavior, "high frequencies" are those satisfying the relation

$$\omega^2 \gg \omega_D^2 \tag{26}$$

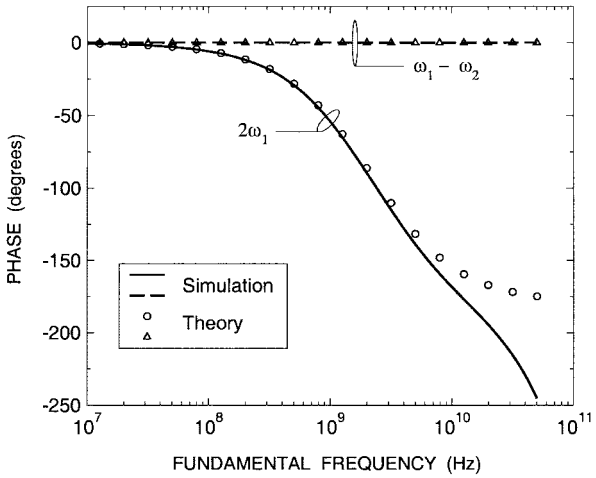
TABLE I
KEY SPICE PARAMETER VALUES FOR AN IBM Si/SiGe HBT [30],[31]

PARAMETER	UNITS	VALUE
IS	A	1.6×10^{17}
NF	—	1.00
T	K	300
BF	—	149.81
VAF	V	117.79
VAR	V	2.10
TF	s	2.15×10^{-12}
PTF	deg	37.00
FC	—	0.5
CJE	F	77.90×10^{-15}
VJE	V	1.48
MJE	—	0.67900
XCJC	—	0.86891
CJC	F	23.95×10^{-15}
VJC	V	0.38600
MJC	—	0.11000
CJS	F	23.05×10^{-15}
VJS	V	0.34000
MJS	—	0.18700
RB	Ω	29.63
RBM	Ω	10.69
IRB	A	0
RC	Ω	3.11
RE	Ω	1.39

[1]Microwave Office Software, Applied Wave Research Inc., El Segundo, CA. [Online]. Available: http://www.mwoffice.com



(a)



(b)

Fig. 2. Simulated and computed [using (24) and (25)] values of the (a) magnitude and (b) phase of the output distortion at the mixing frequencies $2\omega_1$ and $\omega_1 - \omega_2$, plotted versus the fundamental frequency $f_1 \equiv \omega_1/2\pi$. The device is the Si/SiGe HBT described in Table I, and $f_T \approx 27$ GHz at the chosen operating point. The magnitude values are quoted in terms of the power dissipated in a 50-Ω external load.

where ω_D is the critical corner frequency specified in (17). For such fundamental frequencies, (16) and (24) imply

$$|\tilde{i}_{\mathrm{cqs},2\omega_1}| \approx \frac{s^2}{4\omega^3 R_S^{\star 3}} \left[\frac{1}{4}\frac{\dot{g}_m^2}{g_m^6}\hat{\omega}_T^6 + \omega^2 \hat{\omega}_T^2 \hat{\omega}_T'^2 R_S^{\star 2}\right]^{1/2}. \quad (27)$$

The distortion at $2\omega_1$ will thus roll off with fundamental frequency at a rate between -60 and -40 dB/decade, depending on the magnitudes of the two terms in (27). For operation of the transistor at the peak of its $\hat{\omega}_T$ curve, where $\hat{\omega}_T' = 0$, the second term in (27) vanishes, and the rate will be -60 dB/decade. At other operating points, the second term in (27) will eventually dominate (for sufficiently large ω), and the rate will lie closer to -40 dB/decade, as illustrated in Fig. 2(a). An idea of how large ω needs to be for this to occur can be obtained by examining a simplified case. For example, consider a "near-ideal" transistor, where it is possible to write

$$q_T = \tau_F i_{\mathrm{CQS}} + C_{\mathrm{je}} v_{\mathrm{BE}} + C_{\mathrm{jc}} v_{\mathrm{BC}} \quad (28)$$

$$i_{\mathrm{CQS}} = I_S \exp\left[\frac{v_{\mathrm{BE}}}{nV_t}\right] \quad (29)$$

and

$$\hat{\omega}_T^{-1} = \frac{C_{\mathrm{je}} + C_{\mathrm{jc}}}{g_m} + \tau_F + (R_L + R_E) C_{\mathrm{jc}} \quad (30)$$

with C_{je}, C_{jc}, τ_F, I_S, and n all being *constant*, and where V_t is the thermal voltage. For such a device, one then finds that the second term in (27) will dominate when

$$\omega^2 \gg \omega_{D2}^2 \quad (31)$$

where

$$\omega_{D2}^2 \equiv \frac{\left(\frac{\dot{g}_m^2}{g_m^6}\right)\hat{\omega}_T^4}{4\hat{\omega}_T'^2 R_S^{\star 2}} = \frac{1}{4\left[(R_S + R_E)(C_{\mathrm{je}} + C_{\mathrm{jc}})\right]^2}. \quad (32)$$

More generally, with a real device, (28)–(30) can be viewed as *crude* representations of the transistor charge, collector current, and unity–current–gain frequency prior to the onset of high-current effects, and the right-hand side of (32) can be considered as a rough estimate of ω_{D2}^2 that applies for operating points up to the peak in the $\hat{\omega}_T$ curve.

In general, the critical frequencies ω_D and ω_{D2} appearing in (26) and (31) can be quite high. The value of ω_D in (17) is clearly greater than $\hat{\omega}_\beta$, the -3-dB frequency for the loaded current gain; in fact, inspection of (17) reveals that ω_D will fall well below the transistor's $\hat{\omega}_T \equiv \hat{\omega}_\beta \beta_0$ only when the source resistance is large and the emitter resistance is small. For example, for the device considered in Table I and Fig. 2, it turns out $\omega_D \approx \hat{\omega}_T/8.4$. If one decreases the external source resistance from 200 to 50 Ω, and adds 10 Ω of external emitter degeneration, then ω_D will become even larger, with a value of approximately $\hat{\omega}_T/2.4$. Similar comments apply to ω_{D2}, given by (32) (although, in this case, a low value of R_E is not necessary). Such points should be borne in mind when deciding whether a given operating frequency is "high" or "low" from a distortion perspective.

When both (26) and (31) *are* satisfied, the high-frequency value of $|\tilde{i}_{\mathrm{cqs},2\omega_1}|$ that follows from (27) is

$$|\tilde{i}_{\mathrm{cqs},2\omega_1}| \approx \frac{s^2}{4\omega^2 R_S^{\star 2}} \hat{\omega}_T |\hat{\omega}_T'|. \quad (33)$$

Since the fundamental output at high frequencies can be written as

$$|\tilde{i}_{\mathrm{cqs},\omega_1}| \equiv |sG_1(\omega_1)| = \left|\frac{\frac{s\hat{\omega}_T}{R_S^\star}}{\omega_D + j\omega_1}\right| \approx \frac{s\hat{\omega}_T}{\omega R_S^\star} \quad (34)$$

the transistor's output-intercept point becomes

$$\mathrm{OIP}_2(2\omega_1) \approx \left|\frac{4\hat{\omega}_T}{\hat{\omega}_T'}\right|. \quad (35)$$

Thus, *the $\hat{\omega}_T$ versus collector current curve should be as constant as possible to maximize the high-frequency value of* $\mathrm{OIP}_2(2\omega_1)$. Moreover, $\mathrm{OIP}_2(2\omega_1)$ *will reach its absolute maximum value for operation of the transistor near the peak of the curve*. These points are consistent with the experimental observations of Schröter *et al.* [20, p. 1533], who examined integrated silicon transistors, as well as our own measurement results for an InGaP/GaAs HBT, presented in Fig. 3.

