Linearity Analysis of CMOS for RF Application

Sanghoon Kang, Byounggi Choi and Bumman Kim

Department of Electronic and Electrical Engineering and Microwave Application Research Center,
Pohang University of Science and Technology
San 31, Hyoja-dong, Nam-gu, Pohang, Kyungbuk, 790-784, Republic of Korea.
Fax : +82-54-279-2903 e-mail : kangnaru@postech.ac.kr

Abstract— **The linearity of CMOS is analyzed. Transconductance and output conductance are two dominant nonlinear sources of CMOS. Capacitances and substrate leakage network do not generate any significant distortions. But they reduce the output impedance for the best linearity and the power gain at a high frequency and the output conductance nonlinearity is significantly at a high frequency. Up to a few GHz, the output conductance is the dominant nonlinear source, and at a higher frequency, the transconductance is the dominant nonlinearity source. OIP3 is reduced by the effects of those components. OIP3s are calculated for various gate length processes. CMOS linearity is dependent only on current density and drain bias voltage but is not dependent on gate length.**

I. Introduction

There is a tremendous interest in using CMOS technology for RF and microwave applications and many CMOS models for RF application are proposed. But most of the models are focused on small signal characteristics such as small signal gain and noise figure. Recently many researchers have studied about output power, efficiency and coupling to other components, et al. These are well known disadvantages of CMOS process, and various efforts are made to overcome these disadvantages for successful CMOS RFIC implementation. However, linearity is always one of the most important issues for RF circuit.

Some works consider only the transconductance in CMOS linearity analysis and other effects are neglected[1]. For a short gate CMOS, however, the output conductance becomes a more important nonlinear source and it must be considered in CMOS linearity analysis. Capacitances and substrate's effects should be considered at a high frequency.

In this work, we have analyzed CMOS output conductance nonlinearity and have compared it with transconductance nonlinearity. The linearity trend with down-scaled device is also shown in section II. Capacitance effect is discussed at section III. Substrate leakage network is also considered and its effects are analyzed in section IV.

II. Transconductance and Output Conductance Nonlinearities

BSIM3 is widely used in simulation tools, but it requires large data set for accurate modelling and it is very complex. So, it is not suitable for a simple analysis. Taylor series are simple and are widely used for nonlinearity analysis. Drain current in Taylor expansions can be expressed as follows.

$$\begin{aligned} i_{ds}(v_{GS}, v_{DS}) &= I_{DS}(V_{GS}, V_{DS}) \\ &+ G_m v_{gs} + G_d v_{ds} \\ &+ G_{m2} v_{gs}^2 + G_{md} v_{gs} v_{ds} + G_{d2} v_{ds}^2 \\ &+ G_{m3} v_{gs}^3 + G_{m2d} v_{gs}^2 v_{ds} \\ &+ G_{md2} v_{gs} v_{ds}^2 + G_{d3} v_{ds}^3 \\ &+ \cdots \end{aligned} \quad (1)$$

The parameters of this Taylor series can be directly extracted from DC-IV curve but it is hard to extract high order terms and cross modulation terms. These terms can be obtained by a rather simple measurement at a low frequency we have proposed in reference [3].

Assuming that drain is shorted at a signal frequency, all output conductance terms (G_d, G_{d2}, G_{d3} $\cdots$) and cross modulation terms (G_{md}, G_{m2d}, G_{md2} $\cdots$) are vanished and only transconductance terms (G_m, G_{m2}, G_{m3}, $\cdots$) are remained. In this case, the IP3 of gate voltage amplitude given by equation (2) has been used as a device linearity criterion in many previous works.

$$IP3 = \sqrt{\frac{4}{3}\frac{G_m}{G_{m3}}} \quad (2)$$

This model is not accurate enough, however. To compare the output conductance nonlinearity with transconductance nonlinearity, the output nonlinear currents from the components are calculated. For typical CMOS processes, higher order terms and crossmodulation terms in equation(1) are very small and can be ignored to simplify the calculation. In this approximation, the 3rd order intermodulation currents caused by the transconductance and output conductance nonlinearities are given by

$$i_{IM3trans} = \frac{3}{4} G_{m3} A^3 \frac{G_{load}}{G_d + G_{load}} \quad (3)$$

$$i_{IM3cond} = \frac{3}{4} G_{d3} v_{ds}^3 \frac{G_{load}}{G_d + G_{load}} \quad (4)$$

