

Charge-Based Distortion Analysis of Nanoscale MOSFETs

Francesco Chicco¹, *Student Member, IEEE*, Alessandro Pezzotta¹,
and Christian C. Enz¹, *Senior Member, IEEE*

Abstract—This paper presents a study of MOSFETs' linearity, exploiting a simplified version of the charge-based EKV model. It allows to deduce analytically the one-tone and two-tone harmonic distortions introduced by the nonlinear I_D - V_G MOSFET characteristic as a function of the inversion coefficient. The short-channel effects are included in order to address nanoscale MOSFET performance. The analysis is validated through comparisons with the BSIM6 model and measurement results from 28-nm bulk CMOS devices. By means of this model, the designer can choose the appropriate bias region for the critical devices of a circuit depending on the system requirements.

Index Terms—Model, charge-based, EKV, MOSFET, inversion coefficient, distortion, velocity saturation.

I. INTRODUCTION

IN MIXED-SIGNAL integrated circuits, the linearity of an analog building block often represents a crucial aspect, from baseband to RF. Some examples are operational amplifiers, power amplifiers, low-noise amplifiers, analog-to-digital converters, G_m - C filters, etc., which can be found in sensor interfaces, instrumentation for measurements, audio applications, image sensors and wireless transceivers [1]–[9].

The improvement of linearity can be achieved by biasing the transistor in Strong Inversion (SI) with a large overdrive voltage $V_{GS} - V_{T0}$, at the cost of a higher power consumption [10]. Moreover, Velocity Saturation (VS) is becoming dominant in short-channel devices biased in SI, making their I_D - V_G characteristic almost perfectly linear. However, it is not possible to benefit from this improvement, due to the maximum overdrive voltage reduction imposed by technology scaling. Indeed, the supply voltage has reduced, while the threshold voltage has almost remained constant to preserve a low channel leakage current. Consequently, the operating point of MOSFETs has been progressively pushed towards Moderate (MI) and eventually Weak Inversion (WI), regimes in which the distortion caused by the nonlinear I_D - V_G characteristic increases more dramatically. On the other hand, these bias regions are convenient for low-power and low-voltage designs, which are required by applications such as Internet of Things.

Manuscript received March 30, 2018; revised June 24, 2018 and August 17, 2018; accepted August 19, 2018. Date of publication September 26, 2018; date of current version January 18, 2019. This paper was recommended by Associate Editor E. Tlelo-Cuautle. (*Corresponding author: Francesco Chicco.*)

The authors are with the Integrated Circuit Laboratory, École Polytechnique Fédérale de Lausanne, 2002 Neuchâtel, Switzerland (e-mail: francesco.chicco@epfl.ch).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSI.2018.2868387

Moreover, even though several nonlinearities arise as the operation frequency increases, i.e. those related to the parasitic capacitances of the MOSFET, they are typically much smaller than the nonlinearity introduced by the transconductance, as shown in [11]. This is valid especially in MI and WI, which are almost the only choices in a nanoscale technology, as pointed out above.

For these reasons, the analysis of the harmonic distortion in devices and building blocks has been a research topic for decades [10], [12]–[16]. Sansen [10] carried out one of the first systematic distortion analysis on BJTs and MOSFETs in order to explain the origin of frequency spurs in telecommunication circuits. Moreover, in this work all the metrics related to one-tone and two-tone analyses were defined. In [12] a comparison in terms of RF performance among several CMOS nodes (from 350-nm to 50-nm) was presented. The third-order Input Intercept Point was derived using the first- and third-order gate transconductances G_{m1} and G_{m3} obtained from DC measurements. Kang *et al.* took into account also the nonlinear behavior of the output conductance G_{ds} in the Taylor expansion of the drain current. They used the BSIM3 model and compared the simulated results with measurements on 180- and 250-nm devices. In [14] both G_m and G_{ds} nonlinearity were accounted for: the inclusion of cross-terms in the Taylor expansion allowed to optimize the design of a LNA with 65-nm devices reducing the second-order distortion. Cheng *et al.* [15] presented a general model for weak nonlinearity which takes into account all nonlinear sources in the MOSFET, namely both transconductances and parasitic capacitances. They used the PSP model and measured devices and circuits in 90-nm node. Jespers and Murmann [16] used the core long-channel equations of EKV model in order to express the first-, second- and third-order gate transconductances as a function of the normalized inversion charge and then of the transconductance efficiency G_m/I_D . Moreover, they took into account also the nonlinearity introduced by G_{ds} . The analytical results were compared to simulations carried out with the PSP model. However, their approach is proposing a design methodology using G_m/I_D as the main design parameter. The current density is then obtained from G_m/I_D through simulations with a compact model that includes VS and second-order effects. Although this design methodology is effective, it actually relies on a compact model. Instead, the approach proposed in this paper is a simple self-consistent model that includes the impact of VS on distortion without requiring any compact model.

Taking advantage of the theory developed about MOSFETs nonlinearity in the literature aforementioned, improvements

can be introduced in the transistor model, in order to describe more accurately the harmonic distortion behavior in all the bias regions. Besides, it is important to keep the analysis simple, making it a powerful tool in the design phase.

When dealing with older technology nodes, MOSFETs behavior is well-described by the quadratic I_D - V_G expression in SI and by the exponential one in WI. The latter shows better transconductance efficiency with respect to the former, which means higher transconductance for a given current, at the cost of larger area. On the other hand, if linearity is the strongest limitation, SI is the optimal choice. Nevertheless, in the case of nanoscale technologies, the old models are not suited anymore to describe the MOSFET behavior due to the presence of VS.

This is the reason why it is important to analyze the linearity performance of advanced technologies by means of a model which takes into account this effect. On the other hand, as mentioned above, the same model should be essential enough to keep the analysis simple and the results easily employed in the design process. Following the work in [16], a simplified charge-based model would be suited for this work: the best choice would be to exploit the core equations of the EKV model, since they would allow to characterize the devices in all bias regions, from WI to SI [17]. The advantage of the EKV model with respect to other models is the decoupling of the drain current and the terminal voltages through the charge. Consequently, the metrics associated to the distortion analysis could be formulated as a function of the Inversion Coefficient (IC), which gives an immediate indication about the operation region of the device.

Moreover, the focus of this work is to accurately capture the behavior of the nonlinearity associated to G_m at low frequencies. In several circuits with current outputs this one is assumed to be the dominant contribution to the overall harmonic distortion. One of the most common examples of this class of circuits is the Operational Transconductance Amplifier (OTA) [18], which is used in several systems, such as G_m - C filters and switched-capacitor circuits. The extension of this model to voltage-mode circuits (such as Operational Amplifiers), where the additional effect of G_{ds} becomes relevant, is under investigation.

In the end, a test chip in 28-nm Bulk CMOS technology is designed and tested to validate the model on nanoscale MOSFETs.

This paper is structured as follows. Section II describes the charge-based model employed for the analysis, which is detailed in Section III. In Section IV the simulation and measurement setups are described and the results are compared. Section V draws the conclusions.

II. CHARGE-BASED SIMPLIFIED MODEL

The core equations of the simplified model for long-channel MOSFETs are the same as for the complete EKV model [19]:

$$2q_i + \log q_i = v_p - v \quad (1)$$

$$i_d = i_f - i_r = q_s + q_s^2 - q_d - q_d^2 \quad (2)$$

where q_i is the inversion charge density normalized to the specific charge

$$Q_{\text{spec}} \triangleq -2nU_T C_{\text{ox}}, \quad (3)$$

v_p and v are the pinch-off and channel voltages respectively normalized to the thermal voltage $U_T = kT/q$, i_d is the drain current normalized to the specific current

$$I_{\text{spec}} = I_{\text{spec}\square} \frac{W}{L} \quad \text{with} \quad (4)$$

$$I_{\text{spec}\square} \triangleq 2n\mu_0 C_{\text{ox}} U_T^2 \quad (5)$$

and q_s and q_d are the value of q_i at source and drain respectively. n is the slope factor, μ_0 is the constant low-field electron mobility and C_{ox} is the gate capacitance per unit area.

