Linearity and Dynamic Range of Carbon Nanotube Field-Effect Transistors

Stephen Maas
Nonlinear Technologies, Inc., Long Beach, CA 90807

Abstract — **We examine the problem of optimizing the linearity and dynamic range of a FET device, with application to carbon-nanotube (CNT) FETs. To develop insight into the way FET parameters affect linearity, we derive expressions for noise figure, intermodulation distortion, and dynamic range of a FET described by a unilateral equivalent circuit. This exercise identifies criteria for optimizing linearity and comparing the linearity of dissimilar devices. Measurements prove that CNT devices are significantly more linear than modern microwave FETs.**

Index Terms — **Intermodulation distortion, FET linearity, carbon nanotubes.**

I. Introduction

Carbon nanotube (CNT) FETs offer exciting possibilities for high-performance nanoscale circuits [1]. Models and measurements of those devices have indicated the possibility of highly linear operation [2, 3], which could significantly improve the dynamic range of many kinds of components and systems.

Perhaps surprisingly, in view of the maturity of nonlinear circuit theory, clear methods for defining and evaluating linearity have never been developed, and many ordinary criteria that purport to describe linearity are often inadequate. In particular, input and output intercept points alone are poor figures of merit for linearity, and popular ideas about optimizing linearity, based on them, are frequently incorrect. Instead, we view the optimization of dynamic range, within noise-figure constraints, as our criterion for linearity. Indeed, optimizing dynamic range is invariably the goal when linearity is considered in the design of a system or component.

To develop an intuitive understanding of linearity criteria, and to make valid comparisons between dissimilar devices, we calculate noise and distortion properties of a simple, unilateral FET equivalent circuit. This provides valuable insight that is not obtainable from more complex numerical approaches.

With this approach, we develop a deeper understanding of device characteristics that affect linearity, how to optimize linearity, and ways to optimize dynamic range. We also develop ways to compare the linearity of dissimilar devices, and we apply those methods to pHEMT and CNT devices.

II. Gain and Intermodulation Distortion

The analysis of a simplified FET equivalent circuit, shown in Fig. 1 provides considerable insight into the linearity and dynamic range of a CNT FET circuit. Reference [4] includes an analysis of the circuit in Fig. 1 under the assumption that the drain-to-source resistance, R_o, is infinite. This is clearly not the case in CNT FETs, which invariably have low R_o, caused by the presence of metallic tubes in the channel. When finite R_o is included, the expression for third-order IM, the output power PIM_3, of the $2\omega_2 - \omega_1$ tone, is

$$PIM_3 = 144 a_3^2 |H(\omega)|^6 R_s C_r R_L P_{av} \quad (1)$$

The linear transducer gain is

$$G_t = \frac{P_{LIN}}{P_{av}} = 4 a_1^2 |H(\omega)|^2 R_s C_r R_L \quad (2)$$

where P_{LIN} is the linear output power, P_{av} is the available input power, and R_s and R_L are the source and load resistances, respectively. The a_n coefficients are those of a Taylor-series expansion of the gate-to-drain I / V characteristic in the vicinity of the bias point:

$$f(v_g) = a_1 v_g + a_2 v_g^2 + a_3 v_g^3 \quad (3)$$

where v_g is the gate-voltage deviation around the bias point. The quantity C_r is the voltage-divider ratio at the output, $R_o / (R_L + R_o)$, and $H(\omega)$ is the transfer function between the source voltage, V_s, and the internal gate-source voltage, V_g; $H(\omega) = ((R_i + R_s) C_g j\omega)^{-1}$. We assume that the input is *tuned*, so $LC\omega^2 = 1$, but not necessarily *matched*, so R_i and R_s are not, in general, equal. When the output intercept points derived in [4] are modified for finite R_o, we obtain the third-order output intercept point, IP_{o3}, in watts, for the $2\omega_2 - \omega_1$ tone,

$$IP_{o3} = \frac{2}{3} \left| \frac{a_1^3}{a_3} \right| C_r^2 R_L \quad (4)$$

The third-order input intercept point, IP_{i3}, is simply the output intercept point divided by the transducer gain (2). Thus,

$$IP_{i3} = \frac{1}{6} \left| \frac{a_1}{a_3} \right| \frac{1}{R_s |H(\omega)|^2} \quad (5)$$

The absolute value sign is needed because the quantities a_2, a_3 are sometimes negative.