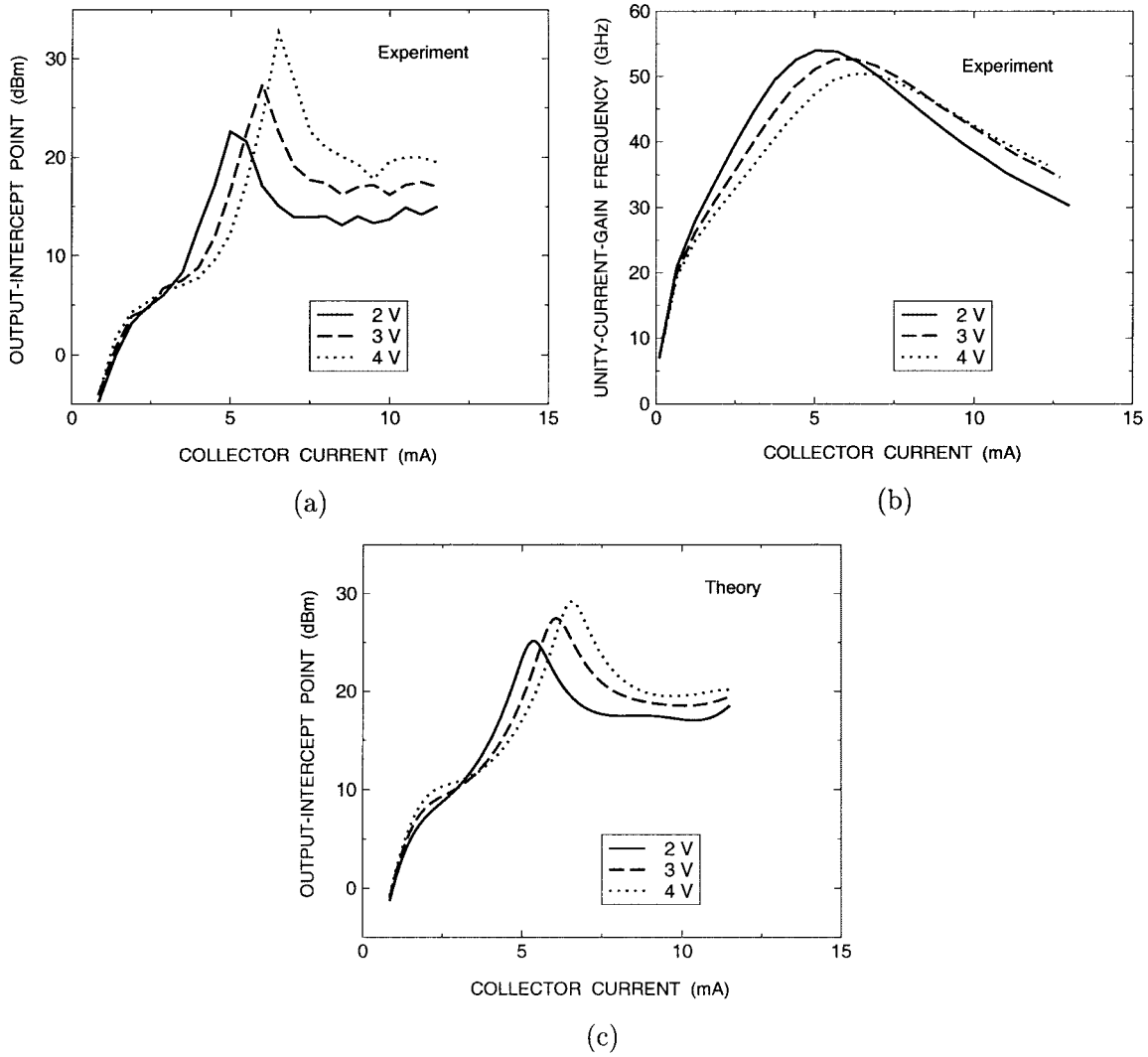


Fig. 3. (a) Measured values of $\mathrm{OIP}_2(2\omega_1)$ versus collector current for an InGaP/GaAs HBT at different collector–emitter bias voltages. The device is "HBT-C" described in [16, Sec. V-A], and the values are quoted in terms of the power dissipated in the external measurement load of 50 Ω. The operating frequency is 5 GHz. (b) Corresponding values of the extrapolated short-circuit unity–current–gain frequency $f_T \equiv \omega_T/2\pi$. (c) Values of $\mathrm{OIP}_2(2\omega_1)$, for the same device, found from (37) as described in the text.

Shown in Fig. 3(a) are high-frequency values of $\mathrm{OIP}_2(2\omega_1)$ versus collector current, at different collector–emitter bias voltages, for the InGaP/GaAs device described as "HBT-C" in [16, Sec. V-A]. The measurement approach was the same as that in [16, Sec. II-B], and the fundamental frequency was 5 GHz. The external source and load were each set to 50 Ω. Rigorously, $\mathrm{OIP}_2(2\omega_1)$ at each point in Fig. 3(a) is related to the loaded unity–current–gain frequency $\hat{\omega}_T$ and its first derivative $\hat{\omega}_T'$, as specified by (35). However, as discussed earlier in Section II-B, for the purposes of examining the qualitative trends predicted by this expression, the short-circuit values ω_T and ω_T' can be used instead; these are readily found from the plot in Fig. 3(b).

Comparing Fig. 3(a) and (b) verifies the main predictions of (35). For example, at each collector–emitter bias considered, the peak in $\mathrm{OIP}_2(2\omega_1)$ coincides with the peak in the corresponding ω_T characteristics. For operation at high currents (following the peaks in the ω_T curves), $\mathrm{OIP}_2(2\omega_1)$ is the worst for a collector–emitter voltage of 2 V, which also exhibits the lowest values of ω_T and the highest values of ω_T'. Finally, setting $\hat{\omega}_T' = 0$, and combining (27) and (34), yields the value of $\mathrm{OIP}_2(2\omega_1)$ to expect when the device is operated at the peak of its $\hat{\omega}_T$ curve:

$$\mathrm{OIP}_2(2\omega_1)|_{\hat{\omega}_T'=0} = (8\omega R_S^\star)\left[\frac{g_m^3}{\dot{g}_m}\right]\left[\frac{1}{\hat{\omega}_T}\right]\Bigg|_{\hat{\omega}_T'=0}. \quad (36)$$

If we assume, for simplicity, that the collector current follows the ideal law (29), then (36) suggests that $\mathrm{OIP}_2(2\omega_1)|_{\hat{\omega}_T'=0}$ will be higher for lower values of peak unity–current–gain frequency occurring at higher currents; the results in Fig. 3(a) and (b) are consistent with this prediction.

A more detailed comparison of the analysis in this study with the results from experiment is facilitated by considering Fig. 3(c), where we have plotted values of $\mathrm{OIP}_2(2\omega_1)$ obtained from the expression

$$\mathrm{OIP}_2(2\omega_1) \equiv \frac{2|G_1(\omega_1)|^2}{|G_2(\omega_1,\omega_1)|} \quad (37)$$

with $G_1(\omega_1) = (\hat{\omega}_T/R_S^\star)/[\omega_D + j\omega_1]$ and $G_2(\omega_1,\omega_1)$ found from (16). Values of $\hat{\omega}_T$ and $\hat{\omega}_T'$ to use on the right-hand side of (37) were estimated by fitting polynomials to the measured ω_T

TABLE II
PARAMETER VALUES FOR THE MEASURED DEVICES

PARAMETER	UNITS	HBT-A	HBT-B	HBT-C
R_B	Ω	60	80	80
R_E	Ω	3.5	3.5	3.5
β_0	—	150	150	150

curves in Fig. 3(b). In each case, the degree of the fitting polynomial was chosen so that the fitted curve would not only return the measured ω_T values with negligible error, but also reproduce the measured trends in the ω_T versus collector current behavior with a minimum amount of "polynomial wiggle" [32, p. 231]. Polynomial wiggle must be avoided to extract derivatives reliably from the measured data. The other parameter values needed to evaluate the right-hand side of (37) were obtained as follows: g_m and its derivatives were found by assuming the collector current obeyed the ideal law (29) with $n = 1$ and $V_t = 0.0259\,\mathrm{V}$; and $R_S = 50\,\Omega + R_B$, R_E, and β_0 were found from knowledge of the device parasitics and current gain, with the relevant values given in Table II. Comparing Fig. 3(a) and (c) makes it clear that (37) is capable of predicting all the measured trends.

While the results in Fig. 3(a) were obtained for operation at 5 GHz, further measurements of $\mathrm{OIP}_2(2\omega_1)$ at 1 GHz, shown in Fig. 4(a), revealed quite different trends; correspondingly, evaluation of (37), carried out in the same manner as that for Fig. 3(c), except at an operating frequency of 1 GHz, yields the curves shown in Fig. 4(b), which again matches experiment. This behavior is to be expected, since the trends in Fig. 3 are predicted to occur only at sufficiently high operating frequencies (above ω_D and ω_{D2}).

3) Output Distortion at $\omega_1 - \omega_2$: Equation (16) can also be used to examine the magnitude of the output distortion at a mixing frequency of $\omega_1 - \omega_2$, given by (25). In this case, in addition to high tone frequencies, as specified by (26), it will be assumed that the tone spacing is small:

$$\Delta\omega^2 \ll \omega_D^2 \tag{38}$$

and

$$\Delta\omega^2 \ll \omega_{D2}^2. \tag{39}$$

Equations (16) and (25) then imply

$$|\tilde{i}_{\mathrm{cqs},\omega_1-\omega_2}| \approx \frac{s^2 \dot{g}_m \hat{\omega}_T^2}{2\omega^2 g_m^2 R_S^{\star 2}\left[1 + g_m\left(\frac{R_S}{\beta_0} + \frac{R_E}{\alpha_0}\right)\right]} \tag{40}$$

where (17) has been used to eliminate ω_D. Thus, the magnitude of the distortion at $\omega_1 - \omega_2$ will always exhibit a -40-dB/decade rolloff with fundamental frequency, as illustrated in Fig. 2(a), independent of the transistor's operating point. The corresponding output-intercept point becomes

$$\mathrm{OIP}_2(\omega_1 - \omega_2) \approx \frac{2g_m^2}{\dot{g}_m}\left[1 + g_m\left(\frac{R_S}{\beta_0} + \frac{R_E}{\alpha_0}\right)\right] \tag{41}$$

which can easily be shown to be the *same* value as that occurring at low fundamental frequencies ($\omega^2 \ll \omega_D^2$). Note that, unlike the situation at $2\omega_1$, neither the distortion in (40), nor the corresponding output-intercept point in (41), is related to $\hat{\omega}_T'$, and hence no advantage is gained by operating the transistor at the peak of its $\hat{\omega}_T$ curve. This can be understood by realizing that the current-cancellation mechanism discussed earlier (between the nonlinear stored charge and nonlinear transconductance) does not control the output distortion when (39) is satisfied.