Since (1) is not invertible analytically, i_d cannot be expressed in closed form as a function of the voltage, unless the expressions are simplified for low or high values of q_i . For this reason, the pinch-off voltage $V_P = (V_G - V_{T0})/n$ is not a convenient way to explore all the bias regions of the MOSFET. Indeed, the proposed method relies on the metric that allows to precisely address the channel inversion level of a MOSFET, namely the inversion coefficient IC ,

$$IC \triangleq \frac{I_D|_{\text{saturation}}}{I_{\text{spec}}} = i_{\text{d}\text{sat}}. \quad (6)$$

Note that being normalized to I_{spec} , IC strips off any size and technology dependence. The inversion regions are classified in terms of IC as follows:

$$IC < 0.1 : \text{Weak Inversion (WI)}$$

$$0.1 < IC < 10 : \text{Moderate Inversion (MI)}$$

$$IC > 10 : \text{Strong Inversion (SI)}$$

In long-channel devices, q_d vanishes at pinch-off giving

$$i_{\text{d}\text{sat}} = q_s + q_s^2. \quad (7)$$

Nevertheless, in the most advanced technology nodes (7) fails in describing correctly the behavior of minimum- and close-to-minimum-length devices due to VS, which has a dramatic impact on the drain current and hence on the transconductance. Indeed, the electron mobility μ is not constant for high values of horizontal electric field in the channel E_x : the shorter the channel, the more this phenomenon affects negatively the devices performance. Consequently, (2) is not valid anymore, being it derived from the drift-diffusion equation without including VS. It is necessary to go back to the drift-diffusion equation and to solve it including the bias dependence of the mobility. The mobility reduction caused by high values of the vertical electric field is not taken into account in this work.

There are several models to describe the dependence of the effective electron mobility μ_{eff} on E_x : in this work a simple piecewise linear model is employed,

$$\mu_{\text{eff}}(E_x) \triangleq \frac{v_{\text{drift}}}{|E_x|} = \begin{cases} \mu_0 & \text{for } E_x < E_c \\ v_{\text{sat}}/|E_x| & \text{for } E_x \geq E_c \end{cases} \quad (8)$$

where v_{sat} is the maximum electron velocity and E_c is the critical electric field, which depends only on the substrate properties,

$$E_c \triangleq \frac{v_{\text{sat}}}{\mu_0}. \quad (9)$$

Including this model in the drift-diffusion equation leads to the same expression as (2); nevertheless, in this case

VS happens before pinch-off and hence q_d saturates to $q_{d,\text{sat}}$, which is a specific value set by the bias conditions and the channel length [19]. If $q_{d,\text{sat}}$ is expressed in terms of q_s , $i_{d,\text{sat}}$ then becomes

$$q_{d,\text{sat}} = \frac{2\lambda_c (q_s + q_s^2)}{2 + \lambda_c + \sqrt{4(1 + \lambda_c) + \lambda_c^2(1 + 2q_s)^2}} \quad (10)$$

$$i_{d,\text{sat}} = \frac{4(q_s + q_s^2)}{2 + \lambda_c + \sqrt{4(1 + \lambda_c) + \lambda_c^2(1 + 2q_s)^2}} \quad (11)$$

where

$$\lambda_c = \frac{L_{\text{sat}}}{L} \quad (12)$$

is the fraction of the channel under full velocity saturation which scales as $1/L$ and L_{sat} the saturated portion of the channel. L_{sat} is another technology parameter extracted from measurements and it is ideally unique for any transistor length. Consequently, it allows to transit smoothly from the drain current for short-channel devices (11) to the one for long-channel devices (7).

For $V_D > V_{D,\text{sat}}$, I_D is taken constant in this model, neglecting the effect of Channel Length Modulation (CLM) and Drain-Induced Barrier Lowering (DIBL). Nevertheless, this choice has little impact on the following analysis: as it will be shown in Section IV, V_D is kept constant in the experimental setup by using a TIA at the drain of the Device Under Test (DUT), behaving as an AC ground. Consequently, the impact of the output conductance G_{ds} is minimized. For the same reason, also the junction capacitance C_{DB} doesn't contribute with additional nonlinearity. Nevertheless, due to the assumption of low-frequency operation, the contribution of all parasitic capacitances is negligible.

Since this simplified EKV model is built on normalized quantities, it has the advantage of being independent from any technology. In fact, in order to employ it, only 4 technology parameters are needed: $I_{\text{spec}\square}$, n , V_{T0} and L_{sat} . They can be easily extracted from the measured I_D - V_G characteristic of the device of interest. These parameters allow to normalize the input terminal voltages and to denormalize the drain current.

III. HARMONIC ANALYSIS

For the scope of this work, only the saturation region of the MOSFET is taken into account: for the rest of the paper $i_d = i_{d,\text{sat}}$. At the gate, the input voltage $V_G = V_{G0} + \Delta V_G$ provides both the bias and the signal. In order to describe the large-signal AC behavior of the MOSFET, first the nonlinear relation between I_D and V_G is expressed in the form of a Taylor expansion around the bias point of the device:

$$I_D = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k I_D}{\partial V_G^k} \bigg|_{V_{G0}} \Delta V_G^k = \sum_{k=0}^{\infty} \frac{1}{k!} G_{mk} \Delta V_G^k, \quad (13)$$

where G_{mk} is the gate transconductance of order k .

Being under low distortion conditions [10], (13) can be approximated to the third-order:

$$I_D \simeq I_{D0} + G_{m1} \Delta V_G + \frac{G_{m2}}{2} \Delta V_G^2 + \frac{G_{m3}}{6} \Delta V_G^3. \quad (14)$$

This choice allows to take into account the effect of both even and odd order harmonics, trading off accuracy and complexity. It has to be noted that in case of strongly nonlinear behaviors, this approximation would provide inaccurate results.

Two different analyses are carried out, namely with a one-tone input signal and with a two-tone one.

A. One-Tone Analysis

Assuming the input voltage signal to be $\Delta V_G|_I = A \cos(\omega t)$, (14) can be decomposed in terms of the three harmonic components; Although well-known, these expressions are reported for the sake of completeness and readability [10].

$$I_D|_I \simeq I_{D(0)}|_I + I_{D(1)}|_I \cdot \cos(\omega t) + I_{D(2)}|_I \cdot \cos(2\omega t) + I_{D(3)}|_I \cdot \cos(3\omega t), \quad (15)$$

where

$$I_{D(0)}|_I = I_{D0} + \frac{G_{m2}A^2}{4}, \quad (16)$$

$$I_{D(1)}|_I = G_{m1}A + \frac{G_{m3}A^3}{8}, \quad (17)$$

$$I_{D(2)}|_I = \frac{G_{m2}A^2}{4}, \quad (18)$$

$$I_{D(3)}|_I = \frac{G_{m3}A^3}{24}. \quad (19)$$

These formulas show that the second-order nonlinear term in (14) generates a DC offset while the third-order one influences the fundamental, either decreasing (compression) or increasing its amplitude (expansion) depending on the sign of G_{m3} . This behavior can be generalized to all nonlinear terms of order higher than three as follows: even-order terms contribute to DC offset while odd-order ones affect the fundamental.

In order to quantify the linearity performance of a device, several metrics can be derived from the output tone amplitudes in (17)-(19). First, the second-order and third-order Harmonic Distortion parameters, i.e. HD_2 and HD_3 , are defined as the ratio of the amplitude of the second and third harmonic versus the amplitude of the fundamental respectively:

$$\text{HD}_2 \triangleq \left| \frac{I_{D(2)}|_I}{I_{D(1)}|_I} \right| = \frac{2G_{m2}A}{8G_{m1} + G_{m3}A^2} \quad (20)$$

$$\text{HD}_3 \triangleq \left| \frac{I_{D(3)}|_I}{I_{D(1)}|_I} \right| = \frac{G_{m3}A^2}{3(8G_{m1} + G_{m3}A^2)}. \quad (21)$$

Note that, in order to keep these expressions as general as possible, $I_{D(1)}|_I$ should not be approximated with the ideal value $G_{m1}A$ because the additional term $G_{m3}A^3$ may be relevant to achieve a better accuracy.

