RF Linearity of Short-Channel MOSFETs

Theerachet Soorapanth and Thomas H. Lee
Center for Integrated Systems, Stanford University
Stanford, CA 94305

Contact: Theerachet Soorapanth, Center for Integrated Systems, Stanford University, CA 94305
Phone: (415) 497-6481 Fax: (415) 725-6278 E-mail: chet@leland.stanford.edu
Area: RF circuits

Abstract— **This paper presents analytical expressions for the 1dB compression and third-order intercept points as a function of DC-bias and technology parameters. These two figures of merit measure the linearity of a system. In particular, the linearity of short-channel MOSFETs is examined.**

I. Introduction

RADIO frequency designs in CMOS are increasingly being explored to take advantage of rapid technology advances. As CMOS technology scales into the sub-micron regime, issues such as excess thermal noise due to hot electron effects have been well documented [2]. However, the equally important issue of linearity is not well studied. The goal of this work is to characterize quantitatively the linearity of sub-micron CMOS devices using widely-known metrics such as 1dB compression and 3rd-order-intermodulation intercept points.

To provide some background, Section II presents general expressions of the linearity metrics of a single-stage system as a function of DC parameters. Section III generalizes the expressions to multiple-stage systems. Since noise performance almost always trades off with linearity, Section IV considers both effects and derives the system gain distribution which yields the maximum linearity-to-noise ratio. In Section V, the linearity measures of short-channel MOSFETs are derived. Section VI discusses the condition for optimal noise and linearity performance in MOSFETs.

II. Background

For a general two-port system, one can write a Taylor series expansion relating input and output variables. Let V and I be the input and output variables, respectively. The value of I at some small signal v_{sig} around an input DC bias V_{DC} can be computed as

$$I(V_{\text{DC}} + v_{\text{sig}}) = c_0 + c_1 v_{\text{sig}} + c_2 v_{\text{sig}}^2 + \cdots + c_n v_{\text{sig}}^n \qquad (1)$$

where

$$c_n \triangleq \frac{I^{(n)}(V_{\text{DC}})}{n!}. \qquad (2)$$

The second and higher-order terms account for system nonlinearity. By using Taylor series the nonlinearity is assumed to be memoryless. Given this assumption, for a narrowband input signal $v_{\text{sig}} = A\cos(\omega_0 t)$, the corresponding narrowband output signal i_{sig} can be determined by summing all $\cos(\omega_0 t)$ terms, yielding

$$i_{\text{sig}} = \left[c_1 A + \frac{3}{4}c_3 A^3 + \cdots + \frac{c_{2k-1}}{2^{2k-2}}\binom{2k-1}{k-1}A^{2k-1}\right]\cos(\omega_0 t). \qquad (3)$$

The 1dB compression point can then be computed by taking the ratio of all terms in A to its linear term in (3) and setting the ratio equal to 0.891 (-1dB). This results in

$$0.109 c_1 A + \sum_{k=2}^{\infty}\left[\frac{c_{2k-1}}{2^{2k-2}}\binom{2k-1}{k-1}A^{2k-1}\right] = 0. \qquad (4)$$

Solving (4) for A, the input-referred 1dB compression point under matched conditions is related through

$$P_{\text{1dB}} = \frac{A^2}{2R_s}. \qquad (5)$$

In general, real systems are either of low-order (< 5) or the coefficients of higher-order terms are small relative to lower-order ones. In such systems, the fifth and higher order terms in (4) can be neglected and the remaining terms yield a simplified solution for A whose corresponding 1dB compression point becomes

$$P_{\text{1dB}} \approx \left|\frac{c_1}{13.8 c_3 R_s}\right| = \left|\frac{I^{(1)}(V_{\text{DC}})}{2.3 I^{(3)}(V_{\text{DC}})R_s}\right|. \qquad (6)$$