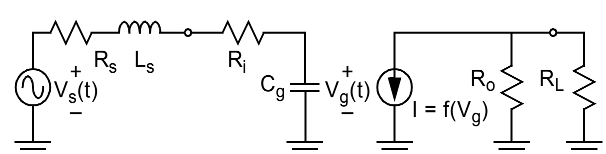


Fig. 1. Simplified pseudolinear small-signal equivalent circuit of a FET. The I / V characteristic, $f(v_g)$, is given by a Taylor-series polynomial in the vicinity of the gate bias point.

 978-1-5090-6360-4/17/$31.00 ©2017 IEEE

III. Noise

We assume the existence of two noise sources in Fig. 1: the thermal noise of the resistance in the input loop, R_i, and a current noise source in parallel with the channel, which comprises the channel noise and the thermal noise of R_o. This noise model is similar to that of Pospieszalski [6]. From [5], the equivalent input noise temperature, T_n, is

$$T_n = T_0\left(\frac{R_i}{R_s} + \frac{G_{nd}}{a_1^2|H(\omega)|^2 R_s}\right) \tag{6}$$

G_{nd} is the noise conductance of the channel noise source:

$$G_{nd} = \frac{\overline{|i|^2}}{4KT_0B} + \frac{1}{R_o} \tag{7}$$

In (7), $\overline{|i|^2}$ is the mean-square value of the channel noise current, K is Boltzmann's constant, $1.37\cdot 10^{-23}$ J / K, and T_0 is the standard temperature of 290K. Eq. (6) is derived under tuned but unmatched conditions.

From (6), achieving a low noise figure requires that $|H(\omega)|$ be maximized by minimizing R_i and C_g, but the situation involving R_s is more complex. At low frequencies the R_i / R_s term in (6) dominates, and the optimum value of R_s is relatively large. As frequency increases, however, the optimum value of R_s decreases, approaching R_i.

IV. Linearity Criteria

A. Intermodulation Intercept Points

A number of criteria for evaluating linearity are in regular use. The most common, IP_{o3}, is, by itself, a poor figure of merit for linearity. To illustrate why, consider two amplifiers having identical IP_{o3} but different gains. While both have the same IM output level, for a given linear output level, the input-referred intercept point of the high-gain amplifier is much lower, implying less dynamic range. Furthermore, a high value of IP_{o3} does not generally imply wide dynamic range, as any linear gain or loss following the device affects IP_{o3} but does not change the signal-to-IM ratio. This is particularly significant in a CNT device, as its low drain-to-source impedance affects IP_{o3} and gain identically, but not its inherent linearity (which we define later).

Another possible figure of merit is IP_{i3}. Input IP has the opposite problem: linear input loss increases IP_{i3}, but it also increases noise figure identically, so it does not affect dynamic range. Indeed, (5) indicates that IP_{i3} favors low $|H(\omega)|$, thus high input loss. High gain and low noise, however, demand high $|H(\omega)|$, so using IP_{i3} as a figure of merit would lead us to favor the worst devices.

B. Dynamic Range

In the design of many types of systems, we are concerned with dynamic range, not simply distortion. One common definition of dynamic range states that the minimum input signal level equals the noise in some bandwidth B; then, the maximum input signal is that which generates IM power equal to that noise level.

With this definition, the lower limit of the dynamic range, P_{min}, from (6), is

$$P_{min} = KBT_n = KBT_0\left(\frac{R_i}{R_s} + \frac{G_{nd}}{a_1^2|H(\omega)|^2 R_s}\right) \tag{8}$$

The maximum input level, P_{max}, occurs when the input-referenced third-order IM component equals P_{min}. From (8), (1), and (2), we obtain

$$P_{max} = \frac{1}{R_s}\left(KT_0B\frac{a_1^2}{a_3^2}\frac{R_s}{36|H(\omega)|^4}\left(\frac{R_i}{R_s} + \frac{G_{nd}}{a_1^2|H(\omega)|^2 R_s}\right)\right)^{1/3} \tag{9}$$

The ratio P_{max} / P_{min} is the dynamic range, DR:

$$DR = \left(\frac{1}{KT_0B}\left|\frac{a_1}{a_3}\right|\frac{1}{6|H(\omega)|^2}\frac{1}{\left(R_i + \frac{G_{nd}}{a_1^2|H(\omega)|^2}\right)}\right)^{2/3} \tag{10}$$

which can also be written

$$DR = \left(\frac{\left|\frac{a_1}{a_3}\right|\frac{1}{6|H(\omega)|^2 R_s}}{KBT_n}\right)^{2/3} \tag{11}$$

C. Discussion

The effect of $|H(\omega)|$ in (9-11) seems counterintuitive. In the development of field-effect transistors, every effort is made to minimize R_i and C_g, for a given device size, thus maximizing $|H(\omega)|$, and we would expect a large $|H(\omega)|$ to affect all aspects of performance positively. Eqs. (9-11), however, show that this is not the case, and the matter is further complicated by R_s, whose effect is more complex. We noted earlier that R_s must be optimized, not minimized, to achieve minimum noise figure; this is consistent with the behavior of real devices. Moreover, the input intercept points increase as $|H(\omega)|$ decreases, as input loss reduces the voltage across the gate-to-source capacitor. Thus, it should be expected that input intercept points, when used as figures of merit, might seem to improve as the FET's input resistance and capacitance increase.