The disadvantage of these two parameters is that they depend on the input signal amplitude: they cannot describe the performance of a device or a circuit with an unique value.

On the contrary, this is achieved by another metric, the 1 dB compression (expansion) point $A_{\mp 1\text{dB}}$, defined as the input amplitude for which the fundamental tone in the output signal, $I_{D(1)}|_I$ is reduced (increased) by 1 dB with respect to the ideal value, $G_{m1}A$:

$$A_{\mp 1\text{dB}} = \sqrt{\pm \left(1 - 10^{\mp \frac{1}{20}}\right) \left| \frac{8G_{m1}}{G_{m3}} \right|}. \quad (22)$$

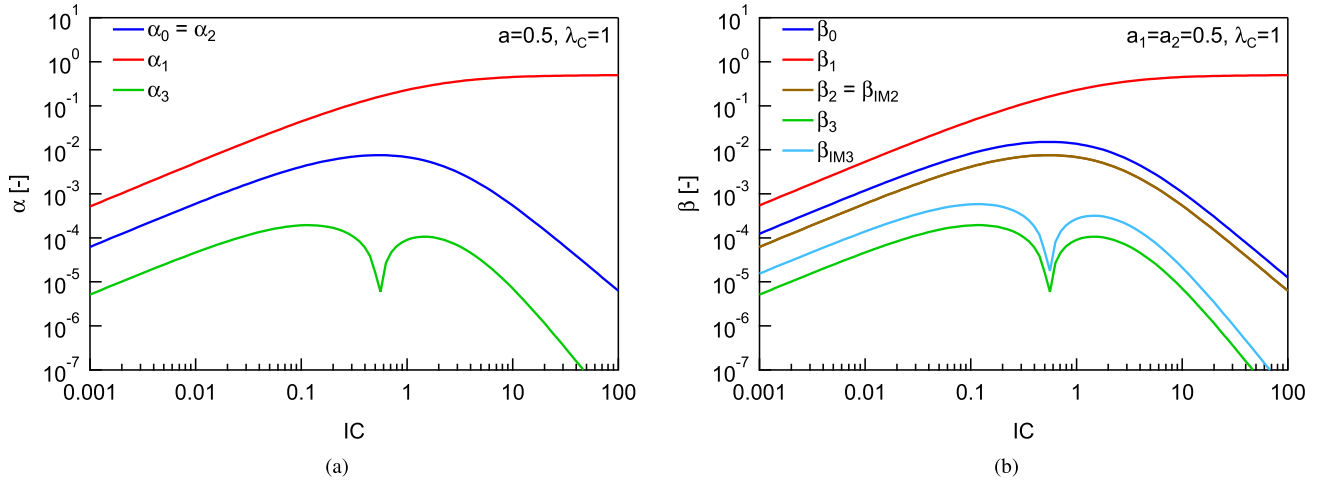


Fig. 1. Amplitude of the harmonics of the output current in the case of (a) one-tone and (b) two-tone analyses, normalized to I_{spec} .

B. Two-Tone Analysis

Assuming the input voltage signal to be $\Delta V_{G|\text{II}} = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t)$, (14) can be decomposed accordingly as $I_{D|\text{II}}$ in terms of the harmonic components, which are reported in the Appendix [10].

Several metrics can be calculated also in the case of two-tone analysis. The most valuable parameters are the second-order and third-order Intercept Point, A_{IP2} and A_{IP3} respectively, defined as the input amplitude for which the ideal fundamental tone, $G_{m1}A_1$ (or $G_{m1}A_2$), and the second- or third-order intermodulation product, $I_{D(\text{IM}2)|\text{II}}$ and $I_{D(\text{IM}3,1)|\text{II}}$ (or $I_{D(\text{IM}3,2)|\text{II}}$) respectively, have the same amplitude in the output signal. Assuming $A = A_1 = A_2$, they are expressed as follows:

$$A_{\text{IP2}} = 2 \left| \frac{G_{m1}}{G_{m2}} \right| \quad (23)$$

$$A_{\text{IP3}} = \sqrt{8 \left| \frac{G_{m1}}{G_{m3}} \right|} \quad (24)$$

C. Derivation of the Normalized Transconductances

The objective of this analysis is to derive analytically the harmonic coefficients, and consequently all the metrics defined in the previous section, as a function of IC , by means of the simplified model shown in Section II. First, the normalized form of G_{m1} , G_{m2} and G_{m3} can be conveniently expressed applying the composite derivative method [20]:

$$g_{mk} \triangleq \frac{G_{mk}}{G_{\text{spec}} / (n^k U_T^{k-1})} = \frac{\partial^k i_{\text{dsat}}}{\partial v_g^k} = \frac{\partial^k i_{\text{dsat}}}{\partial q_s^k} \left(\frac{\partial v_g}{\partial q_s^k} \right)^{-1}. \quad (25)$$

Then, the calculation of the derivatives in (25) is straightforward starting from (1) and (11), leading to the normalized expression of G_{m1} , G_{m2} and G_{m3} as a function of q_s :

$$g_{m1} = \frac{a-1}{\sqrt{4+4\lambda_c+a^2\lambda_c^2}}, \quad (26)$$

$$g_{m2} = \frac{g_{m1}}{a} \frac{4+4\lambda_c+a\lambda_c^2}{4+4\lambda_c+a^2\lambda_c^2}, \quad (27)$$

$$g_{m3} = \frac{g_{m1}}{a^3} \frac{16+32\lambda_c+8b\lambda_c^2+8a^2c\lambda_c^3+a^3d\lambda_c^4}{(4+4\lambda_c+a^2\lambda_c^2)^2}, \quad (28)$$

where

$$\begin{aligned} a &= 1 + 2q_s, \\ b &= (1 - 2q_s)(3 + 7q_s + 6q_s^2), \\ c &= v_1 - 3q_s, \\ d &= 1 - 4q_s, \end{aligned}$$

and $G_{\text{spec}} = I_{\text{spec}}/U_T$ is the specific transconductance.

Finally, the normalized transconductances can be expressed as a function of IC inverting (11) and replacing $q_s(IC)$ in (26)-(28) with

$$q_s(IC) = \frac{\sqrt{(1+\lambda_c IC)^2 + 4IC} - 1}{2}. \quad (29)$$

After the denormalization of $g_{m1}(IC)$, $g_{m2}(IC)$ and $g_{m3}(IC)$, all the parameters shown in Section III-A and III-B can be plotted as a function of IC as well. Fig. 1 shows the harmonics amplitude of the output current normalized to I_{spec} , α and β respectively, resulting from the one-tone and the two-tone analyses. a , a_1 and a_2 are the input voltage amplitudes for the two analyses normalized to U_T .

Note that (26)-(28) are consistent with the results in [16]: the latter can be simply obtained by setting $\lambda_c = 0$ in the former, which is equivalent to impose the long-channel case.

Moreover, the simplified EKV model is capable to reproduce precisely the well-known singularity in α_3 , β_3 and $\beta_{\text{IM}3}$. This behavior is due to the fact that G_{m3} changes sign going from WI to SI and so there is a value of IC for which it is equal to 0, namely IC_{crit} . Note that the value of IC_{crit} depends uniquely on λ_c : since this parameter is by definition always between 0 and 1, it is easy to show that it tends asymptotically to infinity (SI) for long-channel devices, while it gets to MI when λ_c increases. In Fig. 2 the position of IC_{crit} is plotted as a function of λ_c . It is evident that the singularity cannot occur in WI. The relevance of L_{sat} on IC_{crit} proves that the inclusion of VS is indispensable to describe well the behavior of a minimum-length MOSFET. Moreover, Fig. 2 shows that for short-channel devices MI offers a good trade-off for linearity in addition to area and current consumption [21].