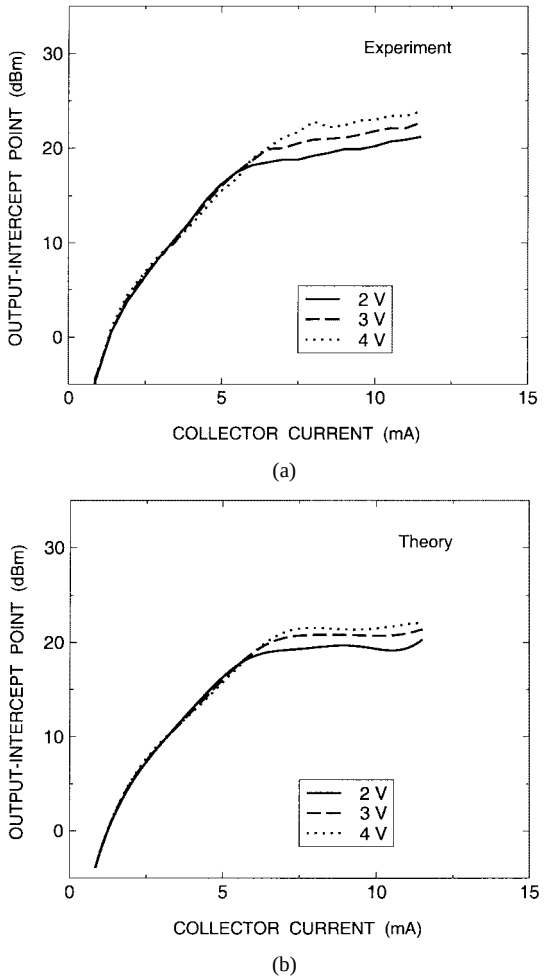


Fig. 4. (a) Measured values of $\mathrm{OIP}_2(2\omega_1)$ for the same device and the same collector–emitter bias voltages as in Fig. 3, but at an operating frequency of 1 GHz; the values are quoted in terms of the power dissipated in the external measurement load of 50 Ω. (b) Values of $\mathrm{OIP}_2(2\omega_1)$ predicted by (37), evaluated as described in the text.

4) Phase Behavior: The phase behavior of the distortion currents at the output follows from (16), (24), and (25) in a straightforward manner. At sufficiently high fundamental frequencies, the phase of the $2\omega_1$ output will ideally fall toward -180° (neglecting excess phase), as illustrated in Fig. 2(b); this will become -270° for operation of the transistor at the peak of its $\hat{\omega}_T$ curve. On the other hand, for sufficiently small tone spacing, the phase of the $\omega_1 - \omega_2$ output will remain at zero, even if the tone frequencies themselves are high; this is also illustrated in Fig. 2(b).

IV. THIRD-ORDER INTERMODULATION DISTORTION

A. Kernel

Of particular interest for modern radio-frequency applications is the third-order intermodulation distortion at $2\omega_1 - \omega_2$, assuming, as before, two equal-amplitude input tones at the fre-

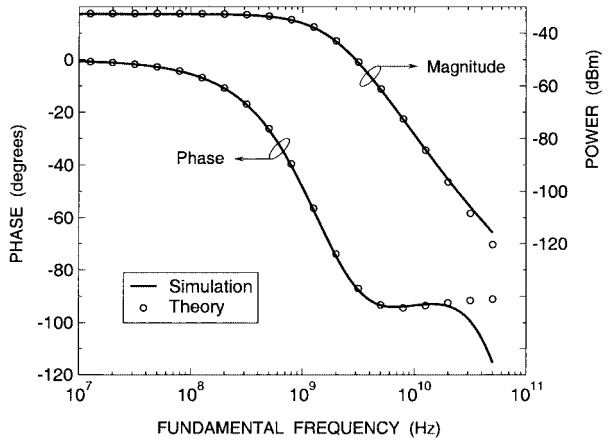


Fig. 5. Simulated and computed [using (44)] magnitude and phase of the output distortion at $2\omega_1 - \omega_2$, plotted versus the fundamental frequency $f_1 \equiv \omega_1/2\pi$. The device is the Si/SiGe HBT described in Table I, and $f_T \approx 27$ GHz at the chosen operating point. The magnitude values are quoted in terms of the power dissipated in a 50-Ω external load.

quencies ω_1 and ω_2. The corresponding value of the third-order Volterra kernel appearing in (15) is (42), shown at the bottom of the following page, where $\kappa_1 \equiv \omega_1 - \omega_2$, $\kappa_2 \equiv 2\omega_1$, and $\kappa_3 \equiv 2\omega_1 - \omega_2$ are various mixing frequencies, and where the symbol Φ is given by the expression

$$\Phi = \left[\frac{1}{R_S^\star}\right]\left[\left(\frac{1}{2}\frac{\dot{g}_m}{g_m^3}\right)\hat{\omega}_T^3\hat{\omega}_T' + \left(\frac{1}{6}\frac{\ddot{g}_m}{g_m^4} - \frac{1}{2}\frac{\dot{g}_m^2}{g_m^5}\right)\hat{\omega}_T^4\right]. \quad (43)$$

The output distortion is then

$$\tilde{i}_{\mathrm{cqs},2\omega_1-\omega_2} \equiv \frac{3}{4}s^3 G_3(\omega_1, \omega_1, -\omega_2) \quad (44)$$

and the corresponding output-intercept point is

$$\mathrm{OIP}_3(2\omega_1 - \omega_2) \equiv \sqrt{\frac{4}{3}} \cdot \frac{|G_1(\omega_1)|^{3/2}}{|G_3(\omega_1, \omega_1, -\omega_2)|^{1/2}} \quad (45)$$

where $G_1(\omega_1)$ was given below (37).

B. Comparison With Simulation and Experiment

Fig. 5 shows the magnitude and phase of the distortion computed from (44) along with simulation results; the parameter values and conditions are the same as those in Section III-C.1 for examining the second-order distortion. There is good agreement in both magnitude and phase, at both low and high fundamental frequencies. As with the second-order results, the discrepancy in the phase at frequencies on the order of $f_T \approx 27$ GHz and above occurs because of the neglected terms in (1) and (2).

Fig. 6(a) shows a comparison of the output-intercept point computed from (45) with results obtained from experiment. The measured devices are "HBT-A," "HBT-B," and "HBT-C," described in [16, Sec. V-A], and the measurement approach was the same as that in [16, Sec. II-B]. The fundamental frequency was 5 GHz and the tone spacing was 1 MHz, and the external source and load resistances were set to 50 and 260 Ω, respectively. The collector–emitter bias voltage was 4 V for HBT-A, and 5 V for HBT-B and HBT-C. Parameter values to use on the right-hand side of (45) were found in the same manner as previously described for the evaluation of (37), with

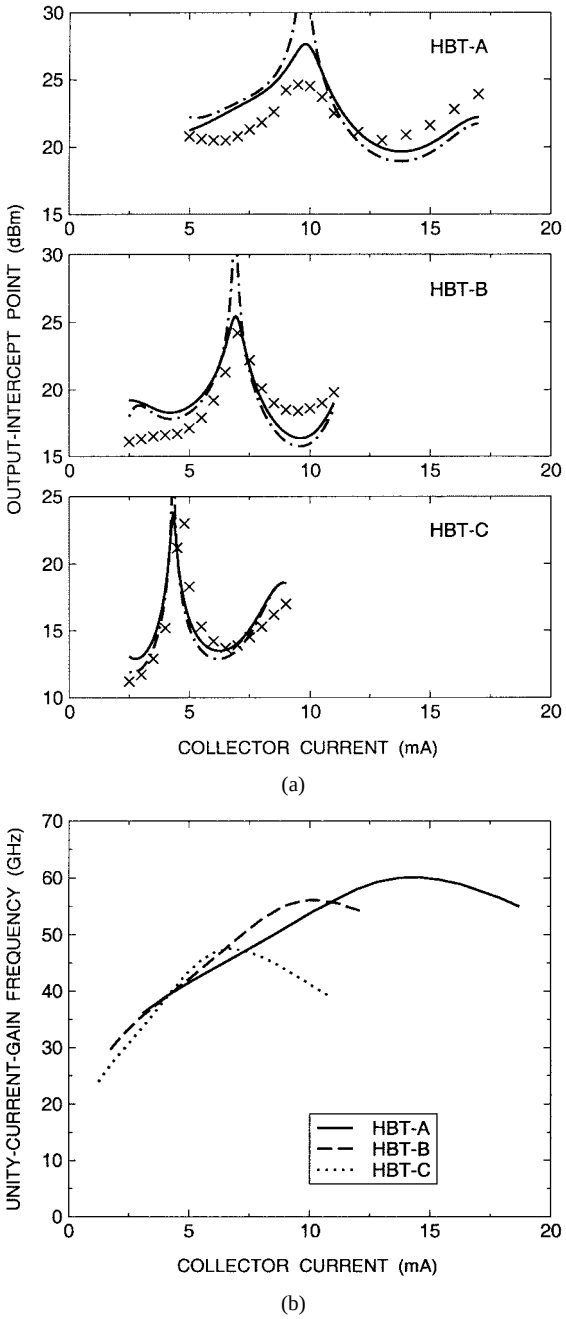


Fig. 6. (a) Measured (symbols) and computed values, using (45) (solid lines) and (53) (stippled lines), of $\mathrm{OIP}_3(2\omega_1 - \omega_2)$ versus collector current for the InGaP/GaAs devices described as "HBT-A," "HBT-B," and "HBT-C" in [16, Sec. V-A]. The operating conditions are described in the text, and the values are quoted in terms of the power dissipated in the external measurement load of 260 Ω. (b) Measured values of the extrapolated short-circuit unity–current–gain frequency $f_T \equiv \omega_T/2\pi$ corresponding to the measured values of $\mathrm{OIP}_3(2\omega_1 - \omega_2)$ in (a).

the ω_T data extracted from the curves in Fig. 6(b). Inspection of Fig. 6(a) reveals that there is good agreement between (45) and the measured values; the complex bias dependence of $\mathrm{OIP}_3(2\omega_1 - \omega_2)$, including the occurrence of distinct peaks, is predicted correctly for all three devices.

