Intermodulation Linearity in High-k/Metal Gate 28 nm RF CMOS Transistors

Zhen Li 1, Guofu Niu 1,*, Qingqing Liang 2 and Kimihiko Imura 2

1 Department of Electrical and Computer Engineering, Alabama Microelectronics Science and Technology Center, Auburn University, 200 Broun Hall, Auburn, AL 36849, USA
2 MaxLinear Inc., Carlsbad, CA 92008, USA

* Author to whom correspondence should be addressed; E-Mail: niuguof@auburn.edu; Tel.: +1-334-844-1856; Fax: +1-334-844-1888.

Academic Editor: Geok Ing Ng

Received: 3 July 2015 / Accepted: 1 September 2015 / Published: 11 September 2015

Abstract: This paper presents experimental characterization, simulation, and Volterra series based analysis of intermodulation linearity on a high-k/metal gate 28 nm RF CMOS technology. A figure-of-merit is proposed to account for both $V_{GS}$ and $V_{DS}$ nonlinearity, and extracted from frequency dependence of measured $IIP^3$. Implications to biasing current and voltage optimization for linearity are discussed.

Keywords: 28 nm; intermodulation; linearity; $IIP^3$; RF CMOS; high-k/metal gate

1. Introduction

Modern CMOS technology scaling is no longer just a matter of shrinking physical dimensions. A key to down scale the equivalent oxide thickness (EOT) in recent technologies is the replacement of classic poly-Si gate/$SiO_2$ gate stack with a high-k dielectric/metal gate stack. Given the tremendous interest in scaled RF CMOS and RF system-on-chip that integrates digital and RF functions, it is necessary to examine the RF performance of the core transistors in these scaled technologies.

In this work, we investigate two-tone intermodulation linearity in a 28 nm high-k/metal gate RF CMOS technology [1], characterized by the intermodulation intercept. Both second and third order intermodulation intercept $IIP^2$ and $IIP^3$ are measured. We focus on $IIP^3$ as it is more relevant. Third order intermodulation products are close to the fundamental frequencies of interest and cannot
be filtered out [2]. Mixing of adjacent channel interferers produces undesired output in the frequency band of interest. Third order nonlinearities are also responsible for desensitization and cross-modulation.

From a gate capacitance perspective, poly depletion effect is no longer present with the use of metal gate, the change of gate-to-source capacitance $C_{gs}$ with gate voltage is less in strong inversion, and linearity should improve compared to poly-gate transistors according to [3]. That analysis, however, assumed velocity saturation at the source, which is not the case in today’s advanced CMOS. Scaling, and the associated changes in doping, effective oxide thickness, strain are all expected to change device $I$–$V$ characteristics as well as the various transconductance nonlinearities, output conductance nonlinearities, and cross nonlinearities.

Harmonic gate voltage $IP3$ of 28 nm RF CMOS devices has been recently examined using third-order derivative of $I_{DS} - V_{GS}$ data [4]. However, no experimental RF measurement of $IP3$ has been reported. Previous investigations using Volterra series analysis [5] showed that such estimation using third-order transconductance nonlinearity alone is not sufficient in characterizing transistor $IP3$. Drain conductance nonlinearity as well as cross terms involving partial derivatives of $I_{DS}$ with respect to both $V_{GS}$ and $V_{DS}$ are also important [6]. Typical compact model parameters are extracted by fitting DC $I$-$V$ curves and sometimes first order derivatives. A good fitting does not necessarily guarantee good accuracy of higher order derivatives, which are difficult to evaluate experimentally due to the increase of numerical and experimental error in differentiation. Direct RF intermodulation measurements are therefore necessary, which we present below, together with simulations using a compact model with DC $I$-$V$ and Y-parameter calibration. As $IP3$ in RF measurements is determined using RF power of the source voltage, the result in general depends on frequency, and cannot be directly compared with traditional gate voltage $IP3$ that is defined using the gate voltage.

We propose below a new figure-of-merit that can be extracted from RF measurements so that meaningful comparison with traditional intermodulation gate voltage $IP3$ can be made with ease. The new figure-of-merit accounts for both $V_{GS}$ and $V_{DS}$ related nonlinearities, and reduces to traditional intermodulation gate voltage $IP3$ when all of the $V_{DS}$ related intermodulation products are neglected.