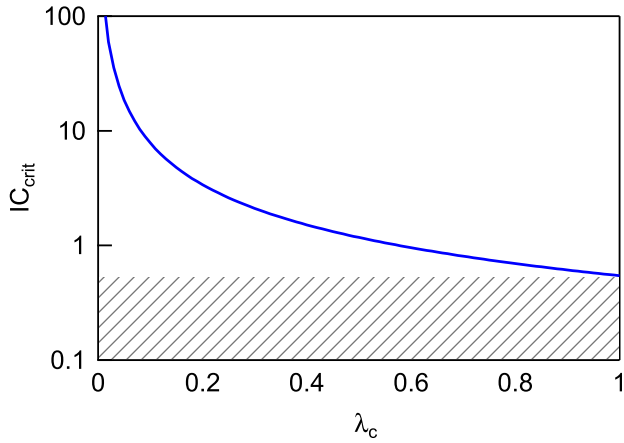


Fig. 2. Position of the singularity in G_{m3} (IC_{crit}) as a function of the saturation of the channel (λ_c).

IV. VALIDATION OF THE ANALYSIS

In order to validate the accuracy of the simplified EKV model shown in Section II in predicting the harmonic distortion in a MOSFET, the analytical expressions are compared to simulations carried out with the same sEKV coded in Verilog-A, with the BSIM6 model and with measurements from 28-nm Bulk CMOS devices.

The DUT is an nMOS with $L = 30$ nm and $W = 3$ μ m. The effective dimensions are $L = 27$ nm and $W = 2.7$ μ m due to the 0.9 shrinking factor of the technology. The device has a single finger and it is wide enough to minimize the contribution of the drain access resistance.

A. Simulation With Verilog-A and BSIM6 Model

The simplified EKV model (sEKV) is coded in Verilog-A to make it available for designers in simulation environments. Moreover, such simplified model is compared with a full compact model, i.e. BSIM6, to show the validity of the proposed approach.

The sEKV Verilog-A code takes the terminal voltages as inputs, together with 4 technology parameters ($I_{spec\Box}$, n , V_{T0} and L_{sat}), the drawn dimensions W and L , the shrink factor of the technology and the thermal voltage U_T ; it provides the drain current as output. Since (1) cannot be inverted analytically, q_s and q_d are computed with a non-recursive function which inverts it numerically achieving very good accuracy in spite of its simplicity. The same function is used in BSIM6 to calculate the inversion charge. Depending on the value of q_d , either (2) or (11) is chosen and then denormalized with I_{spec} to provide the output current. Note that internally the model works with all normalized quantities. The input voltages are normalized to U_T before being used. As mentioned in Section II, isolating all the technology dependence in 4 parameters makes the model easily portable from one technology node to another.

In order to extract the value of the technology parameters for sEKV model and the DC model card for BSIM6, the static I_D - V_G characteristic of the Device Under Test (DUT) is measured. The details about the measurements are reported in Section IV-B. The DC model card for BSIM6 consists of a subset of all the BSIM6 parameters which allows to

TABLE I
THE 4 PARAMETERS OF THE SIMPLIFIED EKV

Parameter	Value
$I_{spec\Box}$	1.07 μ A
n	1.6
V_{T0}	490 mV
L_{sat}	14.5 nm

describe faithfully only the DC behavior of the device. Since the claim of this work is that the first-, second- and third-order gate transconductances are sufficient to describe the harmonic distortion behavior, this kind of model card is supposed to be accurate enough. Both models are fitted to the measured DC drain current (both in linear and logarithmic scale) and to the first-, second- and third-order numerical derivatives (i.e. the three transconductances) by means of an optimization routine in Keysight IC-CAP[®]. In the case of BSIM6, the fitting procedure follows the instructions in [22], while for sEKV it follows [17]. Concerning the bias conditions, the sEKV model is fitted only to the curve at $V_D = 1.1$ V, while the BSIM6 model is fitted for several values of V_D , from linear to saturation region. Table I shows the 4 parameters of the simplified EKV obtained by the fitting. While BSIM6 is a scalable model and it uses a unique set of parameters, sEKV is not: n changes with L and $I_{spec\Box}$ also through n , and consequently they need to be extracted for each length used in the design. Nevertheless, in a consistent extraction, $I_{spec\Box}/n$ should be kept constant, as well as L_{sat} .

The two models are used to carry out simulations in Keysight ADS[®]. First, DC simulations are carried out to extract the I_D - V_G characteristic of the device, both in linear and logarithmic scale. Then, the first-, second- and third-order gate transconductances are obtained by derivating numerically I_D - V_G . Finally, the first, second and third harmonics of I_D are extracted from Harmonic Balance (HB) simulations.

The simulated testbench is built in such a way to mimic as much as possible the experimental setup described in Section IV-B: the TIA is replaced by an ideal operation amplifier with a feedback resistor equal to the inverse of the TIA sensitivity and a constant voltage $V_{bias} = 1.1$ V on the positive terminal to set V_D on the other one.

B. Measurements

In order to validate the analysis presented so far, measurements are carried out on 28-nm Bulk CMOS samples. The nominal maximum voltage which can be applied to the terminals of the devices in such technology is 1 V. Nevertheless, a margin of 10% is allowed and consequently for these measurements V_{DD} is set to 1.1 V in order to explore SI as much as possible.

Regarding the DC measurements, the chip is tested with a probe-card connected to the Keysight B2201A Switching Mainframe. The 4 terminals of the device are controlled by the Keysight Semiconductor Analyzer, which generates the bias voltages and measures the current through them. V_G is swept from 0 to 1.1 V by 25 mV steps and this measurement is repeated sweeping V_D from 0 V to 1.1 V by 100 mV steps.

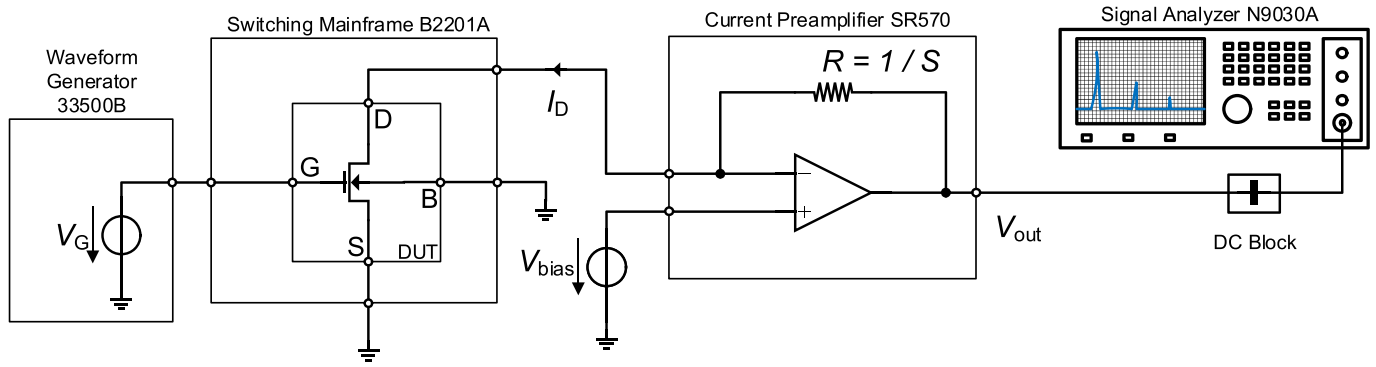


Fig. 3. Experimental setup for AC measurements.

Regarding the large-signal AC measurements, the objective is to measure the amplitude of the first, second and third harmonic of the drain current I_D . Nevertheless, it is way more practical to measure the spectrum of a voltage rather than a current. For this reason, I_D is converted to a voltage by means of a Trans-impedance Amplifier (TIA).