The $|H(\omega)|$ dilemma directly affects the optimization of dynamic range. Invariably we do not want to maximize dynamic range blindly, but instead under a noise-figure constraint. That constraint usually states that the noise figure must be minimized or at least must remain below some particular value. Since $|H(\omega)|$ is constrained by noise-figure requirements, it cannot be freely adjusted to optimize dynamic range.

This leaves us with the quantity a_1 / a_3 as our only degree of freedom to optimize linearity, so we view this as the primary tool to compare the linearity of various devices. For this reason, the next section focusses on measurements of the a_n of the pHEMT and CNT.

V. Optimization of Dynamic Range

Fig. 2 shows the dynamic range at optimum noise figure vs. frequency and scaling. It is noteworthy that, under these conditions, scaling has little effect on dynamic range, a result that conflicts with common expectations. This occurs because the minimum noise temperature increases with S at close to the same rate as the distortion (numerator) term in (11). The output intercept point IP_{3o} increases in proportion to S, as expected from (4), as long as R_L similarly scales with S, and for this reason it is commonly assumed that

978-1-5090-6360-4/17/$31.00 ©2017 IEEE

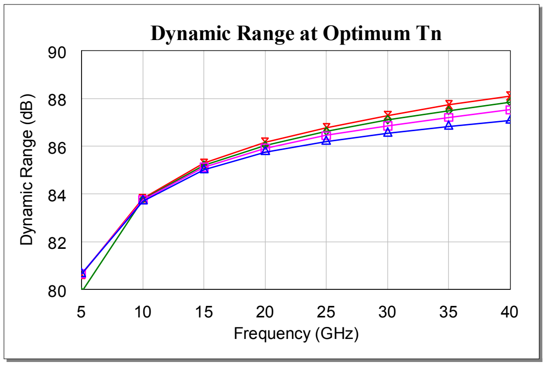


Fig. 2. Dynamic range vs. frequency with R_s optimized for minimum noise figure at each frequency point, for scaling factors of 0.5(Δ), 1.0(❑), 1.5(o) and 2.0(x); also, $a_1 = 0.20$, $a_3 = 0.10$, and $B = 1$ MHz.

dynamic range increases in proportion to S. These results bring that idea into question.

Fig. 3 shows the dynamic range at 12 GHz as a function of R_s. The dynamic range increases monotonically with R_s, although slowly at high R_s values, because $|H(\omega)|$ decreases. However, it also decreases the gain and increases the noise figure, again illustrating our earlier point that dynamic range cannot be optimized blindly. Even so, it does contradict the common idea that dynamic range is optimized by conjugate matching the input and output of the device. From Fig. 3, we see that a conjugate input match ($R_s = R_i = 4\Omega$ when $S = 1.0$) gives poor dynamic range, as it simultaneously increases the noise figure and $|H(\omega)|$. Although conjugate-matching the output optimizes IP_{o3}, from (11) we see that the dynamic range, as defined here, is not affected by output matching. This result underscores, again, the fact that intermodulation intercept points are poor figures of merit for linearity and dynamic range.

VI. CNT FETs

The CNT FETs have an interdigitated source-drain configuration. The gate has 12 fingers that are 360-nm long and 200 µm wide; the total device periphery is 2,400 µm. The gate fingers have a rectangular cross-sectional profile. The gate is deposited on top of high-K Ta_2O_5 dielectric.

The gate-to-source spacing is 0.1 µm and the source-to-drain spacing is 0.7 µm. The source and drain fingers are connected in parallel by their respective bus metals; the gate bus connects all gate fingers and overlaps the source bus. The gate bus crosses over the source bus on a 3-µm thick BCB interlayer dielectric.