An input-referred third-order intercept point (IIP3) can be determined by substituting $v_{\text{sig}} = A\cos(\omega_1 t) + A\cos(\omega_2 t)$ into (1), and the output signal component at frequency $2\omega_1 - \omega_2$ or $2\omega_2 - \omega_1$ determines the third-order intermodulation product, i.e.,

$$i_{\text{IM3}} = \left|\frac{3}{4}c_3 A^3\right|\cos((2\omega_1 - \omega_2)t). \qquad (7)$$

Equating the amplitude of the intermodulation product term with that of $\cos(\omega_1 t)$ term yields IIP3, i.e.,

$$P_{\text{IIP3}} = \left|\frac{2c_1}{3c_3 R_s}\right| = \left|\frac{4I^{(1)}(V_{\text{DC}})}{I^{(3)}(V_{\text{DC}})R_s}\right|. \qquad (8)$$

Taking the ratio of P_{IIP3} to P_{1dB} yields a constant 9.3 (9.64dB) given these assumptions.

III. System with multiple stages

For a system with cascaded stages, the overall system linearity can be approximated from the linearity of individual stages. (To obtain the exact linearity figures, one needs to consider the entire multiple-stage system as a single nonlinear stage and apply the results from the previous section). To estimate IIP3, two boundary values can be obtained based on assumptions of how the IM3 products of each stage combine. An rms bound is obtained when the IM3 product of individual stages are assumed to be uncorrelated, i.e., their powers add to yield the overall rms IM3. On the other hand, a minimum bound is obtained when the IM3 products are fully correlated and thus their rms values add to yield the maximum IM3. The overall rms and maximum IM3 at the final stage can therefore be written as

$$\mathrm{IM3_{rms}} = P_{\mathrm{in}}^3 G_T \left(\frac{1}{\mathrm{IIP3_1}^2} + \sum_{i=2}^{N} \frac{\prod_{j=1}^{i-1} {G_j}^2}{\mathrm{IIP3}_i{}^2} \right) \quad (9)$$

$$\mathrm{IM3_{max}} = P_{\mathrm{in}}^3 G_T \left(\frac{1}{\mathrm{IIP3_1}} + \sum_{i=2}^{N} \frac{\prod_{j=1}^{i-1} G_j}{\mathrm{IIP3}_i} \right)^2 . \quad (10)$$

Here, G_T is the overall available power gain, while G_i and $\mathrm{IIP3}_i$ are the available power gain and IIP3 of stage i, respectively. P_{in} is the available source power, and N denotes the number of stages. From the IM3 values, the corresponding estimates for the system IIP3 are

$$\mathrm{IIP3_{rms}} = \left(\frac{1}{\mathrm{IIP3_1}^2} + \sum_{i=2}^{N} \frac{\prod_{j=1}^{i-1} {G_j}^2}{\mathrm{IIP3}_i{}^2} \right)^{-\frac{1}{2}} \quad (11)$$

$$\mathrm{IIP3_{min}} = \left(\frac{1}{\mathrm{IIP3_1}} + \sum_{i=2}^{N} \frac{\prod_{j=1}^{i-1} G_j}{\mathrm{IIP3}_i} \right)^{-1} . \quad (12)$$

According to (11) and (12), the overall IIP3 will be maximized by maximizing the IIP3 of individual stages (especially of the final stages) since the term associated with the IIP3 of stage i is multiplied by the total gain of the preceding $i-1$ stages. In addition, individual gains should be minimized, especially the initial gain stages since they appear in all subsequent terms. This requirement contradicts with the minimum noise condition which requires that the initial stages have the largest gain.

IV. Optimization of Linearity and Noise Performance

To quantify the tradeoff between noise and linearity performance, let us define a figure of merit called IP3-to-noise-figure ratio (IP3NR) as

$$\mathrm{IP3NR} \triangleq \frac{\mathrm{IIP3_{min}}}{F} \quad (13)$$

with overall noise figure defined as

$$F = 1 + \frac{1}{n_o} \sum_{i=1}^{N} \frac{n_i}{\prod_{j=1}^{i} G_j} . \quad (14)$$