Fig. 3 shows the experimental setup for the large-signal AC measurements. The chip is still accessed through the Switching Mainframe and the probe-card. The sinusoidal input signal as well as the DC component is generated with the Keysight 33500B Waveform Generator and connected to the gate probe. Source and bulk probes are connected to ground and the N-well probe for the ESD diodes is biased to supply voltage $V_{DD} = 1.1$ V, both generated by the Keysight E2646A Power Supply. The drain probe is connected to the negative input of the Stanford SR570 Low-Noise Current Preamplifier, used as TIA. In order to bias this node to the proper voltage, the positive input of the TIA is connected to V_{DD} and the negative feedback is exploited to fix the other input. The sensitivity is set to $500 \mu\text{A V}^{-1}$. In the end, the output of the TIA is connected to the Keysight N9030A PXA Signal Analyzer. A DC block capacitor is placed before the PXA to allow the use of the DC-coupled mode.

In order to explore all the operation regions, from WI to SI, the DC component of the input signal V_{G0} is swept from 0.1 V to 1.1 V by 25 mV steps. Moreover, the amplitude of the sinusoid is set to 50 mV for both the one-tone and the two-tone analysis. The frequency of the one-tone signal is set to 6 kHz. Instead, the two tones are generated from one tone at 5.5 kHz amplitude modulated by another tone at 500 Hz, resulting in two tones at 5 and 6 kHz. The amplitude of the modulated signal is set to twice the target amplitude for the subcarriers being the modulation coefficient equal to 0.5.

C. Comparison

The results obtained from the simulations and the measurements are processed and compared to the analytical expressions. The latter are based on the normalized transconductances shown in (26)-(28), which are calculated using the 4 parameters of the sEKV model (Table I) and denormalized following (25).

Fig. 4 shows the comparison among the simulated I_D - V_G curve with sEKV and BSIM6 and the measurements, in both linear and logarithmic scale. The match among the three curves

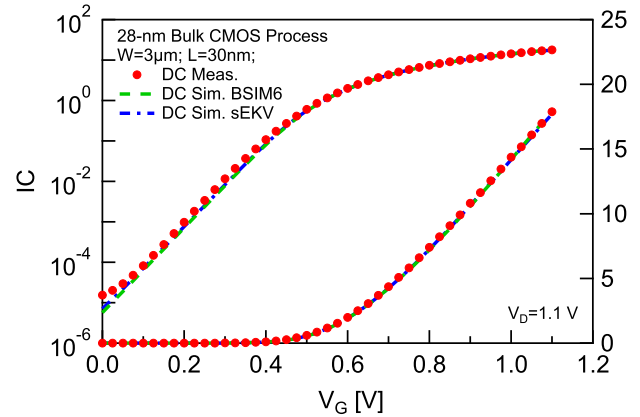
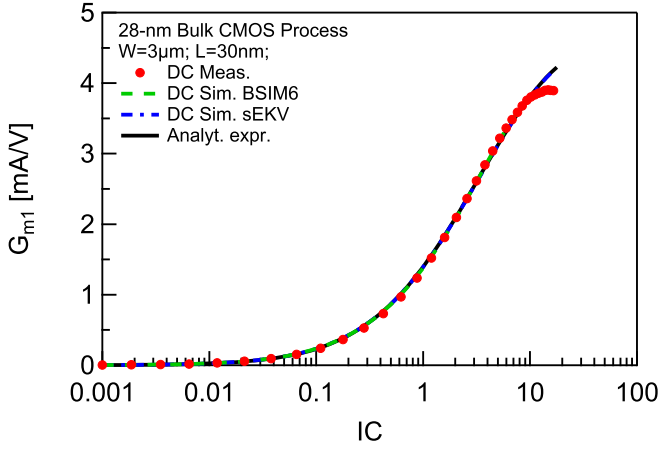


Fig. 4. Comparison among simulated and measured I_D - V_G , in linear scale (left axis) and logarithmic scale (right axis).

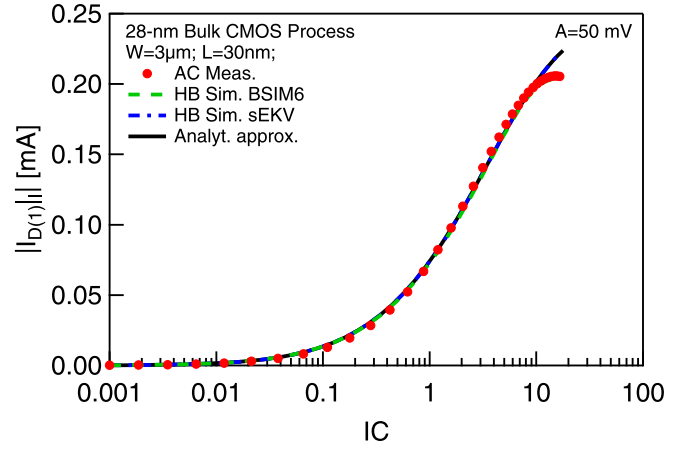
is very good from WI to SI in both scales. The analytical expression (11) is not plotted because equal to the core equation of the sEKV Verilog-A model. It is evident the difficulty in biasing the transistor in SI, as pointed out in the introduction. Indeed, the highest IC achievable in lower than 20 at V_G and V_D equal to 1.1 V, which is beyond the nominal V_{DD} as mentioned at the beginning of Section IV-B.

The analytical expressions of G_{m1} , G_{m2} and G_{m3} (26)-(28) are compared to the numerical derivation of the simulated I_D using the sEKV and the BSIM6 models and those obtained derivating the measured I_D in Fig. 5. The match between the curves is very good: this proves that the sEKV model is accurate in describing the DC behavior of the MOSFET. It can be noticed that the measured and the BSIM6 G_{m2} change sign for IC close to 20: it is due to mobility reduction caused by the vertical electric field. Indeed, the sEKV G_{m2} remains positive because it does not include such effect.

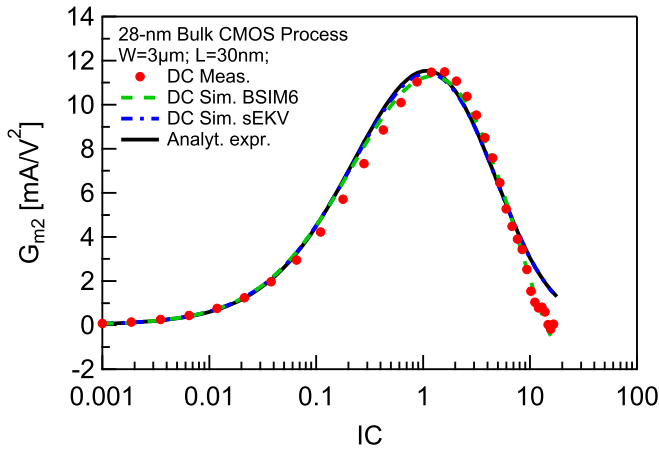
The analytical approximation of the amplitudes of the three drain current harmonics, namely $I_{D(1)|I}$, $I_{D(2)|I}$ and $|I_{D(3)}|I$ (17)-(19), are compared to those obtained by a one-tone HB simulation using the two models and those measured on the DUT with a one-tone test in Fig. 6. The 4 curves clearly match from WI to SI. The meaning of this result is twofold: it confirms that the sEKV Verilog-A model is compatible with a HB simulation regarding the nonlinear behavior of the device, and it also supports the claim that the DC model is sufficient to describe such behavior precisely.