The involved nature of $G_3(\omega_1, \omega_1, -\omega_2)$ in (42) makes it difficult to obtain insight into all aspects of the distortion behavior in Figs. 5 and 6. However, some useful information can be found by examining a few asymptotic limits of this expression.

C. Asymptotic Limits

1) High-Frequency Limit: In the limit of infinite tone frequencies ($\omega \equiv \omega_1 \approx \omega_2 \to \infty$) and vanishing tone spacing ($\Delta\omega \equiv \omega_1 - \omega_2 \to 0$), the magnitude of (42) takes on a value $G_{3\mathrm{HF}}$ given by

$$G_{3\mathrm{HF}} \equiv \lim_{\substack{\omega\to\infty \\ \Delta\omega\to 0}} |G_3(\omega_1,\omega_1,-\omega_2)| = \frac{1}{\omega^3 R_S^{\star 3}} \left| -\frac{1}{6}\hat{\omega}_T^2\hat{\omega}_T'' + \hat{\omega}_T\hat{\omega}_T' \times \left\{ \frac{1}{6}\hat{\omega}_T' - \frac{1}{3}\frac{\dot{g}_m}{g_m^2}\left[\frac{\hat{\omega}_T}{1+g_m\left(\frac{R_S}{\beta_0}+\frac{R_E}{\alpha_0}\right)}\right]\right\}\right|. \tag{46}$$

At sufficiently high tone frequencies, and for sufficiently small tone spacing, the output distortion at $2\omega_1-\omega_2$ will hence roll off with fundamental frequency at -60 dB/decade, as illustrated in Fig. 5(a).

2) Ideal-Transistor Limit: In the case of an ideal transistor, where the charge function is given by (22), both $\hat{\omega}_T'$ and $\hat{\omega}_T''$ vanish. While (46) is thus zero, the output distortion in such a device actually takes on a *nonzero* value determined by the high-frequency limit of (42) found under the constraint that $\hat{\omega}_T', \hat{\omega}_T'' = 0$:

$$G_{3\mathrm{IT}} \equiv \lim_{\substack{\omega\to\infty \\ \Delta\omega\to 0}} |G_3(\omega_1,\omega_1,-\omega_2)|_{\hat{\omega}_T',\hat{\omega}_T''=0} = \frac{\left(\frac{\hat{\omega}_T}{g_m}\right)^4}{\omega^4 R_S^{\star 4}} \left| \frac{1}{6}\ddot{g}_m - \frac{1}{2}\frac{\dot{g}_m^2}{g_m}\left[\frac{\frac{1}{3}+g_m\left(\frac{R_S}{\beta_0}+\frac{R_E}{\alpha_0}\right)}{1+g_m\left(\frac{R_S}{\beta_0}+\frac{R_E}{\alpha_0}\right)}\right]\right| \tag{47}$$

which implies a -80-dB/decade rolloff for the output distortion. Note that this result vanishes completely *only* if, in addition to (22), the collector current is governed by the ideal law (29), the low-frequency gain is infinitely high ($\beta_0 \to \infty$), and there is no emitter degeneration ($R_E = 0$).

3) Actual Value at High Frequencies: The actual value of $|G_3(\omega_1,\omega_1,-\omega_2)|$ at high frequencies can be written in terms of the limits in (46) and (47), plus an extra term. Provided the operating frequency is well above the critical distortion frequency ω_D, as specified by (26), and the tone spacing is small

$$\Delta\omega \ll \omega_D \tag{48}$$

then expanding (42), and using (26) and (48) to discard the less important terms, after considerable manipulation, one finds

$$|G_3(\omega_1,\omega_1,-\omega_2)|^2 \approx G_{3\mathrm{HF}}^2 + G_{3\mathrm{IT}}^2 + \xi \tag{49}$$

where ξ is an additional term given by

$$\xi = \frac{\theta}{\omega^8 R_S^{\star 8}} \tag{50}$$

and θ can be written in terms of $\hat{\omega}_T$, g_m, and their derivatives. The result for θ, which is somewhat untidy, is given in Appendix III; note that $\theta = 0$ for an ideal transistor (where $\hat{\omega}_T', \hat{\omega}_T'' = 0$), and that θ simplifies considerably when the device is operating at the peak of its $\hat{\omega}_T$ curve (where $\hat{\omega}_T' = 0$). When θ is nonzero, it can be positive or negative, so that ξ in (50) provides a positive or negative adjustment to the limits in (46) and (47) that diminishes with fundamental frequency at -80 dB/decade.

D. Discussion

Using (34) to write $G_1(\omega_1) \approx \hat{\omega}_T/\omega R_S^\star$ and employing (49) for $G_3(\omega_1,\omega_1,-\omega_2)$, (45) yields

$$\mathrm{OIP}_3(2\omega_1-\omega_2) \approx \sqrt{\frac{4}{3}} \cdot \frac{\left(\frac{\hat{\omega}_T}{\omega R_S^\star}\right)^{3/2}}{(G_{3\mathrm{HF}}^2 + G_{3\mathrm{IT}}^2 + \xi)^{1/4}} \tag{51}$$

as the value of the output-intercept point at high frequencies. In general, all three terms in the denominator must be retained, making (51) quite involved. However, further insight into this expression can be obtained by extending the cancellation theory proposed by Maas *et al.* [10].

As detailed in Appendix IV, *the limit $G_{3\mathrm{HF}}$ embodies a cancellation of third-order intermodulation currents generated by the transistor's nonlinear transconductance and nonlinear stored charge*; furthermore, $G_{3\mathrm{HF}}$ can be expressed in terms of the series coefficients in (18) and (19), and in terms of the lower order mixing voltages appearing across the transistor's base–emitter terminals, as follows:

$$G_{3\mathrm{HF}} = \lim_{\substack{\omega\to\infty \\ \Delta\omega\to 0}} \left| \left[\frac{4}{3s^3}\right]\left[\tilde{i}_{\mathrm{cqs},2\omega_1-\omega_2}^{\mathrm{NL}} - \frac{\alpha_0 g_1}{j(2\omega_1-\omega_2)c_1}\times\tilde{i}_{q_t,2\omega_1-\omega_2}\right]\right| = \lim_{\substack{\omega\to\infty \\ \Delta\omega\to 0}} \left| \left[\frac{\alpha_0 g_1}{s^3 c_1}\right]\left[\left(\frac{g_3 c_1}{g_1}-c_3\right)\tilde{v}_{\mathrm{be},\omega_1}^2\tilde{v}_{\mathrm{be},\omega_2}^* + \frac{4}{3}\left(\frac{g_2 c_1}{g_1}-c_2\right) \cdot \left(\tilde{v}_{\mathrm{be},2\omega_1}\tilde{v}_{\mathrm{be},\omega_2}^* + \tilde{v}_{\mathrm{be},\omega_1}\tilde{v}_{\mathrm{be},\omega_1-\omega_2}\right)\right]\right| \tag{52}$$

$$G_3(\omega_1,\omega_1,-\omega_2) = \left[\frac{\frac{1}{R_S^\star}}{\omega_D+j\omega_1}\right]^2\left[\frac{\frac{1}{R_S^\star}}{\omega_D-j\omega_2}\right] \times \frac{\left\{\left[\Phi + j\frac{(\hat{\omega}_T'^2+\hat{\omega}_T\hat{\omega}_T'')\hat{\omega}_T}{6}\kappa_3\right](\omega_D+j\kappa_1)(\omega_D+j\kappa_2) - \left[\omega_D\hat{\omega}_T' - \left(\frac{1}{R_S^\star}\right)\left(\frac{\dot{g}_m}{g_m^3}\right)\hat{\omega}_T^2\right]\left[\omega_D + j\frac{2\kappa_2+\kappa_1}{3}\right]\left[\frac{1}{2}\left(\frac{1}{R_S^\star}\right)\left(\frac{\dot{g}_m}{g_m^3}\right)\hat{\omega}_T^3 + j\frac{\hat{\omega}_T\hat{\omega}_T'}{2}\kappa_3\right]\right\}}{(\omega_D+j\kappa_1)(\omega_D+j\kappa_2)(\omega_D+j\kappa_3)} \tag{42}$$

where $\tilde{i}^{\rm NL}_{cqs,2\omega_1-\omega_2}$ and $\tilde{i}_{q_t,2\omega_1-\omega_2}$ refer to the intermodulation currents, at a mixing frequency of $2\omega_1-\omega_2$, generated by the nonlinear transconductance and the nonlinear stored charge, respectively. In the case of an ideal transistor, where the charge function is given by (22), the cancellation is *always* perfect and (52) *always* vanishes, since $g_2c_1/g_1-c_2=0$ and $g_3c_1/g_1-c_3=0$; alternatively, $\hat{\omega}_T'=0$ and $\hat{\omega}_T''=0$, so that (46) always vanishes. More practically, in a real transistor, $G_{3\rm HF}$ will vanish and the cancellation will be perfect *only* when the bias point is chosen such that the coefficients and lower order mixing voltages appearing in (52) have the appropriate magnitude and phase; mathematically, a condition for this to occur, in terms of g_m, $\hat{\omega}_T$, and their derivatives, can be found by setting the right-hand side of (46) to zero. Generally, no simple solution exists for the resulting equation, but if we assume that the collector current is governed by the ideal law (29), then the required information is contained in the transistor's $\hat{\omega}_T$ curve.