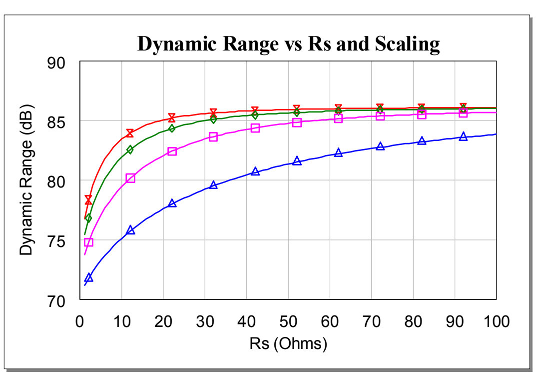


Fig. 3. Dynamic range vs. R_s at 12 GHz with the same scalings. The device parameters and graph markers are those of Figure 2.

VII. Device Linearity

In many cases we wish to compare the linearity of dissimilar devices. From (11), we have four terms that determine dynamic range: a_1 / a_3, $|H(\omega)|$, R_s, and T_n (or F). The latter three are constrained by the need to achieve an acceptable noise figure, as is gate-width scaling. This leaves us with a_1 / a_3 as the most important basis for comparing linearity; it is especially useful, as it is an inherent property of a device, largely independent of device width and frequency.

In all cases, devices can be compared on the basis of a noise quantity and a linearity term. T_n obviously should be that noise quantity, and L, where

$$L = \left|\frac{a_1}{a_3}\right| \frac{1}{6|H(\omega)|^2 R_s} \qquad (12)$$

is the linearity term.

VIII. Measurements of A_n Terms

We use an earlier method to measure values of a_n [7], appropriately updated for devices, such as carbon nanotube FETs, having low drain-to-source resistance. The method consists of exciting the device, in a test fixture, at a frequency low enough that the reactive parasitics are negligible. Harmonics of that excitation are measured at the device output, and the a_n terms are found from a Volterra analysis.

IX. Device Comparison

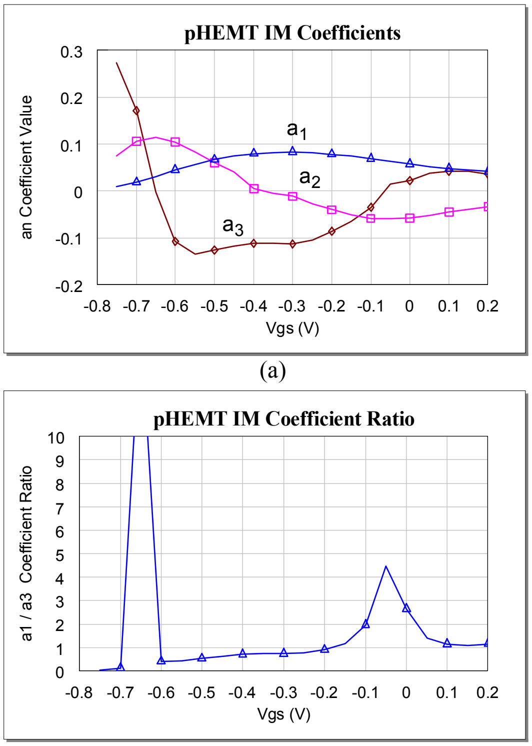
In this section we compare the a_n measurements of a modern, conventional pHEMT, the Renesas NE3512, and our CNT device.

Fig. 4 (a) shows the a_n values of the HEMT device at a drain bias voltage of 2.5V, and Fig. 4 (b) shows a_1 / a_3. The behavior is typical of modern FET devices. The pHEMT's a_1 / a_3 peaks at -0.65V or -0.07V exist only in ideally unilateral circuits; in a real amplifier, IM3 components are generated by second-order mixing between first- and second-order voltage components at the gate. This is dominated by a_2, and since $a_3 \sim 0$ is the a_2 maximum, those bias points do not result in low distortion. Between those points lies the practical gate bias region, where a_1 / a_3 is a little less than 1.0.

Fig. 5 shows the measured a_n coefficients of the CNT device and the a_1 / a_3 ratio. These measurements are somewhat less accurate than the pHEMT measurements, as the device is subject to slow drift, on the order of minutes, and fine fluctuations of the gate I / V characteristic, on the order of seconds. These phenomena are responsible for the variations in the a_n characteristics above approximately 0.2V.

The behavior of the coefficients for the CNT FET is markedly different from those of the pHEMT device. At gate-bias voltages below zero, the curves are relatively flat and the values a_2 and a_3 are quite small. The a_3 coefficient, in particular, is positive over most of this region, so amplifiers using CNT devices should exhibit modest gain enhancement before saturation. Above zero, the transconductance rolls off and the a_3 coefficient increases significantly. Even in that region, however, the CNT's a_1 / a_3 ratio is greater

 978-1-5090-6360-4/17/$31.00 ©2017 IEEE



(b)

Fig. 4. (a) Measured a_n coefficients of an NE3512 FET as a function of gate-bias voltage. (b) The ratio a_1 / a_3. The peaks are largely fictitious, as they occur at gate voltages near the points where a_3 = 0 (see text).

than the pHEMT's, showing more than an order of magnitude superiority.