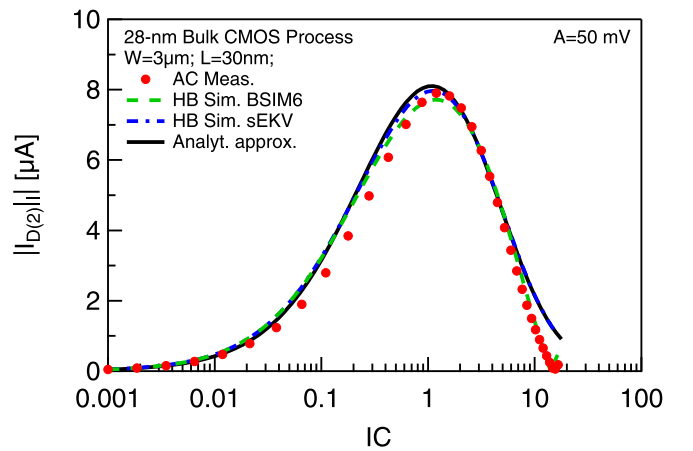
(a)



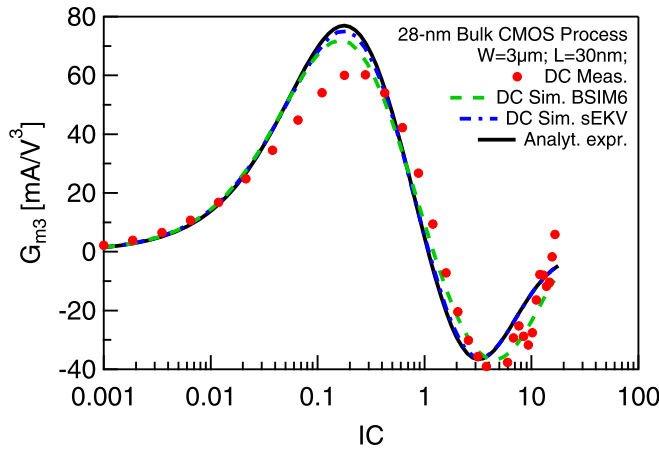
(a)



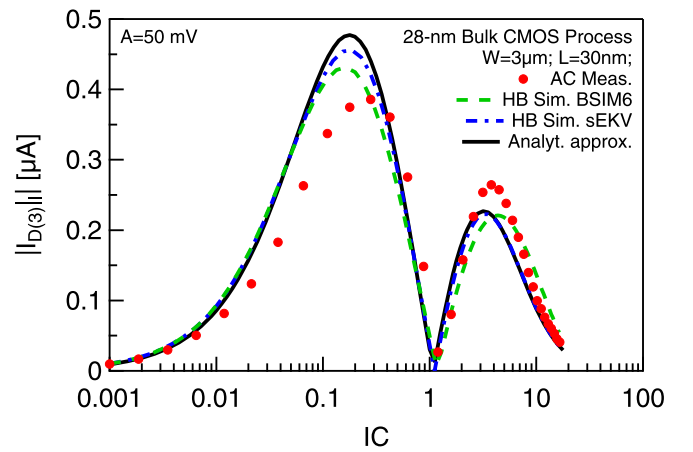
(b)



(b)



(c)



(c)

Fig. 5. Comparison among the DC measured, the DC simulated and the analytical (a) G_{m1} , (b) G_{m2} and (c) G_{m3} .

Fig. 6. Comparison among the AC measured, the HB simulated and the analytically approximated (a) $I_{D(1)|I}$, (b) $I_{D(2)|I}$ and (c) $I_{D(3)|I}$.

It is noted again that these results are valid in the assumption of low-frequency operation.

The analytical approximation of HD_2 and HD_3 obtained from (20)-(21) are compared to simulations and measurements in Fig. 7; the latter are calculated from the drain current harmonics (Fig. 6) using the definition of HD_2 and HD_3 . The match among the curves is very good for both parameters. This result follows exactly what stated above regarding Fig. 6.

Moreover, the singularity mentioned in Section III-C is effectively caught and it is located in the middle of the MI region. This confirms once more that this bias region is very convenient for several trade-offs [21], [23]. Nevertheless, in practice it is quite difficult to exploit such singularity, due to process and temperature variations. Still, even if the third-order distortion is not fully canceled, IC values around this point represent an interesting trade-off [15], [24].

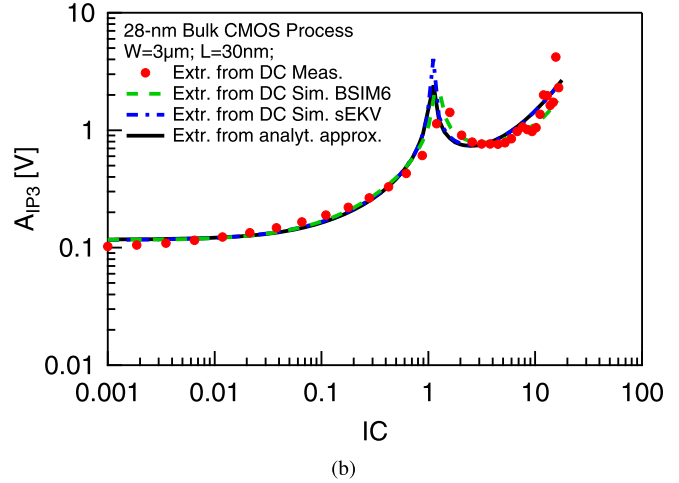
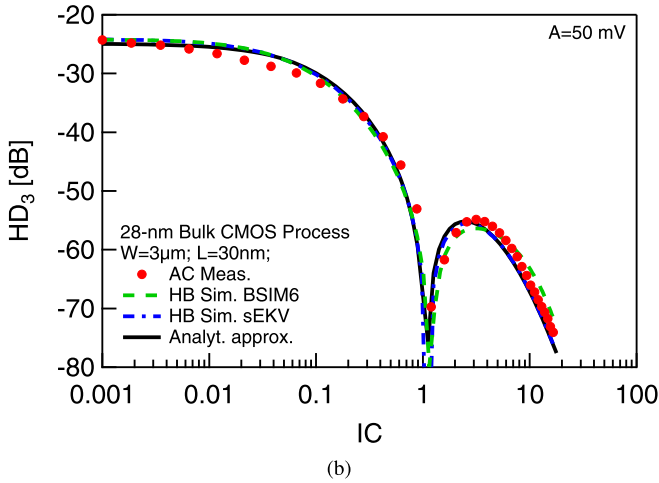
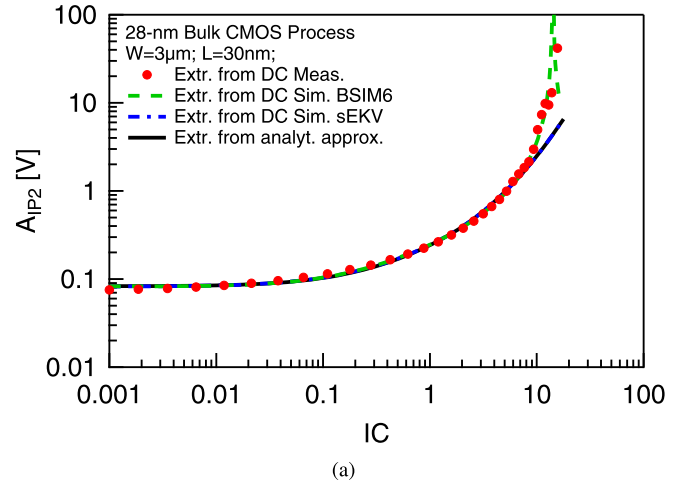
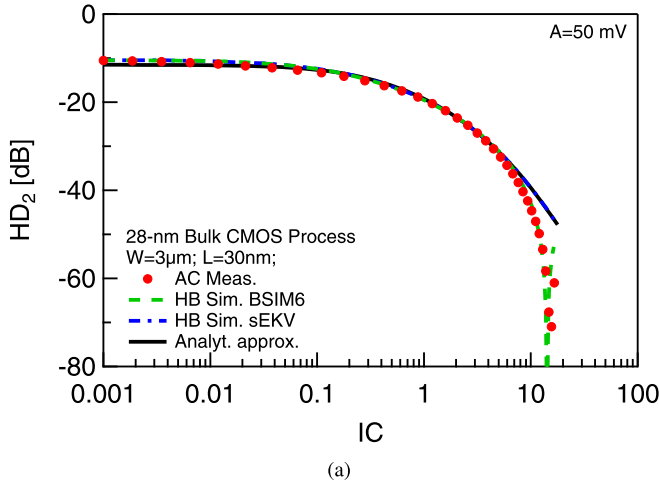


Fig. 7. Comparison among the AC measured, the HB simulated and the analytically approximated (a) HD_2 and (b) HD_3 .