Intuitively, it might be expected that $G_{3\rm HF}$ is the most important term in the denominator of (51), since it represents the effects of third-order cancellation in the high-frequency limit. The corresponding assumption has already been shown to be true for the second-order distortion at $2\omega_1$; referring back to Section III-C.2, recall (33) and (35), and note that these expressions essentially represent the distortion and output-intercept point as $\omega\to\infty$, and that they are both governed by $\hat{\omega}_T'$, the sole determinant of cancellation in the second-order case. Thus, neglecting the last two terms in the denominator of (51), the output-intercept point becomes

$$\mathrm{OIP}_3(2\omega_1-\omega_2)\approx \mathrm{OIP}_{3\rm HF}\equiv\sqrt{\frac{4}{3}}\cdot\frac{\left(\frac{\hat{\omega}_T}{\omega R_S'}\right)^{3/2}}{G_{3\rm HF}^{1/2}}. \qquad (53)$$

This result predicts that $\mathrm{OIP}_3(2\omega_1-\omega_2)$ will become a constant at sufficiently high frequencies, and is hence consistent with previous experimental observations [10, p. 445], [33, Fig. 4]. However, its utility is best demonstrated by applying it to our own measurements.

In Fig. 6(a), we have superimposed the predictions of (53), evaluated in the same way as (45), onto the existing curves, and it becomes evident that (53) is capable of predicting all the qualitative trends. More importantly, it is clear that *the peaks in* $\mathrm{OIP}_3(2\omega_1-\omega_2)$ *occur at those bias points where* $G_{3\rm HF}\to 0$ *and* $\mathrm{OIP}_{3\rm HF}\to\infty$, *where there is perfect cancellation between the intermodulation distortion currents generated by the nonlinear transconductance and nonlinear stored charge*. These observations are further validated by the results in Fig. 7(a), where we have compared the predictions of (53), evaluated using the ω_T curves in Fig. 7(b), with experimental data for the $\mathrm{OIP}_3(2\omega_1-\omega_2)$ of HBT-C working at two collector–emitter bias voltages. When the collector–emitter bias is changed, the peak in $\mathrm{OIP}_3(2\omega_1-\omega_2)$ shifts, and this is correctly predicted by (53).

Equation (53) also leads to a useful conclusion regarding the distortion performance at the peak of the $\hat{\omega}_T$ curve, where the high-frequency gain is the highest. Setting $\hat{\omega}_T'=0$, and employing (46) in (53), yields

$$\mathrm{OIP}_3(2\omega_1-\omega_2)|_{\hat{\omega}_T'=0}\approx \mathrm{OIP}_{3\rm HF}|_{\hat{\omega}_T'=0}\equiv\left|\frac{8\hat{\omega}_T}{\hat{\omega}_T''}\right|^{1/2}\Bigg|_{\hat{\omega}_T'=0}. \qquad (54)$$

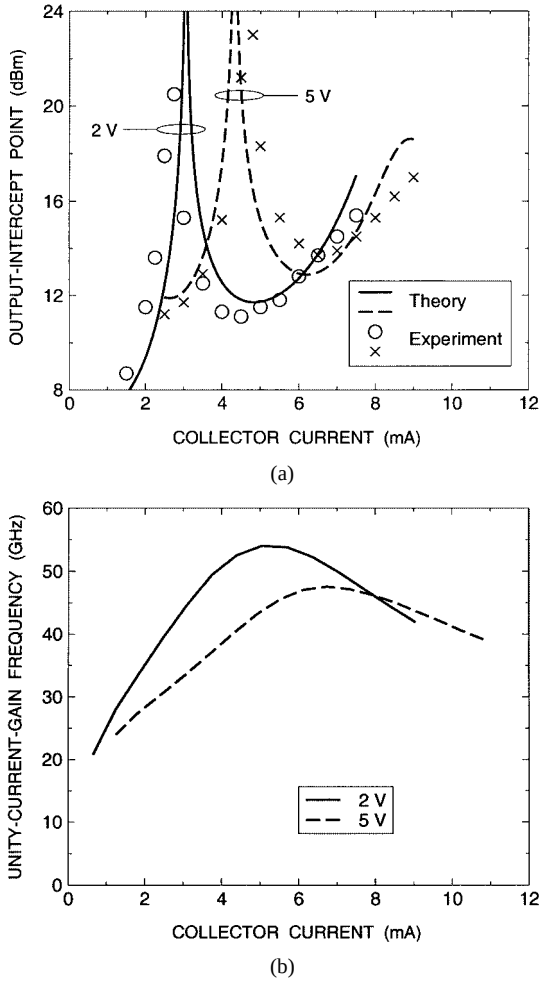


Fig. 7. (a) Measured and computed [using (53)] values of $\mathrm{OIP}_3(2\omega_1-\omega_2)$ versus collector current, at two different collector–emitter bias voltages, for the InGaP/GaAs device described as "HBT-C" in [16, Sec. V-A]. The values are quoted in terms of the power dissipated in the external measurement load of 260 Ω. (b) Measured values of the extrapolated short-circuit unity–current–gain frequency $f_T\equiv\omega_T/2\pi$ corresponding to the measured values of $\mathrm{OIP}_3(2\omega_1-\omega_2)$ in (a).

Therefore, *for optimum distortion performance when the transistor is operated at the peak of its* $\hat{\omega}_T$ *curve, the curve should be as flat as possible*. Since $\hat{\omega}_T'=0$ at the peak, this merely reflects the fact that as $|\hat{\omega}_T''|\to 0$, the transistor behaves more ideally, causing the third-order current cancellation between the nonlinear transconductance and the nonlinear stored charge to become more complete, and the distortion performance to improve. The results in Fig. 6 are consistent with (54). For operation of the transistor near the peak of its ω_T curve, HBT-A exhibits the best performance, and HBT-C exhibits the worst performance; correspondingly, HBT-A has the highest value of peak ω_T and the least curvature in its ω_T plot, and HBT-C has the lowest value of peak ω_T and the greatest curvature in its ω_T plot.

V. Conclusions

The following conclusions can be drawn from this study of high-frequency distortion in bipolar transistors.

1) By employing the charge-control approach originally suggested by Poon and Narayanan [5]–[7], it is possible

to relate the transistor's distortion to its transconductance g_m, its "loaded" unity–current–gain frequency $\hat{\omega}_T$, and the derivatives of these quantities with respect to base–emitter voltage and collector current, respectively. For the purposes of identifying qualitative trends in the distortion performance, a plot of the short-circuit unity–current–gain frequency versus collector current can usually be used in lieu of knowledge of $\hat{\omega}_T$ and its derivatives.

2) For the second-order distortion, the extent of current cancellation between the nonlinear transconductance and nonlinear stored charge [10] is determined by the slope of the $\hat{\omega}_T$ versus collector current curve; the cancellation is perfect only where the slope is zero, i.e., at the peak of the curve.
3) For operating frequencies sufficiently above a critical distortion frequency ω_D, the magnitude of the second-order distortion at $2\omega_1$ will fall with fundamental frequency at a rate between -40 and -60 dB/decade, and the phase will fall toward -180 and -270°, with each of the latter quantities occurring for operation of the transistor at the peak of its $\hat{\omega}_T$ curve, which also yields the maximum value for the output-intercept point $\mathrm{OIP}_2(2\omega_1)$.
4) For high operating frequencies (above ω_D) and small tone spacing (below ω_D and ω_{D2}), the magnitude of the second-order distortion at $\omega_1 - \omega_2$ will always fall at -40 dB/decade, regardless of the transistor's operating point. The phase of the distortion is zero, even at high fundamental frequencies, and the output-intercept point $\mathrm{OIP}_2(\omega_1 - \omega_2)$ has the same value at both low and high fundamental frequencies.
5) For high operating frequencies (above ω_D) and small tone spacing (below ω_D), the magnitude of the third-order intermodulation distortion at $2\omega_1 - \omega_2$ will roll off with fundamental frequency at a rate between -60 and -80 dB/decade, the latter being achieved at those points on the $\hat{\omega}_T$ curve where the high-frequency limit $G_{3\mathrm{HF}}$ in (46) vanishes.
6) The limit $G_{3\mathrm{HF}}$ embodies a cancellation of third-order intermodulation currents generated by the nonlinear transconductance and the nonlinear stored charge [10], and the high-frequency output-intercept point $\mathrm{OIP}_3(2\omega_1 - \omega_2)$ thus peaks wherever $G_{3\mathrm{HF}}$ vanishes. To maximize $\mathrm{OIP}_3(2\omega_1 - \omega_2)$ when the transistor is operated at the peak of its $\hat{\omega}_T$ curve, the curve should be as flat as possible.

Appendix A
Expressions for the Taylor-Series Coefficients

From basic charge-control theory and the definition of ν_1, it follows that[2]

$$\nu_1 \equiv \frac{di_{\mathrm{CQS}}}{dq_T} = \hat{\omega}_T \tag{55}$$

[2]Each Taylor-series coefficient is to be evaluated at the transistor's operating point.

where $\hat{\omega}_T$ is the "loaded" unity–current–gain frequency. The conventional expression for $\hat{\omega}_T$ follows from (5) and (6):

$$\hat{\omega}_T^{-1} = \frac{C_{\mathrm{je}} + C_{\mathrm{de}} + C_{\mathrm{dc}} + C_{\mathrm{jc}}}{g_m} + (R_L + R_E)\, C_{\mathrm{jc}} \tag{56}$$

where the device capacitances are $C_{\mathrm{je}} \equiv dq_{\mathrm{JE}}/dv_{\mathrm{BE}}$, $C_{\mathrm{de}} \equiv dq_{\mathrm{DE}}/dv_{\mathrm{BE}}$, $C_{\mathrm{dc}} \equiv \partial q_C/\partial v_{\mathrm{BE}}$, and $C_{\mathrm{jc}} \equiv \partial q_C/\partial v_{\mathrm{BC}}$, and where the transconductance is $g_m \equiv di_{\mathrm{CQS}}/dv_{\mathrm{BE}}$. The values of ν_2 and ν_3 follow from (55) and their definitions:

$$\nu_2 \equiv \frac{1}{2}\frac{d^2 i_{\mathrm{CQS}}}{dq_T^2} = \frac{1}{2}\hat{\omega}_T \hat{\omega}_T' \tag{57}$$

$$\nu_3 \equiv \frac{1}{6}\frac{d^3 i_{\mathrm{CQS}}}{dq_T^3} = \frac{1}{6}\hat{\omega}_T(\hat{\omega}_T'^2 + \hat{\omega}_T \hat{\omega}_T'') \tag{58}$$

where $\hat{\omega}_T'$ and $\hat{\omega}_T''$ are the first and second derivatives, respectively, of $\hat{\omega}_T$ with respect to i_{CQS}, found under the circuit constraint of constant V_{CC}.