Here, n_o represents the noise power due to source impedance at the input stage, n_i is the output noise power of stage i, and $\mathrm{IIP3_{min}}$ is defined in (12). The optimal gain distribution maximizing IP3NR is

$$G_1 = \sqrt{\frac{n_1 G_T \mathrm{IIP3_2}}{(n_0 G_T + n_N)\mathrm{IIP3_1}}} \quad (15)$$

$$G_i = \sqrt{\frac{n_i \mathrm{IIP3}_{i+1}}{n_{i-1}\mathrm{IIP3}_i}} \quad ; i = 2 \ldots N-1 \quad (16)$$

$$G_N = \sqrt{\frac{G_T (n_0 G_T + n_N)\mathrm{IIP3_1}}{n_{N-1}\mathrm{IIP3}_N}} . \quad (17)$$

Substituting (15)-(17) into (12) and (14), the resulting optimal IIP3 and F are

$$\mathrm{IIP3_{opt}} = \left(\frac{1}{\mathrm{IIP3_1}} + \sqrt{\frac{G_T}{(n_0 G_T + n_N)\mathrm{IIP3_1}}} \sum_{i=1}^{N-1} \sqrt{\frac{n_i}{\mathrm{IIP3}_{i+1}}} \right)^{-1} \quad (18)$$

$$F_{\mathrm{opt}} = 1 + \frac{n_N}{G_T n_o} + \frac{1}{n_o} \sqrt{\frac{(n_0 G_T + n_N)\mathrm{IIP3_1}}{G_T}} \sum_{i=1}^{N-1} \sqrt{\frac{n_i}{\mathrm{IIP3}_{i+1}}} . \quad (19)$$

The above equations can be simplified in a special case where all n_i and $\mathrm{IIP3}_i$ are assumed to be identical. With that assumption, the gain distribution, the optimal IIP3 and F simplify to:

$$G_1 = \sqrt{\frac{G_T}{\alpha G_T + 1}} \quad ; \alpha = n_0 / n_1 \quad (20)$$

$$G_i = 1 \quad ; i = 2 \ldots N-1 \quad (21)$$

$$G_N = \sqrt{G_T (\alpha G_T + 1)} \quad (22)$$

$$\mathrm{IIP3_{opt}} = \frac{\mathrm{IIP3_1}}{1 + (N-1)\sqrt{\frac{G_T}{\alpha G_T + 1}}} \quad (23)$$

$$F_{\mathrm{opt}} = 1 + \frac{1}{\alpha G_T} + \frac{N-1}{\alpha} \sqrt{\frac{\alpha G_T + 1}{G_T}} . \quad (24)$$

In such a simplified case, the resulting expressions suggest that only two gain stages are sufficient to obtain the maximum IP3NR. This makes intuitive sense since the input stage is mainly responsible for noise performance, while the output stage provides the necessary gain. That is, if the first gain stage generates more noise than the noise due to source impedance (i.e. $\alpha < 1$), G_1 will have to be scaled up by approximately a factor of $1/\sqrt{\alpha}$, assuming large G_T.

V. Short-channel MOSFETs

Consider the I-V characteristics of short-channel MOSFETs [1] whose drain current and gate overdrive voltage are related through

$$I_{\mathrm{D,SAT}} = W v_{\mathrm{sat}} C_{\mathrm{ox}} \frac{{V_{\mathrm{od}}}^2}{V_{\mathrm{od}} + E_{\mathrm{sat}} L} \quad (25)$$

with

$$E_{\text{sat}} = \frac{2v_{\text{sat}}}{\mu_{\text{eff}}} \quad (26)$$

$$\mu_{\text{eff}} = \frac{\mu_0}{1+\theta V_{\text{od}}} \quad (27)$$

$$V_{\text{od}} = V_{\text{gs}} - V_t. \quad (28)$$

Here, E_{sat} is the velocity saturation field strength and v_{sat} is the saturation velocity. From (25)-(28), the coefficients c_1 and c_3 can be determined and substituted in (6) and (8), yielding