2. Tested Technology and Measurement System

Figure 1a shows typical $I_{DS} - V_{GS}$ characteristics of a 30 nm device from the examined 28 nm technology. Figure 1b shows measured cut-off frequency $f_T$ as a function of $I_{DS}$. A 304 GHz peak $f_T$ is reached at 0.45 mA/µm at $V_{DS} = 1.05$ V. Figure 1c shows typical $I_{DS} - V_{DS}$ characteristics.

Figure 2a shows the experimental setup used, which is similar to the setup in [7]. Broadband 50 Ω terminations are used so that they do not filter out the second order harmonics which may remix with the fundamental output to produce third order intermodulation ($IM3$). Devices are probed on-wafer using Cascade Infinity GSG probes. Two Agilent signal sources are synchronized and combined using a power combiner to produce a two tone input. Attenuators are used to reduce the intermodulation within the sources. Automatic level control in the sources is turned off to minimize intermodulation generated by the sources. An HP-6625 power supply is used to provide precision DC biases. A spectrum analyzer is used to measure the output spectrum. Power meters are used for calibration of power loss on cables and probes. Analyzer setting is optimized for each measurement to minimize analyzer $IM3$ and maximize
signal to noise ratio. For each bias point and frequency, the input power is swept and the third order intercept is obtained by extrapolation. The analyzer setting is optimized dynamically for each input power level. The measurement system intermodulation is verified to be well below the intermodulation from the device under test. The upper and lower $IM3$ are the same in our measurements.

Figure 1. Measured (a) $I_{DS}$ versus $V_{GS}$; (b) $f_T$ versus $I_{DS}$ and (c) $I_{DS}$ versus $V_{DS}$.

Figure 2. (a) Measurement setup and (b) extrapolation illustration for $IP3$ and $IP2$.

Figure 2b illustrates how $IP3$ and $IP2$ are determined for a 30 nm device biased at $V_{GS} = 0.44$ V, $V_{DS} = 0.6$ V. Device total width is 256 $\mu$m. Gate finger width $W_f$ is 1 $\mu$m, number of finger $N_f$ is 16, and multiplicity $M = 16$. At low $P_{in}$, first order output $P_{out,1st}$ increases linearly with $P_{in}$ at a slope of 1:1, while the third and the second order intermodulation output ($P_{out,3rd}$ and $P_{out,2nd}$) increase at slopes of 3:1 and 2:1, respectively. $IP3$ is obtained as the extrapolated intercept of $P_{out,1st}$ and $P_{out,3rd}$ in a region of $P_{in}$ where the ideal slopes are observed. The input and output powers at $IP3$ are denoted as $IIP3$ and $OIP3$. Their difference is gain. Similarly, we can obtain $IIP2$ and $OIP2$ from the extrapolation intercept of $P_{out,1st}$ and $P_{out,2nd}$. 

- $f_1$=5GHz
- $f_2$=5GHz
- $df$=1MHz
- $W_f$=4$\mu$m
- $L$=30nm
- $N_f$=16
- $M$=4
- $V_{GS}$=0.44V
- $V_{DS}$=0.6V
- $P_{out,1st}$
- $P_{out,2nd}$
- $P_{out,3rd}$
- $P_{in}$
- $P_{out}$
- $OIP2$
- $OIP3$
- $IIP2$
- $IIP3$
3. Results and Analysis

As mentioned earlier, in RF measurement, the intercept point is defined using RF input power. The input third order intermodulation intercept point, IIP3, is thus dependent on frequency, because of finite source impedance, which for our case, is a 50 Ω resistance. For a given RF input power, the RF gate voltage varies with frequency, as transistor input impedance varies with frequency. For analysis as well as estimation of IIP3 at another design frequency from measurement at one frequency, it is desirable to find a figure-of-merit that does not depend on frequency. Such figure-of-merit is more useful if it can relate to the traditional figure-of-merit, gate voltage VIP3, but also include effects of drain voltage related nonlinearities. We derive such a figure-of-merit below using Volterra series analysis.