An expression for ς_1 in terms of $\hat{\omega}_T$ and g_m follows by writing

$$\varsigma_1 \equiv \frac{dv_{\mathrm{BE}}}{dq_T} = \frac{dv_{\mathrm{BE}}}{di_{\mathrm{CQS}}}\frac{di_{\mathrm{CQS}}}{dq_T} = \frac{1}{g_m}\hat{\omega}_T. \tag{59}$$

Similarly, one can find

$$\varsigma_2 \equiv \frac{1}{2}\frac{d^2 v_{\mathrm{BE}}}{dq_T^2} = \frac{1}{2}\left[\frac{1}{g_m}\hat{\omega}_T' - \frac{\dot{g}_m}{g_m^3}\hat{\omega}_T\right]\hat{\omega}_T \tag{60}$$

and

$$\varsigma_3 \equiv \frac{1}{6}\frac{d^3 v_{\mathrm{BE}}}{dq_T^3} = \frac{1}{6}\frac{1}{g_m}\hat{\omega}_T(\hat{\omega}_T'^2 + \hat{\omega}_T\hat{\omega}_T'') + \left(\frac{1}{2}\frac{\dot{g}_m^2}{g_m^5} - \frac{1}{6}\frac{\ddot{g}_m}{g_m^4}\right)\hat{\omega}_T^3 - \frac{1}{2}\frac{\dot{g}_m}{g_m^3}\hat{\omega}_T^2\hat{\omega}_T' \tag{61}$$

where $\dot{g}_m$ and $\ddot{g}_m$ are the first and second derivatives, respectively, of g_m with respect to v_{BE}.

The coefficient b_1 is the reciprocal of the common-emitter low-frequency current gain:

$$b_1 \equiv \frac{di_{\mathrm{BQS}}}{di_{\mathrm{CQS}}} = \beta_0^{-1}. \tag{62}$$

The coefficients b_2 and b_3 can then be written as follows:

$$b_2 \equiv \frac{1}{2}\frac{d^2 i_{\mathrm{BQS}}}{di_{\mathrm{CQS}}^2} = \frac{1}{2}\left[\frac{-\beta_0'}{\beta_0^2}\right] \tag{63}$$

$$b_3 \equiv \frac{1}{6}\frac{d^3 i_{\mathrm{BQS}}}{di_{\mathrm{CQS}}^3} = \frac{1}{6}\left[\frac{2\beta_0'}{\beta_0^3} - \frac{\beta_0''}{\beta_0^2}\right] \tag{64}$$

where β_0' and β_0'' are the first and second derivatives, respectively, of β_0 with respect to i_{CQS}.

Appendix B
Visualization of the Cancellation Mechanism

The cancellation or subtraction of terms in (21) can be viewed as a manifestation of voltage feedback enabled by the presence of the source resistance. This can be illustrated by considering the origin of the total output current at a mixing frequency of $2\omega_1$ in response to a single-tone input at ω_1. The equivalent circuit for the calculation, based on the "method of nonlinear currents" [28, pp. 190–207], is shown in Fig. 8(a); for simplicity, we have assumed $R_E = 0$.

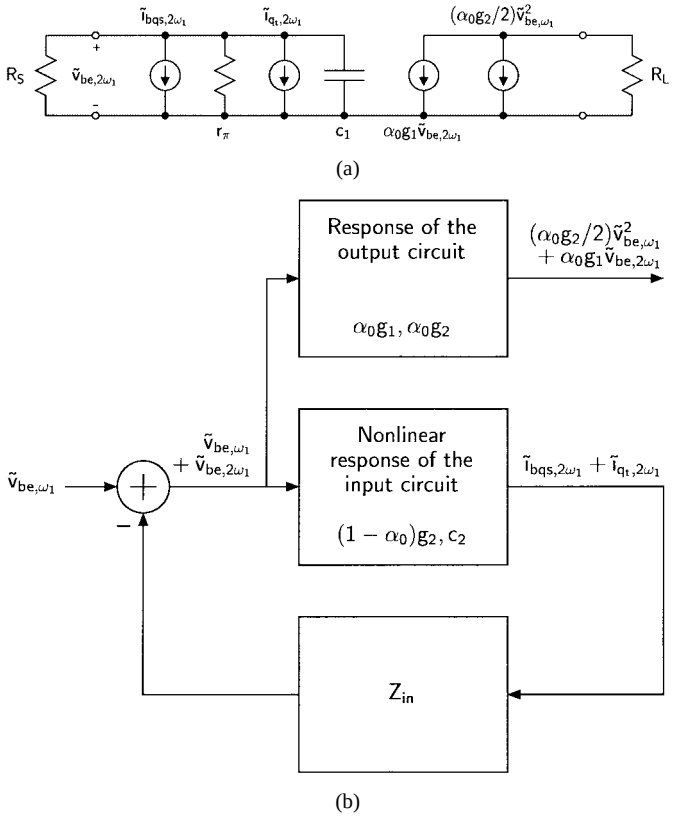


Fig. 8. (a) Equivalent circuit, based on the method of nonlinear currents, for computing the output distortion at a mixing frequency of $2\omega_1$ in response to a single-tone input at ω_1. (b) One manner of visualizing the feedback mechanism described in the text; the only signals shown are those involved in the second-order response at $2\omega_1$. The charge coefficients c_1 and c_2 are effective values, specified by the relations below (23).

The feedback mechanism, which is illustrated schematically in Fig. 8(b), can be understood as follows. The application of a source voltage causes a first-order base–emitter voltage at a frequency ω_1 to appear across the transistor's input terminals, with a phasor amplitude that we shall denote $\tilde{v}_{\mathrm{be},\omega_1}$. In response to $\tilde{v}_{\mathrm{be},\omega_1}$, the transistor's nonlinear transconductance causes an output current at $2\omega_1$, with the phasor amplitude $(\alpha_0 g_2/2)\tilde{v}^2_{\mathrm{be},\omega_1}$. However, the appearance of $\tilde{v}_{\mathrm{be},\omega_1}$ also causes nonlinear currents at $2\omega_1$ to flow into the input of the transistor; these are the quasi-static base current $\tilde{i}_{\mathrm{bqs},2\omega_1}$, and the charging current $\tilde{i}_{q_t,2\omega_1}$. The presence of the source resistance then causes a feedback voltage $\tilde{v}_{\mathrm{be},2\omega_1}$, at a frequency $2\omega_1$, which *opposes* the effects of $\tilde{v}_{\mathrm{be},\omega_1}$, to develop across the transistor's input terminals:

$$\tilde{v}_{\mathrm{be},2\omega_1} = -(\tilde{i}_{\mathrm{bqs},2\omega_1} + \tilde{i}_{q_t,2\omega_1})Z_{\mathrm{in}} \tag{65}$$

where Z_{in} is the equivalent impedance seen at the input, given by

$$Z_{\mathrm{in}} = \frac{r_\pi R_S}{r_\pi + R_S + j2\omega_1 r_\pi c_1 R_S} = \frac{\frac{1}{c_1}}{\omega_D + j2\omega_1} \tag{66}$$

with $r_\pi \equiv 1/g_1(1-\alpha_0)$; note that if $R_S = 0$, then $Z_{\mathrm{in}} = 0$, and hence $\tilde{v}_{\mathrm{be},2\omega_1} = 0$. The transistor's linear transconductance then acts on $\tilde{v}_{\mathrm{be},2\omega_1}$, creating an output current $\alpha_0 g_1 \tilde{v}_{\mathrm{be},2\omega_1}$ at $2\omega_1$ that *opposes* the original current $(\alpha_0 g_2/2)\tilde{v}^2_{\mathrm{be},\omega_1}$. The net distortion current at the output is thus

$$\begin{aligned}\tilde{i}_{\mathrm{cqs},2\omega_1} &= \left(\frac{\alpha_0 g_2}{2}\right)\tilde{v}^2_{\mathrm{be},\omega_1} + \alpha_0 g_1 \tilde{v}_{\mathrm{be},2\omega_1} \\ &= \alpha_0\left[\left(\frac{g_2}{2}\right)\tilde{v}^2_{\mathrm{be},\omega_1} - g_1(\tilde{i}_{\mathrm{bqs},2\omega_1} + \tilde{i}_{q_t,2\omega_1})Z_{\mathrm{in}}\right]\end{aligned} \tag{67}$$

where (65) has been used. It is *this* subtraction, in (67), that ultimately leads to the subtraction, or cancellation, identified in (21). This is readily verified by substituting (66), along with the values $\tilde{v}_{\mathrm{be},\omega_1} = (1/c_1)(s/R_S)/[\omega_D + j\omega_1]$, $\tilde{i}_{\mathrm{bqs},2\omega_1} = (1-\alpha_0)(g_2/2)\tilde{v}^2_{\mathrm{be},\omega_1}$, and $\tilde{i}_{q_t,2\omega_1} = j\omega_1 c_2 \tilde{v}^2_{\mathrm{be},\omega_1}$, all of which follow from the method of nonlinear currents, into (67), and then comparing the result to $(1/2)s^2 G_2(\omega_1,\omega_1)$, with $G_2(\omega_1,\omega_1)$ found from (20).