$$P_{\text{1dB}} \approx 0.29 \frac{v_{\text{sat}} L}{\mu_1 R_s} V_{\text{od}} \left(1 + \frac{\mu_1 V_{\text{od}}}{4 v_{\text{sat}} L}\right)\left(1 + \frac{\mu_1 V_{\text{od}}}{2 v_{\text{sat}} L}\right)^2 \quad (29)$$

$$P_{\text{IIP3}} = \frac{8}{3} \frac{v_{\text{sat}} L}{\mu_1 R_s} V_{\text{od}} \left(1 + \frac{\mu_1 V_{\text{od}}}{4 v_{\text{sat}} L}\right)\left(1 + \frac{\mu_1 V_{\text{od}}}{2 v_{\text{sat}} L}\right)^2 \quad (30)$$

where

$$\mu_1 \triangleq \mu_0 + 2\theta v_{\text{sat}} L. \quad (31)$$

According to (29) and (30), linearity can be improved by increasing the gate overdrive voltage. However, increasing the gate overdrive will increase power consumption, which may be unacceptable in some applications. Another observation is that W and C_{ox} do not appear in the P_{1dB} and P_{IIP3} equations. This suggests that IIP3 and device transconductance g_m can be decoupled through W and C_{ox} factors. Note that the above derivation is based on the assumption of quasi-static nonlinearity which is valid when the operating frequency is well below the transition frequency ω_T of the device.

A more accurate, but less intuitive, expression for P_{1dB} can be obtained by directly solving the infinite series in (4) which, in this case, converges to

$$P_{\text{1dB}} = \frac{\left(1 + \frac{\mu_1 V_{\text{od}}}{2 v_{\text{sat}} L}\right)^4}{2R_s \left(\frac{\mu_1}{2 v_{\text{sat}} L}\right)^2 \left[V_{\text{od}}\left(1 + \frac{\mu_1 V_{\text{od}}}{4 v_{\text{sat}} L}\right) + \frac{6.88 v_{\text{sat}} L}{\mu_1 \left(1 + \frac{\mu_1 V_{\text{od}}}{2 v_{\text{sat}} L}\right)^2}\right]}. \quad (32)$$

The ratio of P_{IIP3} to P_{1dB} now becomes

$$9.17 \left[1 + 0.145 \frac{\mu_1 V_{\text{od}}}{v_{\text{sat}} L}\left(1 + \frac{\mu_1 V_{\text{od}}}{4 v_{\text{sat}} L}\right)\right], \quad (33)$$

which is no longer a constant but depends on the overdrive voltage. Figure 3 illustrates this dependency.

As an example, let us consider a 0.5μm (0.35μm L_{eff}) technology (t_{ox}=9.7nm), where $v_{\text{sat}}=2.4\times10^7$ cm/sec and μ_0=495 cm^2/Vsec (NMOS). Figures 1 and 2 illustrate the corresponding P_{1dB} and P_{IIP3} as a function of V_{od}. Also shown in the figures are results from Spice simulations with MOSIS HP 0.5μm CMOS models. The simulation results show a close match with the theoretical calculations.

An interesting result regarding the effect of technology scaling on P_{1dB} and P_{IIP3} is shown in Figures 4 and 5. According to Figure 4, as the technology scales down (assuming v_{sat} and μ_0 stay approximately the same), the 1dB compression point decreases for the case of overdrive voltage below 1.5 volts. For overdrive voltage above 2 volts, the 1dB compression point stays relatively constant independent of scaling. On the other hand, IIP3 stays relatively constant for effective gate length above 0.6 μm. As technology scales below 0.6 μm, IIP3 will start to increase for the case of overdrive voltage greater than 1.5 volts, while decreasing at first but eventually rising for low overdrive case ($< 1V$).