A simplified equivalent circuit as shown in Figure 3 is used. Gate-drain capacitance (C_{gd}) is omitted, as the result is much simpler and sufficient for most purposes [5]. R_S = 50 Ω. C_{gs} is gate-to-source capacitance. C_d is drain capacitance. R_L = 50 Ω is load resistance.

\[ i_{ds} = g_m v_{gs} + K_{2gm} v_{gs}^2 + K_{3gm} v_{gs}^3 + g_0 v_{ds} + K_{2g0} v_{ds}^2 + K_{3g0} v_{ds}^3 + K_{2gmg0} v_{gs} v_{ds} + K_{3gm2g0} v_{gs}^2 v_{ds} \]

(1)

and \( K_{2gm}, K_{3gm}, K_{2g0}, K_{3g0}, K_{2gmg0}, K_{3gm2g0} \) are nonlinearity coefficients that relate to higher order partial derivatives as defined in [8] using Taylor expansion. For instance,

\[ g_m = \frac{\partial I_{DS}}{\partial V_{GS}} \]  \hspace{1cm} (2)

\[ K_{3gm} = \frac{1}{6} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3} \]  \hspace{1cm} (3)

Using the nonlinear current source method, IIP3 can be derived [5]:

\[ IIP3 = \frac{1 + \omega^2 C_{gs}^2 R_S^2}{6 R_S} \frac{1}{|K_{3gm} g_m + \Delta|} \]  \hspace{1cm} (4)

where \( \Delta = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 \). \( \Delta_1 \) through \( \Delta_4 \) are functions of nonlinear output conductance, its high order terms and cross terms with transconductance nonlinearity as follows:

\[ \Delta_1 = -\frac{1}{3} K_{2gmg0} g_m Z_1 - \frac{1}{3} K_{3gmg0} Z_2 \]  \hspace{1cm} (5)

\[ \Delta_2 = \frac{2}{3} K_{2gm} K_{2g0} Z_3 + \frac{1}{3} K_{3gmg0} g_m Z_4 + \frac{1}{3} K_{2gmg0}^2 Z_5 \]  \hspace{1cm} (6)

\[ \Delta_3 = -K_{3g0} g_m^2 Z_6 - \frac{1}{3} K_{2g0} K_{2gm} g_m Z_7 \]  \hspace{1cm} (7)

\[ \Delta_4 = \frac{2}{3} K_{2g0}^2 g_m^3 Z_8 \]  \hspace{1cm} (8)
Z₁ through Z₈ are given by:

\[ Z_1 = Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2) \]  
\[ Z_2 = Z_L(\omega_1) + [Y_S(-\omega_2)Y_S^{-1}(\omega_1) + 2] \]  
\[ Z_3 = 2Z_L(\omega_1 - \omega_2)Z_L(\omega_1) + Z_L(2\omega_1)Z_L(-\omega_2) \]  
\[ Z_4 = Z_L^2(\omega_1)[2Y_S(-\omega_2)Y_S^{-1}(\omega_1) + 1] \]  
\[ Z_5 = 2Z_L(\omega_1 - \omega_2)Z_L(-\omega_2) + Z_L(2\omega_1)Z_L(-\omega_1) \]  
\[ Z_6 = Z_L^2(\omega_1)Z_L(-\omega_2) \]  
\[ Z_7 = Z_L^2(\omega_1)[Z_L(2\omega_1) + 2Z_L(2\omega_1) + 6Z_L(\omega_1 - \omega_2)] \]  
\[ Z_8 = Z_L^2(\omega_1)Z_L(-\omega_2)[Z_L(2\omega_1) + 2Z_L(\omega_1 - \omega_2)] \]

with \( Z_L(\omega) = \frac{1}{\frac{1}{R_S}+g_0+j\omega C_d} \) and \( Y_S(\omega) = \frac{1}{R_S} + j\omega C_{gs} \).

A close inspection of the Volterra series based derivation details shows that at the intermodulation IP₃ point, the first order \( v_{gs} \) has an amplitude of:

\[ V_{GS,IP₃} = \sqrt{\frac{1}{\frac{3}{4}K_{3gm} + \Delta}} \]  

(17)

For typical transistor sizes of interest, the \( \Delta \) term is found to have a negligibly weak frequency dependence, making \( V_{GS,IP₃} \) nearly frequency independent in practice. We thus propose to use \( V_{GS,IP₃} \) as a figure-of-merit as it includes output conductance effect, and is more general than the traditional VIP₃ defined solely using \( K_{3gm} \) and \( g_m \). The designation \( GS \) in the subscript refers to the fact that this is the \( V_{GS} \) amplitude at the intercept. The value of \( V_{GS,IP₃} \), however, is clearly a function of the \( V_{DS} \) dependence of \( I_{DS} \), through the \( \Delta \) term.