Fig. 9. Comparison among extrapolated (a) A_{IP2} and (b) A_{IP3} from DC measurements, DC simulations and analytical expressions.

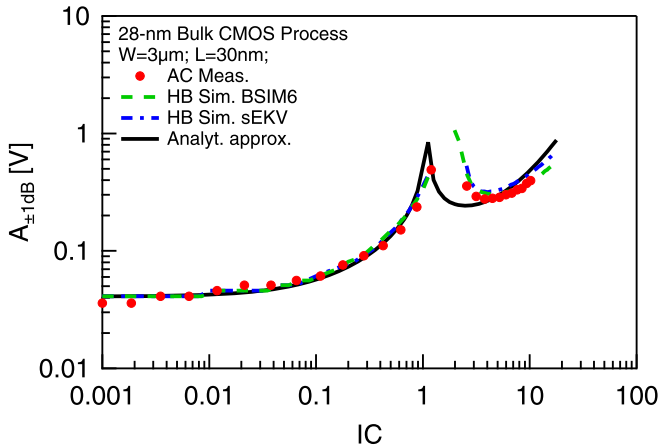


Fig. 8. Comparison among the AC measured, the HB simulated and the analytically approximated $A_{\pm 1dB}$.

Another singularity appears in HD_2 with BSIM6: it comes from the change of sign of G_{m2} , which has been already discussed above.

Fig. 8 and Fig. 9 report three RF design metrics, namely $A_{\pm 1dB}$, A_{IP2} and A_{IP3} , as a function of IC . The goal is to prove that the proposed model is a powerful tool to gain insight in the device and circuit performance during the design process. The analytical approximation of $A_{\pm 1dB}$ is compared

to simulations and measurements in Fig. 8. For the latter, the amplitude of the fundamental is measured and simulated increasing the signal amplitude until reaching the 1 dB deviation with respect to the ideal linear extrapolation. In the measurements of $A_{\pm 1dB}$, the signal amplitude superimposed to the bias voltage is limited to 30% above V_{DD} to avoid deteriorating the device. There are hence no values beyond 0.5 V but still the trend is very clear. The presence of a low impedance at the drain of the device allows to measure $A_{\pm 1dB}$ without the occurrence of voltage clipping. Fig. 8 shows a good match up to the peak; however, the simple analytical approximation slightly underestimate IC_{crit} . This is due to the fact that (22) does not account for higher order harmonics above the third, which become important particularly at this critical point where amplitude grows significantly. Since G_{m3} changes sign, in the same plot there are both A_{+1dB} and A_{-1dB} . For IC values smaller than the sweet spot IC_{crit} , G_{m3} is positive and therefore the device shows expansive behavior (A_{+1dB}), while for IC values greater than IC_{crit} , G_{m3} becomes negative and the device has compressive behavior (A_{-1dB}).

A_{IP2} and A_{IP3} are instead extrapolated from (23)-(24) using the gate transconductances obtained from the analytical expressions, DC simulations and DC measurements. The match is once more very good, especially in WI and MI.

A_{IP2} presents a singularity similarly to HD_2 due to mobility degradation which is well-predicted only by BSIM6, while for A_{IP3} all the curves describe accurately the peak in MI.

The outcome of this comparison proves that it is possible both to estimate the amplitude of the current harmonics by means of a DC simulation by computing the three gate transconductances from the DC drain current and also to extract a small-signal parameters from a HB simulation. Moreover, there is a direct relation between the IC and the most relevant RF design metrics for linearity through the analytical expressions for G_{m1} , G_{m2} , G_{m3} and q_s (26)-(29).

V. CONCLUSION

The MOSFET linearity performance is analyzed through a charge-based model. The advantage of the simplified EKV model is to decouple the drain current and the terminal voltages through the charge. This enables the derivation of analytical expressions for the first-, second- and third-order gate transconductances including the effect of Velocity Saturation, one of the main improvements with respect to previous works. This feature allows to describe the behavior of nanoscale technologies while keeping the model simple.

All the metrics for one- and two-tone analyses are derived as a function of the Inversion Coefficient, enabling a full device characterization in all inversion conditions. Indeed, a DC model is demonstrated to be sufficient for the analysis at low-frequency operation, since all the metrics depend only on G_{m1} , G_{m2} and G_{m3} . In detail, the evident ‘‘sweet spot’’ in HD_3 , A_{IP3} and A_{1dB} is caused by the change of G_{m3} sign, which is effectively captured only when VS is included. Such singularity is located at a specific IC value, defined as IC_{crit} , whose value depends only on λ_c .

The model is implemented in Verilog-A and compared with the analytical expressions, a DC BSIM6 model and with measurements carried out on a 28-nm Bulk CMOS technology. The match between the three cases is very good from WI to SI for all metrics. For this technology IC_{crit} is close to 1, confirming the MI region as an interesting operating point in terms of trade-offs.

Circuit designers can benefit from this analysis to assess the performance of a given technology or to choose the bias region of devices which are critical from the linearity point of view.

APPENDIX

TWO-TONE ANALYSIS: HARMONIC COMPONENTS

In this appendix the harmonic components of $I_{D|II}$ are reported:

DC term:

$$\omega = 0 \Rightarrow I_{D(0)|II} = I_{D0} + \frac{G_{m2}(A_1^2 + A_2^2)}{4} \quad (30)$$

1st Harmonics:

$$\omega = \omega_1 \Rightarrow I_{D(1,1)|II} = G_{m1}A_1 + \frac{G_{m3}}{4} \left(A_1A_2^2 + \frac{A_1^3}{2} \right) \quad (31a)$$

$$\omega = \omega_2 \Rightarrow I_{D(1,2)|II} = G_{m1}A_2 + \frac{G_{m3}}{4} \left(A_1^2A_2 + \frac{A_2^3}{2} \right) \quad (31b)$$

2nd Harmonics:

$$\omega = 2\omega_1 \Rightarrow I_{D(2,1)|II} = \frac{G_{m2}A_1^2}{4} \quad (32a)$$

$$\omega = 2\omega_2 \Rightarrow I_{D(2,2)|II} = \frac{G_{m2}A_2^2}{4} \quad (32b)$$

3rd Harmonics:

$$\omega = 3\omega_1 \Rightarrow I_{D(3,1)|II} = \frac{G_{m3}A_1^3}{24} \quad (33a)$$

$$\omega = 3\omega_2 \Rightarrow I_{D(3,2)|II} = \frac{G_{m3}A_2^3}{24} \quad (33b)$$

2nd-order Intermodulation products:

$$\omega = \omega_1 \pm \omega_2 \Rightarrow I_{D(IM2)|II} = \frac{G_{m2}A_1A_2}{2} \quad (34)$$

3rd-order Intermodulation products:

$$\omega = 2\omega_1 \pm \omega_2 \Rightarrow I_{D(IM3,1)|II} = \frac{G_{m3}A_1^2A_2}{8} \quad (35a)$$

$$\omega = 2\omega_2 \pm \omega_1 \Rightarrow I_{D(IM3,2)|II} = \frac{G_{m3}A_1A_2^2}{8} \quad (35b)$$

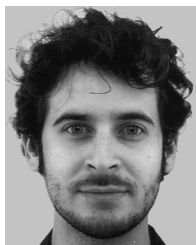
ACKNOWLEDGMENT

The sEKV Verilog-A code is available upon e-mail request to the authors.