0-7803-7239-5/02/$10.00 © 2002 IEEE

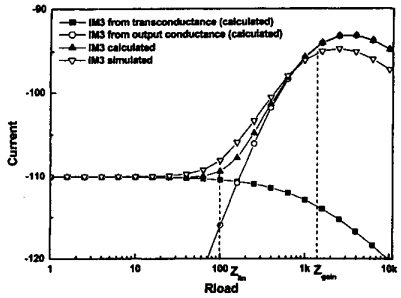

Fig. 1. IM3 currents vs. load impedance

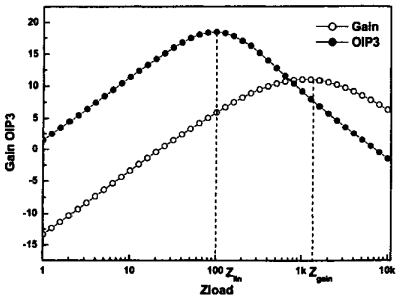

Fig. 2. Gain and OIP3 vs. load impedance

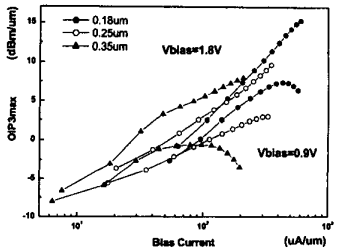

Fig. 3. Linearity of various gate length MOS (TSMC process)

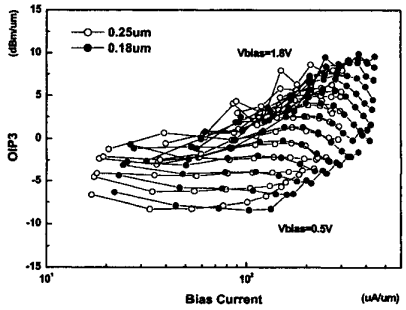

Fig. 4. Linearity of various gate length MOS (Hynix process)

where A is the fundamental voltage amplitude at the gate and v_{ds} is the fundamental voltage at the drain, which can be expressed as

$$v_{ds} = \frac{G_m A}{G_{load} + G_d} \tag{5}$$

TSMC 0.25umx10x10um CMOS BSIM3v3 model is used for the study[4]. All Taylor series parameters are extracted from the DC IV curve of the model. Using these parameters and equations (3)–(5), we calculate the output 3rd order intermodulation currents for various load impedances and compare them with harmonic balance simulation. Source voltage of 2mV (=-50dBm in 50Ω) is selected to ensure a weakly nonlinear operation. The result is shown in figure 1, where the intermodulation currents are shown in dB scale. As shown, the nonlinear currents are dependent on the load impedance. Drain voltage swing is small at a low load impedance and the transconductance nonlinearity dominates the device nonlinearity. At a high load impedance, however, the output conductance dominates the device nonlinearity.

Only with G_m, G_{m3}, G_d and G_{d3} coefficients, the calculation result matches the simulation one very well. When we include G_{md} and other higher components, the calculations predict the simulated results even better. However, their effects are small and can be neglected in saturation region. But in linear region, the cross modulation terms are large and should be included for an accurate calculation.

In figure 2, the output 3rd order intercept point (OIP3) and power gain are shown for a 300 ohm source resistance. As shown, the maximum OIP3 is obtained at a load impedance less than the value for maximum gain (= Z_{gain}). The maximum linearity is achieved at the load impedance where the output conductance nonlinearity increases to the level comparable to the transconductance nonlinearity (= Z_{lin}). Up to Z_{lin}, the fundamental and intermodulation currents (=i_{IM3}) remain the same and their powers increase at the same rate, proportional to the load resistance. Therefore, OIP3 increases. The fundamental current stays constant up to Z_{gain}, but the intermodulation current due to the output conductance nonlinearity (=$i_{IM3cond}$) increases rapidly in this region and OIP3 decreases. Above Z_{gain}, i_{IM3} and power gain both decrease. In RF applications, the load impedance is generally matched to the device output impedance for a maximum power transfer. But if we want a highly linear operation, the output impedance should be reduced from the power matching impedance. It is desirable to select a output load between Z_{lin} and Z_{gain} for good RF performances.

To understand the process dependent linearity trend, we evaluate the maximum OIP3 for TSMC 0.35um, 0.25um, 0.18um processes. Total gate width is 100um with 10 fingers. Because the linearity varies with output load resistance, we calculate the maximum OIP3s and corresponding load impedances for various current levels and drain bias voltages. As the gate length becomes shorter, the transconductance becomes more linear but the output conductance becomes more nonlinear. As a result, the overall linearity

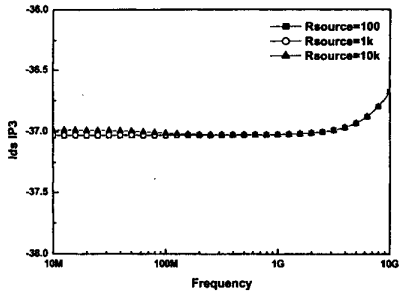

Fig. 5. C_{gs} nonlinearity effect on output current IP3

of CMOS does not change for the gate length variations as shown in figure 3. A shorter gate MOSFET delivers a little better linearity at a high current density, but the improvement is negligible and at lower current density, the same or less linearity is achieved. If IIP3 is considered, a short gate device is more nonlinear at the same input current level.