Appendix C
Value of θ

The value of θ appearing in (50) is

$$\begin{aligned}\theta = &\left(\frac{1}{12}\frac{\dot{g}_m^2}{g_m^6} - \frac{1}{36}\frac{\ddot{g}_m}{g_m^5}\right)\left[1 + g_m\left(\frac{R_S}{\beta_0} + \frac{R_E}{\alpha_0}\right)\right]\hat{\omega}_T^6\hat{\omega}_T'^2 \\ &+ \frac{\dot{g}_m}{g_m^7}\left\{\frac{1}{18}\ddot{g}_m - \frac{1}{12}\frac{\dot{g}_m^2}{g_m}\left[\frac{\frac{4}{3} + 2g_m\left(\frac{R_S}{\beta_0} + \frac{R_E}{\alpha_0}\right)}{1 + g_m\left(\frac{R_S}{\beta_0} + \frac{R_E}{\alpha_0}\right)}\right]\right\}\hat{\omega}_T^7\hat{\omega}_T' \\ &- \frac{1}{36}\frac{\dot{g}_m^2}{g_m^6}\hat{\omega}_T^7\hat{\omega}_T''.\end{aligned} \tag{68}$$

Appendix D
Third-Order Current Cancellation in the High-Frequency Limit

The high-frequency limit of the third-order kernel, given by (46), embodies a cancellation of intermodulation currents generated by the nonlinear transconductance and nonlinear stored charge. This can be shown by applying the method of nonlinear currents to compute the output distortion, at the intermodulation mixing frequency of $2\omega_1 - \omega_2$, that exists in the limit of infinite tone frequencies ($\omega \equiv \omega_1 \approx \omega_2 \to \infty$) and vanishing tone spacing ($\Delta\omega \equiv \omega_1 - \omega_2 \to 0$). To simplify the discussion, we will again assume $R_E = 0$, although the result holds even when $R_E \neq 0$.

As suggested in [10, p. 444], the circuit of Fig. 8 can be used for the calculation, provided that each of the current sources is replaced by an appropriate one at $2\omega_1 - \omega_2$. In the input circuit, the required nonlinear currents, which arise from both third-order nonlinearities and second-order interactions, can be written in terms of lower order mixing voltages, as follows:

$$\begin{aligned}&\tilde{i}_{\mathrm{bqs},2\omega_1-\omega_2} \\ &= (1-\alpha_0)\left[\frac{3}{4}g_3\tilde{v}^2_{\mathrm{be},\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + g_2\left(\tilde{v}_{\mathrm{be},2\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2}\right.\right. \\ &\qquad\left.\left. + \tilde{v}_{\mathrm{be},\omega_1}\tilde{v}_{\mathrm{be},\omega_1-\omega_2}\right)\right]\end{aligned} \tag{69}$$

$$\tilde{i}_{q_t,2\omega_1-\omega_2} = j(2\omega_1-\omega_2)\left[\frac{3}{4}c_3\tilde{v}^2_{\mathrm{be},\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + c_2\left(\tilde{v}_{\mathrm{be},2\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + \tilde{v}_{\mathrm{be},\omega_1}\tilde{v}_{\mathrm{be},\omega_1-\omega_2}\right)\right] \tag{70}$$

where "$*$" denotes complex conjugation. In the high-frequency limit, the quasi-static base current $\tilde{i}_{\mathrm{bqs},2\omega_1-\omega_2}$ is clearly negligible in comparison with the current $\tilde{i}_{q_t,2\omega_1-\omega_2}$ generated by the nonlinear stored charge; additionally, the impedance between the input terminals approaches $1/j(2\omega_1-\omega_2)c_1$. Therefore, the input voltage is determined solely by the transistor's nonlinear stored charge:

$$\lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} \tilde{v}_{\mathrm{be},2\omega_1-\omega_2} = \lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} \left[\frac{-1}{j(2\omega_1-\omega_2)c_1}\times\tilde{i}_{q_t,2\omega_1-\omega_2}\right]. \tag{71}$$

In the output circuit, the required currents are those due to the nonlinear transconductance acting on lower order mixing voltages, and the linear transconductance acting on $\tilde{v}_{\mathrm{be},2\omega_1-\omega_2}$. These currents are given by the following expressions:

$$\tilde{i}^{\mathrm{NL}}_{\mathrm{cqs},2\omega_1-\omega_2} = \frac{3}{4}\alpha_0 g_3\tilde{v}^2_{\mathrm{be},\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + \alpha_0 g_2 \times\left(\tilde{v}_{\mathrm{be},2\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + \tilde{v}_{\mathrm{be},\omega_1}\tilde{v}_{\mathrm{be},\omega_1-\omega_2}\right) \tag{72}$$

$$\tilde{i}^{L}_{\mathrm{cqs},2\omega_1-\omega_2} = \alpha_0 g_1\tilde{v}_{\mathrm{be},2\omega_1-\omega_2}. \tag{73}$$

The net distortion at the output is then

$$\begin{aligned}\lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} &\left|\tilde{i}_{\mathrm{cqs},2\omega_1-\omega_2}\right| \\ &= \lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} \left|\tilde{i}^{\mathrm{NL}}_{\mathrm{cqs},2\omega_1-\omega_2} + \tilde{i}^{L}_{\mathrm{cqs},2\omega_1-\omega_2}\right| \\ &= \lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} \left|\tilde{i}^{\mathrm{NL}}_{\mathrm{cqs},2\omega_1-\omega_2} - \frac{\alpha_0 g_1}{j(2\omega_1-\omega_2)c_1}\times\tilde{i}_{q_t,2\omega_1-\omega_2}\right|\end{aligned} \tag{74}$$

where (71) and (73) have been used; hence, the output distortion involves a cancellation, or subtraction, of intermodulation currents originating from the nonlinear transconductance and the nonlinear stored charge. Employing (70) and (72), this result can be expressed in the alternative form

$$\begin{aligned}\lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} &\left|\tilde{i}_{\mathrm{cqs},2\omega_1-\omega_2}\right| \\ &= \lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} \left|\left[\frac{\alpha_0 g_1}{c_1}\right]\left[\frac{3}{4}\left(\frac{g_3c_1}{g_1}-c_3\right)\tilde{v}^2_{\mathrm{be},\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + \left(\frac{g_2c_1}{g_1}-c_2\right)\left(\tilde{v}_{\mathrm{be},2\omega_1}\tilde{v}^*_{\mathrm{be},\omega_2} + \tilde{v}_{\mathrm{be},\omega_1}\tilde{v}_{\mathrm{be},\omega_1-\omega_2}\right)\right]\right|.\end{aligned} \tag{75}$$

Combining (44) and (46), it is also possible to write

$$\lim_{\substack{\omega\to\infty\\ \Delta\omega\to 0}} \left|\tilde{i}_{\mathrm{cqs},2\omega_1-\omega_2}\right| = \frac{3}{4}s^3 G_{3\mathrm{HF}}. \tag{76}$$

By equating (76) with (74) and (75), one then obtains (52) in Section IV-D.

ACKNOWLEDGMENT

The authors would like to thank K. Walter, IBM Microelectronics, Burlington, VT, for providing the SPICE parameters in Table I, and J. Scott, C. Hutchinson, D. Root, D. D'Avanzo, and D. Hornbuckle, all of Agilent Technologies, Santa Rosa, CA, for assistance with the distortion measurements.

REFERENCES

[1] S. Narayanan, "Transistor distortion analysis using Volterra series representation," *Bell Syst. Tech. J.*, pp. 991–1024, May-June 1967.

[2] ——, "Intermodulation distortion of cascaded transistors," *IEEE J. Solid-State Circuits*, vol. SC-4, pp. 97–106, June 1969.

[3] S. H. Chisholm and L. W. Nagel, "Efficient computer simulation of distortion in electronic circuits," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 742–745, Nov. 1973.

[4] Y. L. Kuo, "Distortion analysis of bipolar transistor circuits," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 709–716, Nov. 1973.

[5] H. C. Poon, "Modeling of bipolar transistor using integral charge-control model with application to third-order distortion studies," *IEEE Trans. Electron Devices*, vol. ED-19, pp. 719–731, June 1972.

[6] S. Narayanan and H. C. Poon, "An analysis of distortion in bipolar transistors using integral charge control model and Volterra series," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 341–351, July 1973.

[7] H. C. Poon, "Implication of transistor frequency dependence on intermodulation distortion," *IEEE Trans. Electron Devices*, vol. ED-21, pp. 110–112, Jan. 1974.

[8] H. E. Abraham and R. G. Meyer, "Transistor design for low distortion at high frequencies," *IEEE Trans. Electron Devices*, vol. ED-23, pp. 1290–1297, Dec. 1976.

[9] H. K. V. Lotsch, "Theory of nonlinear distortion produced in a semiconductor diode," *IEEE Trans. Electron Devices*, vol. ED-15, pp. 294–307, May 1968.

[10] S. A. Maas, B. L. Nelson, and D. L. Tait, "Intermodulation in heterojunction bipolar transistors," *IEEE Trans. Microwave Theory Tech.*, vol. 40, pp. 442–448, Mar. 1992.