From (12), we must consider $|H(\omega)|$ as well as a_1 / a_3. The parameters of interest are shown in Table 1., including the

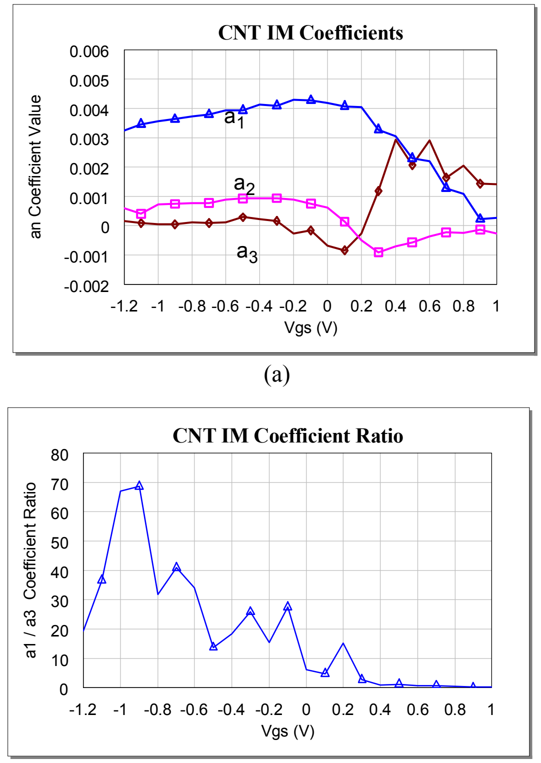
Table 1: FET parameter Comparison

Parameter	NE3512	CNT FET
R_g (Ω)	1.9	1.4
R_d (Ω)	0.8	6.1
R_s (Ω)	1.0	0.47
C_{gs} (pF)	0.28	1.95
$\|H(\omega)\|$ @ 1 GHz	10.9	1.57
L (R_s = 50Ω, f = 1 GHz	$2.6 \cdot 10^{-5}$	0.046

value of $|H(\omega)|$ with a 50-ohm source. Because of its greater gate-to-source capacitance, the CNT device's $|H(\omega)|$ is approximately a factor of six smaller than that of the pHEMT, making the linearity term (12) ~10^3 that of the pHEMT. It is difficult to estimate the improvement in dynamic range, as noise figure of the CNT device was not measured. However, from (11), the CNT noise temperature would have to be approximately 30 dB greater than that of the pHEMT to have the same dynamic range

X. Conclusions

Linearity measurements show that CNT devices have extraordinary inherent linearity, significantly greater than



(b)

Fig. 5. (a) Measured a_n coefficients of a CNT FET as a function of gate-bias voltage; V_{ds} = 3.3V. (b) The ratio a_1 / a_3. Values above 0V are affected by drift and small-scale I / V instability.

that of modern microwave FETs. This result confirms earlier predictions [2, 3].

References

[1] J. Appenzeller, "Carbon Nanotubes for High-Performance Electronics—Progress and Prospects," *Proc. IEEE*, vol. 96, no. 2, p. 201, 2008.

[2] C. Maneux et al., "Analysis of CNTFET Physical Compact Model," *DTIS 2006*, p. 40, 2006.

[3] J. E. Baumgardner et al., "Inherent Linearity in Carbon Nanotube Field Effect Transistors," *Applied Phys. Ltrs.*, vol. 91, no.5, 2007.

[4] S. Maas, *Nonlinear Microwave and RF Circuits,* Artech House, Norwood, MA, 2003.

[5] S. Maas, *Noise in Linear and Nonlinear RF and Microwave Circuits*, Artech House, Norwood, MA, 2005.

[6] M. W. Pospieszalski, "Modeling of Noise Parameters of MESFETs and MODFETs and Their Frequency and Temperature Dependence," *IEEE Trans. Microwave Theory Tech*., vol. 37, p. 1340, 1989.

[7] S. A. Maas, D. Neilson, "Modeling GaAs MESFETs for Intermodulation Analysis," *Microwave J*., vol. 34, no. 5, p. 295, 1991.

978-1-5090-6360-4/17/$31.00 ©2017 IEEE