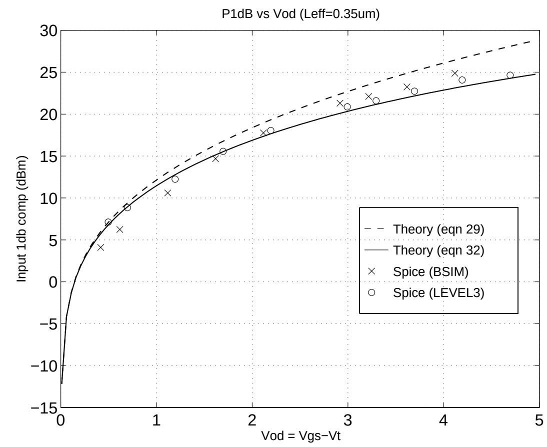

Fig. 1. 1dB compression point vs V_{od}

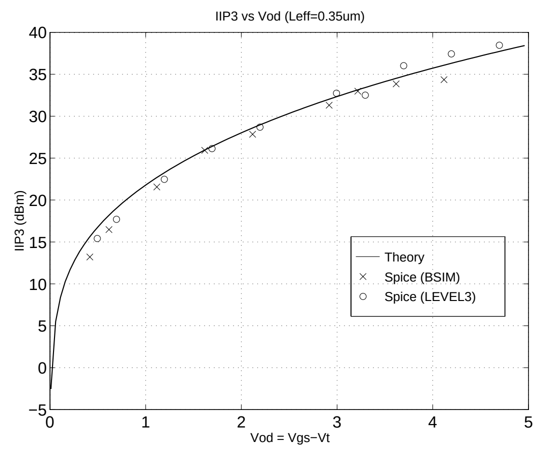

Fig. 2. IIP3 vs V_{od}

Figure 6 shows the corresponding P_{IIP3}-to-P_{1dB} ratio as gate length scales. One observation is that as the gate length reduces, the ratio deviates from the ~10dB constant value. This is because the higher-order terms in (4) become more significant and thus cannot be neglected.

VI. Noise Consideration

In [2], the expression for the minimum noise figure at RF for short-channel MOSFETs is given as

$$F_{\min} = 1 + \frac{2}{\sqrt{5}}\left(\frac{\omega}{\omega_T}\right)\sqrt{\delta\gamma(1-|c|^2)} \quad (34)$$

where δ and γ are the non-ideality factors in the gate and drain current noise power spectral density, respectively, and c is the correlation coefficient between the two noise