[11] A. Samelis and D. Pavlidis, "Mechanisms determining third order intermodulation distortion in AlGaAs/GaAs heterojunction bipolar transistors," *IEEE Trans. Microwave Theory Tech.*, vol. 40, pp. 2374–2380, Dec. 1992.

[12] S. A. Maas, B. Nelson, and D. Tait, "Comments on 'Mechanisms determining third-order intermodulation distortion in AlGaAs/GaAs heterojunction bipolar transistors'," *IEEE Trans. Microwave Theory Tech.*, vol. 41, pp. 2038–2039, Nov. 1993.

[13] A. Samelis and D. Pavlidis, "Reply to 'Comments on mechanisms determining third order intermodulation distortion in AlGaAs/GaAs heterojunction bipolar transistors'," *IEEE Trans. Microwave Theory Tech.*, vol. 41, pp. 2039–2040, Nov. 1993.

[14] J. Lee, W. Kim, Y. Kim, T. Rho, and B. Kim, "Intermodulation mechanism and linearization of AlGaAs/GaAs HBT's," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 2065–2072, Dec. 1997.

[15] N. L. Wang, W. J. Ho, and J. A. Higgins, "AlGaAs/GaAs HBT linearity characteristics," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 1845–1850, Oct. 1994.

[16] M. Iwamoto, P. M. Asbeck, T. S. Low, C. P. Hutchinson, J. B. Scott, A. Cognata, X. Qin, L. H. Camnitz, and D. C. D'Avanzo, "Linearity characteristics of InGaP/GaAs HBT's and the influence of collector design," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 2377–2388, Dec. 2000.

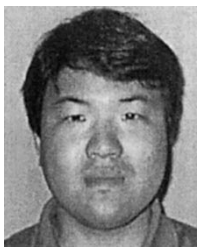
[17] K. W. Kobayashi, J. C. Cowles, L. T. Tran, A. Gutierrez-Aitken, M. Nishimoto, J. H. Elliott, T. R. Block, A. K. Oki, and D. C. Streit, "A 44-GHz-high IP3 InP HBT MMIC amplifier for low DC power millimeter-wave receiver applications," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1188–1195, Sept. 1999.

[18] L. C. N. de Vreede, H. C. de Graaf, J. A. Willemen, W. van Noort, R. Jos, L. E. Larson, J. W. Slotboom, and J. L. Tauritz, "Bipolar transistor epilayer design using the MAIDS mixed-level simulator," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1331–1338, Sept. 1999.

[19] M. Iwamoto, T. S. Low, C. P. Hutchinson, J. B. Scott, A. Cognata, X. Qin, L. H. Camnitz, P. M. Asbeck, and D. C. D'Avanzo, "Influence of collector design on InGaP/GaAs HBT linearity," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 2000, pp. 757–760.

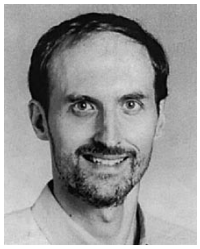
[20] M. Schröter, D. R. Pehlke, and T.-Y. Lee, "Compact modeling of high-frequency distortion in silicon integrated bipolar transistors," *IEEE Trans. Electron Devices*, vol. 47, pp. 1529–1535, July 2000.
[21] J. C. Pedro and J. Perez, "Accurate simulation of GaAs MESFET's intermodulation distortion using a new drain-source current model," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 25–33, Jan. 1994.
[22] J. C. Pedro, "Evaluation of MESFET nonlinear intermodulation distortion reduction by channel-doping control," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 1989–1997, Nov. 1997.
[23] J. A. Garcia, A. Mediavilla, J. C. Pedro, N. B. de Carvalho, A. T. Puente, and J. L. Garcia, "Characterizing the gate-to-source nonlinear capacitor role on GaAs FET IMD performance," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 2344–2355, Dec. 1998.
[24] M. Iwamoto, C. P. Hutchinson, J. B. Scott, T. S. Low, M. Vaidyanathan, P. M. Asbeck, and D. C. D'Avanzo, "Optimum bias conditions for linear broad-band InGaP/GaAs HBT power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2954–2962, Dec. 2002.
[25] K. L. Fong and R. G. Meyer, "High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages," *IEEE J. Solid-State Circuits*, vol. 33, pp. 548–555, Apr. 1998.
[26] V. Aparin and C. Persico, "Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 977–980.
[27] K. Lu, P. M. McIntosh, C. M. Snowden, and R. D. Pollard, "Low-frequency dispersion and its influence on the intermodulation performance of AlGaAs/GaAs HBTs," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1996, pp. 1373–1376.
[28] S. A. Maas, *Nonlinear Microwave Circuits*. Piscataway, NJ: IEEE Press, 1997.
[29] G. Massobrio, "Bipolar junction transistor (BJT)," in *Semiconductor Device Modeling with SPICE*, P. Antognetti and G. Massobrio, Eds. New York: McGraw-Hill, 1988, ch. 3.
[30] D. L. Harame, J. H. Comfort, J. D. Cressler, E. F. Crabbé, Y.-C. Sun, B. S. Meyerson, and T. Tice, "Si/SiGe epitaxial-base transistors-Part I: Materials, physics, and circuits," *IEEE Trans. Electron Devices*, vol. 42, pp. 455–468, Mar. 1995.
[31] ——, "Si/SiGe epitaxial-base transistors—Part II: Process integration and analog applications," *IEEE Trans. Electron Devices*, vol. 42, pp. 469–482, Mar. 1995.
[32] J. H. Mathews, *Numerical Methods for Computer Science, Engineering, and Mathematics*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[33] D. A. Teeter, M. Karaküçük, J. R. East, and G. I. Haddad, "Analysis of intermodulation distortion in GaAs/AlGaAs HBT's," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1992, pp. 263–266.

Mani Vaidyanathan (S'95–M'99) is currently an Assistant Research Scientist with the University of California at San Diego, La Jolla. His research interests are the theory and modeling of semiconductor devices, where he has been involved with topics ranging from describing carrier transport in small-dimension transistors to characterizing nonlinear behavior for RF applications.



Masaya Iwamoto (S'99) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1997, the M.S. degree in electrical engineering from the University of California at San Diego (UCSD), La Jolla, in 1999, and is currently working toward the Ph.D. degree at UCSD. His doctoral dissertation concerns the distortion characteristics of InGaP/GaAs HBTs and their applications to power amplifiers.

During the summers of 1997–2002, he was an Intern with Agilent Technologies (formerly the Hewlett-Packard Company), Santa Rosa, CA, where his responsibilities included HBT large-signal modeling, distortion characterization, and broad-band power-amplifier design.



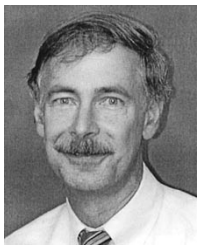
Lawrence E. Larson (S'82–M'86–SM'90–F'00) received the B.S. and M. Eng. degrees in electrical engineering from Cornell University, Ithaca, NY, in 1979 and 1980, respectively, the Ph.D. degree in electrical engineering and MBA degree from the University of California at Los Angeles (UCLA), in 1986 and 1996, respectively.

From 1980 to 1996, he was with Hughes Research Laboratories, Malibu, CA, where he directed the development of high-frequency microelectronics in GaAs, InP, and Si/SiGe and MEMS technologies. In 1996, he joined the faculty of the University of California at San Diego (UCSD), La Jolla, where he is currently the Inaugural Holder of the Communications Industry Chair. He is currently Director of the UCSD Center for Wireless Communications. During the 2000–2001 academic year, he was on leave with IBM Research, San Diego, CA, where he directed the development of RF integrated circuits (RFICs) for third-generation (3G) applications. He has authored or coauthored over 150 papers. He holds 24 U.S. patents.

Dr. Larson was the recipient of the 1995 Hughes Electronics Sector Patent Award for his work on RF MEMS technology. He was corecipient of the 1996 Lawrence A. Hyland Patent Award of Hughes Electronics for his work on low-noise millimeter-wave high electron-mobility transistors (HEMTs), and the 1999 IBM Microelectronics Excellence Award for his work in Si/SiGe HBT technology.

Prasad S. Gudem (M'96) received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 1996. His doctoral research concerned the modeling and numerical simulation of amorphous silicon devices.

After a brief time with Mitel Semiconductors, Kanata, ON, Canada, he joined Cadence Design Systems, San Diego, CA, where he was involved with semiconductor-device modeling for analog and RF applications. In August 2000, he joined the Watson Research Center, IBM, where he was involved with RF integrated-circuit design for WCDMA applications using IBM's SiGe technology. He was also appointed a Lecturer with the University of California at San Diego, La Jolla, where he taught a three-quarter course on communication circuit design. He recently joined Qualcomm Inc., San Diego, CA, where he continues to work on RF circuit design for wireless applications.



Peter M. Asbeck (M'75–SM'97–F'00) received the B.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1969 and 1975, respectively.

His professional experience includes working with the Sarnoff Research Center, Princeton, NJ, and Philips Laboratory, Briarcliff Manor, NY. In 1978, he joined the Rockwell International Science Center, Thousand Oaks, CA, where he was involved in the development of high-speed devices and circuits using III–V compounds and heterojunctions. He pioneered the effort to develop HBTs based on GaAlAs/GaAs and InAlAs/InGaAs materials. In 1991, he joined the University of California at San Diego, La Jolla, as a Professor in the Department of Electrical and Computer Engineering. His research has led to over 220 publications.

Dr. Asbeck is a Distinguished Lecturer of the IEEE Electron Devices Society and the IEEE Microwave Theory and Techniques Society (IEEE MTT-S).