REFERENCES

- [1] S. Pernici, G. Nicollini, and R. Castello, ‘‘A CMOS low-distortion fully differential power amplifier with double nested Miller compensation,’’ *IEEE J. Solid-State Circuits*, vol. 28, no. 7, pp. 758–763, Jul. 1993.
- [2] F. N. L. O. Eynde, P. F. M. Ampe, L. Verdeyen, and W. Sansen, ‘‘A CMOS large-swing low-distortion three-stage class AB power amplifier,’’ *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 265–273, Feb. 1990.
- [3] T. Kaneko, Y. Kimura, K. Hirose, M. Miyahara, and A. Matsuzawa, ‘‘A 76-dB-DR 6.8-mW 20-MHz bandwidth CT $\Delta\Sigma$ ADC with a high-linearity Gm-C filter,’’ in *Proc. 42nd Eur. Solid-State Circuits Conf. (ESSCIRC Conf.)*, Sep. 2016, pp. 253–256.
- [4] V. Aparin, G. Brown, and L. E. Larson, ‘‘Linearization of CMOS LNA’s via optimum gate biasing,’’ in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 4, May 2004, pp. IV-748-1–IV-748-51.
- [5] W.-C. Wang and Y.-H. Lin, ‘‘A 118 dB PSRR, 0.00067% (–103.5 dB) THD+N and 3.1 W fully differential class-D audio amplifier with PWM common mode control,’’ *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 2808–2818, Dec. 2016.
- [6] A. Selvakumar, M. Zargham, and A. Liscidini, ‘‘Sub-mW current re-use receiver front-end for wireless sensor network applications,’’ *IEEE J. Solid-State Circuits*, vol. 50, no. 12, pp. 2965–2974, Dec. 2015.
- [7] T. W. Kim, B. Kim, and K. Lee, ‘‘Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors,’’ *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 223–229, Jan. 2004.
- [8] C. Andrews and A. C. Molnar, ‘‘A passive mixer-first receiver with digitally controlled and widely tunable RF interface,’’ *IEEE J. Solid-State Circuits*, vol. 45, no. 12, pp. 2696–2708, Dec. 2010.
- [9] F. Beffa *et al.*, ‘‘A receiver for WCDMA/EDGE mobile phones with inductorless front-end in 65 nm CMOS,’’ in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2011, pp. 370–372.
- [10] W. Sansen, ‘‘Distortion in elementary transistor circuits,’’ *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 46, no. 3, pp. 315–325, Mar. 1999.
- [11] C. Enz, ‘‘An MOS transistor model for RF IC design valid in all regions of operation,’’ *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 1, pp. 342–359, Jan. 2002.
- [12] P. H. Woerlee *et al.*, ‘‘RF-CMOS performance trends,’’ *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, Aug. 2001.

- [13] S. Kang, B. Choi, and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 3, pp. 972–977, Mar. 2003.
- [14] S. C. Blaakmeer, E. A. M. Klumperink, D. M. W. Leenaerts, and B. Nauta, "Wideband balun-LNA with simultaneous output balancing, noise-canceling and distortion-canceling," *IEEE J. Solid-State Circuits*, vol. 43, no. 6, pp. 1341–1350, Jun. 2008.
- [15] W. Cheng, M. S. O. Alink, A. J. Annema, J. A. Croon, and B. Nauta, "RF circuit linearity optimization using a general weak nonlinearity model," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 10, pp. 2340–2353, Oct. 2012.
- [16] P. G. A. Jespers and B. Murmann, "Calculation of MOSFET distortion using the transconductance-to-current ratio (g_m/I_D)," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 529–532.
- [17] C. Enz, F. Chicco, and A. Pezzotta, "Nanoscale MOSFET modeling: Part 1: The simplified EKV model for the design of low-power analog circuits," *IEEE Solid-State Circuits Mag.*, vol. 9, no. 3, pp. 26–35, 2017.
- [18] S.-H. Yang, K.-H. Kim, Y.-H. Kim, Y. You, and K.-R. Cho, "A novel CMOS operational transconductance amplifier based on a mobility compensation technique," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 52, no. 1, pp. 37–42, Jan. 2005.
- [19] C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling: The EKV Model for Low-Power and RF IC Design*. Hoboken, NJ, USA: Wiley, 2006.
- [20] A. Mangla, M.-A. Chalkiadaki, F. Fadhuile, T. Taris, Y. Deval, and C. C. Enz, "Design Methodology for ultra low-power analog circuits using next generation BSIM6 MOSFET compact model," *Microelectron. J.*, vol. 44, no. 7, pp. 570–575, 2013.
- [21] C. Enz, F. Chicco, and A. Pezzotta, "Nanoscale MOSFET modeling: Part 2: Using the inversion coefficient as the primary design parameter," *IEEE Solid-State Circuits Mag.*, vol. 9, no. 4, pp. 73–81, 2017.
- [22] H. Agrawal, S. Khandelwal, J. P. Duarte, Y. S. Chauhan, A. Niknejad, and C. Hu, "BSIM6.1.1 MOSFET compact model," Dept. Elect. Eng. Comput. Sci., Univ. California, San Diego, San Diego, CA, USA, Tech. Rep., 2015.
- [23] T. Taris, J. B. Begueret, and Y. Deval, "A 60 μ W LNA for 2.4 GHz wireless sensors network applications," in *Proc. IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, Jun. 2011, pp. 1–4.
- [24] B. Toole, C. Plett, and M. Cloutier, "RF circuit implications of moderate inversion enhanced linear region in MOSFETs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 2, pp. 319–328, Feb. 2004.



Francesco Chicco (S'16) received the B.Sc. degree in physical engineering from the Polytechnic University of Turin, Italy, in 2013, and the M.Sc. degrees in micro and nanotechnologies for integrated systems from the Polytechnic University of Turin, Italy, the Grenoble Institute of Technology, France, and the École Polytechnique Fédérale de Lausanne, Lausanne (EPFL), Switzerland, in 2015. He is currently pursuing the Ph.D. degree with the Integrated Circuit Laboratory, EPFL. He is involved in the analysis and design of low-power analog and RF integrated circuits for wireless communications.



Alessandro Pezzotta received the M.Sc. and Ph.D. degrees in solid-state physics from the University of Milano-Bicocca, Italy, in 2012 and 2015, respectively. While working toward the Ph.D. degree, he developed a mixed-signal CMOS front-end ASIC devoted to the gas-electron multiplier detector readout, employed in neutron beam monitoring applications. He also participated into the Research and Development activities regarding CMOS continuous-time analog filters and sensors analog frontends. Since 2016, he has been with the Integrated Circuit Laboratory, École Polytechnique Fédérale de Lausanne as a Scientist, involved in analog IC design and semiconductor device modeling, especially in design methodologies based on the inversion coefficient in terms of dc characteristics, linearity, noise, and RF performance. He has authored over 30 scientific papers.



Christian C. Enz (S'83–M'84–SM'11) received the Ph.D. degree from the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 1989. He is currently a Professor with EPFL, where he is also the Director of the Institute of Microengineering and the Head of the Integrated Circuits Laboratory. Until 2013, he was the Vice President of the Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland, where he was the Head of the Integrated and Wireless Systems Division. Prior to joining CSEM, he was a Principal Senior Engineer at Conexant (formerly Rockwell Semiconductor Systems), Newport Beach, CA, USA, where he was responsible for the modeling and characterization of MOS transistors for RF applications. He is the developer of the EKV MOS transistor model. He has co-authored over 250 scientific papers and has contributed to numerous conference presentations. His technical expertise is in low-power analog and RF integrated circuits, wireless sensor networks, and semiconductor device modeling. He is an individual member of the Swiss Academy of Engineering Sciences. He was an elected member of the IEEE Solid-State Circuits Society, AdCom, from 2012 to 2014. He is the Chair of the IEEE SSCS Chapter of Switzerland.