Using \( V_{GS,IP₃} \), Equation (4) can then be rewritten as

\[ IIP₃ = \frac{C_{gs}^2R_SV_{GS,IP₃}^2}{8}\omega^2 + \frac{V_{GS,IP₃}^2}{8R_S} \]  

(18)

Equation (18) indicates that \( IIP₃ \) increases linearly with \( \omega^2 \) and \( V_{GS,IP₃} \) can be obtained experimentally by plotting measured \( IIP₃ \) as a function of \( \omega^2 \), as shown in Figure 4a. A linear fitting is made. The intercept with the \( IIP₃ \) axis gives \( V_{GS,IP₃}^2/8R_S \). Note that the unit used for \( IIP₃ \) is watt instead of dBm. As measured \( IIP₃ \) in dBm is shown in Figure 4b. The device has a drawn gate length of 30 nm. \( W_f = 4 \mu m \). \( N_f = 16 \). Multiplicity \( M = 4 \). The total width \( W_{total} = 256 \mu m \). \( V_{GS} = 0.7 \) V and \( V_{DS} = 1.0 \) V. Measurement frequency ranges from 100 MHz to 10 GHz. Within measurement uncertainty, the data shows an expected linear dependence on the square of fundamental angular frequency. This linear dependence of \( IIP₃ \) on \( \omega^2 \) is found to be valid for other bias points as well. The slope is given by \( \frac{C_{gs}^2R_SV_{GS,IP₃}^2}{8} \) from which \( C_{gs} \) can be extracted. The \( C_{gs} \) calculated is fairly close to that extracted from S-parameter measurements, thus supporting the validity of the proposed technique.

If we ignore the \( \Delta \) term that originates from the \( v_{ds} \) dependence of \( i_{ds} \), \( V_{GS,IP₃} \) reduces to

\[ VIP₃ = \sqrt{\frac{4}{3}\frac{g_m}{K_{3gm}}} \]  

(19)
This is essentially the $V_{GS,IP3}$ one would get if transistor drain current depends on $V_{GS}$ only. This $VIP3$ for intermodulation distortion differs from the third order harmonic distortion $VIP3$ in [4,9] by a constant.

The transistor model used to evaluate the derivatives needed in Equation (4) is a PSP model, with initial parameter values for base line digital CMOS transistors of the same technology. In this work, device model parameters are tuned to better fit the $I$-$V$ characteristics and S-parameters. Figure 5a,b compare simulated $I_{DS}$ versus $V_{GS}$ with measurement using linear and log $I_{DS}$ scales, respectively. Good agreement is achieved. To simulate $IP3$, quasi periodic steady state (QPSS) analysis is used in Cadence SpectreRF to calculate two-tone large signal behavior [10]. For each bias point, a series of input power level is swept. The output is plotted using ipnVRI function to ensure the extrapolation point for $IP3$ is within the linear range, in the same manner $IP3$ is determined in measurement illustrated earlier in Figure 2b.

Figure 6a shows both measured and simulated $IP3$ at 5 GHz as a function of $V_{GS}$ at $V_{DS} = 0.6$ V for the same device in Figure 5. Measurements and simulations are also made at 2 and 10 GHz. At each $V_{GS}$, from frequency dependence of $IP3$, a $V_{GS,IP3}$ is extracted. From 0.5 to 0.7 V, simulated $IP3$ is higher than measured $IP3$ by as much as 3.8 dB. This indicates that simulated $IP3$ for such technologies may be optimistic. In future work, model parameters can be further optimized to see if $IP3$ can be better fitted. To our knowledge, there are no direct knobs to turn to tune higher order derivatives in compact models. Improvement of $IP3$ simulation may require new improvements of the model formulation itself in addition to better parameter extraction and optimization. Figure 6b shows the $VIP3$ calculated from $K_{3gm}$ and $g_m$ using Equation (19). Fitting of $VIP3$, which is determined by the first and third order derivatives of $I_{DS}$-$V_{GS}$, is clearly worse than the fitting of $I_{DS}$-$V_{GS}$ itself shown earlier in Figure 5. Figure 6c,d show $V_{GS,IP3}$ and $K_{3gm}$ as a function of $V_{GS}$. The $K_{3gm} = 0$ point is clearly different from the measured $IP3$ and $V_{GS,IP3}$ peak positions. The peak $IP3$ $V_{GS}$ is 55 mV lower than the peak $VIP3$ $V_{GS}$. As was observed in 90 nm technology [5], $VIP3$ does not correctly predict the linearity sweet spot, due to omission of the $\Delta$ term. Around $V_{GS} = 0.6$ V, $V_{GS,IP3}$ and the traditional $VIP3$ are close to each other, as the $\Delta$ term is small. Beyond its peak, $IP3$ drops to a valley and starts rising
slowly. However, when $V_{GS} > 0.65$ V, as the device gets closer to linear operation region, $I_{IP3}$ shows a slight decrease.