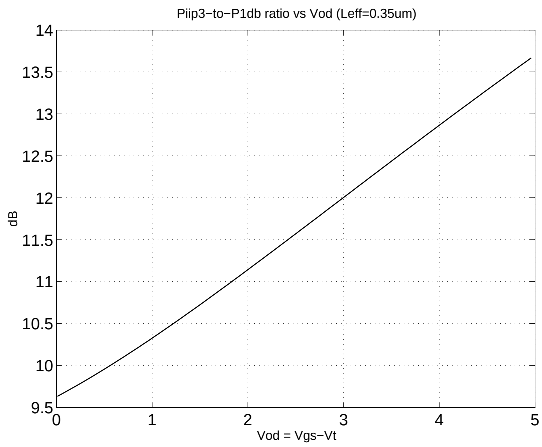

Fig. 3. IIP3-to-P1dB ratio vs V_{od}

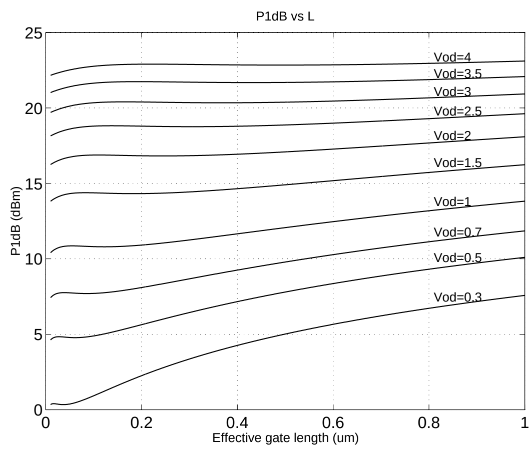

Fig. 4. P1dB vs L

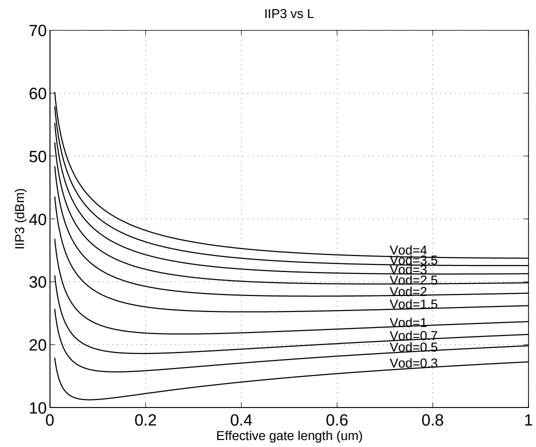

Fig. 5. IIP3 vs L

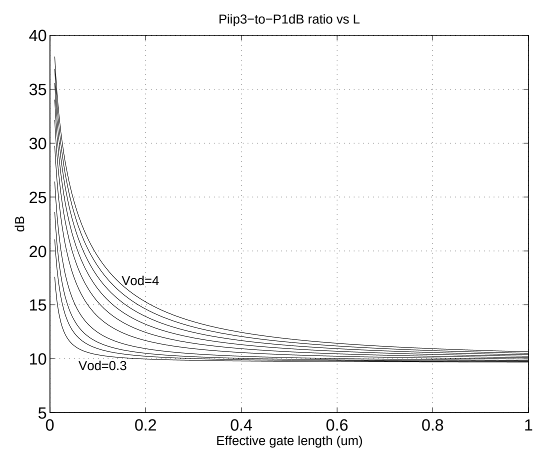

Fig. 6. IIP3-to-P1dB ratio vs L

sources. The noise performance is improved with increasing ω_T which, in turn, improves with increasing g_m whose value is given by

$$g_{m,\mathrm{sat}} = W v_{\mathrm{sat}} C_{\mathrm{ox}} \mu_0 V_{\mathrm{od}} \; \frac{\mu_1 V_{\mathrm{od}} + 4 v_{\mathrm{sat}} L}{(\mu_1 V_{\mathrm{od}} + 2 v_{\mathrm{sat}} L)^2}. \qquad (35)$$

Consider a special case in which $g_{m,\mathrm{sat}}$ can be simplified to

$$g_{m,\mathrm{sat}} \approx W v_{\mathrm{sat}} C_{\mathrm{ox}} \frac{\mu_0}{\mu_1} \qquad ; \text{for } V_{\mathrm{od}} \gg \frac{4 v_{\mathrm{sat}} L}{\mu_1}. \qquad (36)$$

According to (35) and (36), as V_{od} increases, $g_{m,\mathrm{sat}}$ will increase (linearly for small V_{od}) and eventually flatten out when the condition in (36) is satisfied. Therefore, only when $V_{\mathrm{od}} < 4 v_{\mathrm{sat}} L/\mu_1$ can ω_T, and thus F_{min}, be improved by increasing the overdrive voltage. To keep the power dissipation constant, the device's width has to be reduced. As we have seen earlier, changing the width does not affect IIP3, but will, at some point, affect the noise match condition causing a deviation from the given F_{min} until the benefit from ω_T increase is outpaced by the degradation due to the noise mismatch. Otherwise, overall dynamic range significantly improves with overdrive voltage over the range below the saturation point of ω_T. For 0.35μm L_{eff}, this range is $V_{\mathrm{od}} < 4.2V$.

VII. Conclusions

We have determined a direct relationship between linearity measures and DC bias parameters of a general two-port system as well as the condition for achieving optimal linearity and noise performance. We have derived the corresponding relation for short-channel MOS transistors. The results verify that CMOS technology is promising for RF designs.

Acknowledgments

The authors would like to thank Derek Shaeffer and Kevin Yu for many helpful and enlightening discussions.

References

[1] Ping K. Ko, *VLSI Electronics Microstructure Science*, vol. 18, Academic Press, Inc, 1989.

[2] Derek K. Shaeffer, Thomas H. Lee, "A 1.5V 1.5GHz CMOS Low Noise Amplifier," *IEEE Journal of Solid State Circuits,* vol. 32, no. 5, May 1997.

[3] I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, Inc, 1980.