**Figure 5.** Comparison of simulated $I_{DS}$ versus $V_{GS}$ with measurement on (a) linear and (b) log $I_{DS}$ scales at $V_{DS} = 0.6$ and 1.0 V.

Figure 6a–d show measured $I_{IP3}$, $I_{IP2}$, $V_{GS,IP3}$, and $V_{IP3}$ as a function of $V_{GS}$ at $V_{DS} = 0.6$ and 1.0 V. The same device as in Figure 6 is used. As can be seen from Figure 7a, $I_{IP3}$ curves at high $V_{DS}$ are shifted towards low $V_{GS}$ direction due to decreased threshold voltage, a consequence of drain induced barrier lowering. In strong inversion region, at the same $V_{GS}$, a higher $V_{DS}$ results in a higher $I_{IP3}$. For instance, at $V_{GS} = 0.8$ V, $I_{IP3}$ increases by 7.7 dB when $V_{DS}$ increases from 0.6 to 1.0 V. As shown in Figure 7b, $I_{IP2}$ has a clear peak, though not as sharp as $I_{IP3}$, around $V_{GS} = 0.6$ V, in strong inversion. If both high $I_{IP3}$ and high $I_{IP2}$ are desired, the transistor should be biased around $V_{GS} = 0.6$ V, which is approximately 200 mV above threshold voltage. A comparison of Figure 7c,d shows that the $V_{DS}$ dependence of $V_{GS,IP3}$ and hence $I_{IP3}$ is insufficiently captured by $V_{IP3}$, due to lack of $v_{ds}$ related terms, as expected.

Figure 8a shows measured $I_{IP3}$ at 5 GHz for devices with $W_{total} = 153.6$ and 256 µm. Note that the device finger widths are 0.3 and 1 µm respectively. At both very low and high $I_{DS}$, a large device gives a large $I_{IP3}$. Both peak $I_{IP3}$ value and peak $I_{IP3} I_{DS}$ decrease with device width. Narrow width effect clearly plays a role in affecting the position of the linearity peak. Figure 8b shows measured $I_{IP3}$ as a function of $V_{GS}$ for two 30 nm MOSFETs with the same total width of 256 µm. As the device finger
widths are both large, 2 and 4 µm respectively, no narrow width effect is observed, and $I_{IP3}$ is largely the same for the two devices as expected.

![Figure 7](image1)

**Figure 7.** Measured (a) $I_{IP3}$; (b) $I_{IP2}$ and (c) $V_{GS,IP3}$ and (d) $V_{IP3}$ as a function of $V_{GS}$ for different $V_{DS}$.

![Figure 8](image2)

**Figure 8.** Measured width impact on $I_{IP3}$. (a) Measured $I_{IP3}$ at 5 GHz for two 30 nm devices with different total width; (b) Measured $I_{IP3}$ as a function of $V_{GS}$ for two 30 nm devices with same total width.

### 4. Conclusions

We have presented experimental measurements and simulation of RF intermodulation linearity as a function of biases and frequencies on a 28 nm high-k/metal gate RF CMOS technology. Using Volterra series analysis, a new figure-of-merit, $V_{GS,IP3}$, is proposed. $V_{GS,IP3}$ can be experimentally determined from RF $IP3$ measurements, circuit simulations, or calculated from DC $I-V$ characteristics, and reduces to traditional $V_{IP3}$ when $V_{DS}$ dependence of $I_{DS}$ is neglected. Due to stronger impact of $V_{DS}$ on $I_{DS}$, a stronger impact of $V_{DS}$ on $I_{IP3}$ is observed compared to 90 nm technologies. The strong output conductance and related nonlinearities also cause a large separation between the $V_{GS,IP3}$/$I_{IP3}$ peak and $V_{IP3}$ peak. $V_{GS}$ dependence of $I_{IP2}$ is independent on that of $I_{IP3}$. A higher $V_{DS}$ is found to improve $I_{IP3}$ as well as $I_{IP2}$. 
Author Contributions

Zhen Li performed measurement and prepared manuscript. Guofu Niu supervised this research and finalized manuscript. Qingqing Liang contributed to technical oversight. Kimihiko Imura provided technical guidance for manuscript preparation.

Conflicts of Interest

Authors declare no conflict of interest.

References


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).