To verify the above conclusions from measured data, we evaluate 0.18um and 0.25um NMOSs from Hynix. We measure $v_{gs} - i_{ds}$ curve and $v_{ds} - i_{ds}$ curve using Agilent 4155 parameter analyzer, and extract G_m, G_{m3}, G_d, and G_{d3} from the measured DC IV data. To get smooth data from the measurement, we use a data processing program based on MATLAB. Maximum OIP3s are calculated for various current densities and bias voltages using equation (3)(4) and are depicted in figure 4. As shown, the device output linearity depends only on the current density and bias voltage, but does not depend on the gate length. These trends are identical to the simulation results.

III. Intrinsic capacitance effect

Previous experimental result shows that CMOS input impedance is a major source of nonlinearity at a high frequency [2]. The input impedance of CMOS consists of two nonlinear sources, gate-source capacitance and feedback component through gate-drain capacitance. These two components can not be easily separated experimentally and are simulated numerically. The effects of C_{gs} on the input impedance nonlinearities are evaluated for various source impedances. To reduce the effect of C_{dg}, drain is shorted and gate voltage is varied by the various source impedances. The simulated result in figure 5 shows that the gate voltage intermodulation remains constant, i.e., C_{gs} by itself does not degrade the device linearity up to several GHz.

The effect of the nonlinear input impedance can be explained by the gate-drain capacitance. The gate-drain capacitance is generally neglected for a simple linearity analysis, but it forms a feedback loop for a harmonics and disturbs device linearity at a high frequency. We simulate the gate fundamental voltage and intermodulation components for various load impedances. The intermodulation components at the gate are dramatically increased for a large load impedance. When we simulate with C_{gd} zero, the gate voltage intermodulation components are reduced by about 8dB. But the harmonic generation by C_{gd} is minimal compared to the other two sources we montioned. The drain-source capacitance is small and remains almost constant in saturation region, and its effect is negligible. In short, these capacitances do not generate significant harmonics but modify the harmonics by the feedback though C_{dg} and by reducing output impedance. The output IM3 current modified by these capacitances can be obtained from following calculation.

$$v_d = \frac{-G_m}{j\omega C_{gd}(1-\frac{1}{A_v}) + j\omega C_{ds} + G_d + G_{load}} \cdot v_g \quad (6)$$

$$v_g = \frac{1}{1 + Z_s(j\omega C_{gs} + j\omega C_{gd}(1-A_v))} \cdot v_{in} \quad (7)$$

$$A_v = \frac{-G_m + j\omega C_{gd}}{G_{load} + G_d + j\omega C_{ds} + j\omega C_{gd}} \quad (8)$$

Where A_v is a voltage gain from gate to drain. The intermodulation currents are give by

$$|i_{IM3trans}| = \left|\frac{3}{4}\frac{G_{load}}{G_{load}+Y_{out}} G_{m3} v_g^3\right|$$

$$\angle i_{IM3trans} = \angle\left[\frac{G_{load}}{G_{load}+Y_{out}} v_g\right] \quad (9)$$

$$|i_{IM3cond}| = \left|\frac{3}{4}\frac{G_{load}}{G_{load}+Y_{out}} G_{d3} \cdot v_d^3\right|$$

$$\angle i_{IM3cond} = \angle\left[\frac{G_{load}}{G_{load}+Y_{out}} G_{d3} v_d\right] \quad (10)$$

$$Y_{out} = j\omega C_{gd}(1-\frac{1}{A_v}) + j\omega C_{ds} + G_d \quad (11)$$

The drain intermodulation voltage is feedback to the gate through C_{gd} and modifies intermodulation current. This feedback effect can be calculated as

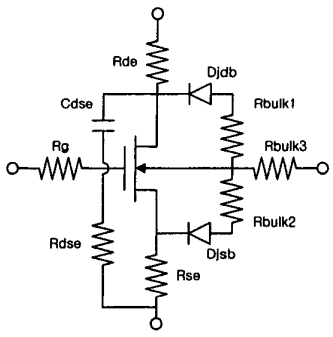

Fig. 6. MOSFET substrate network model

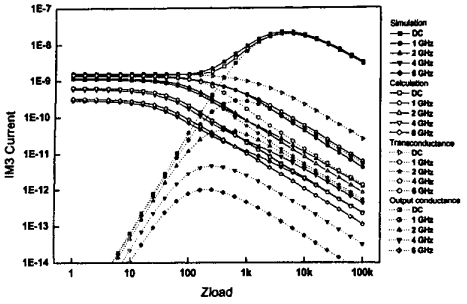

Fig. 7. The 3rd order intermodulation current with substrate network

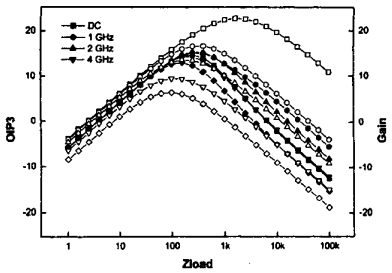

Fig. 8. Gain and OIP3 including substrate network effect

$$i_{IM3transmod} = \frac{i_{IM3trans}}{1 - A_v\beta} \quad (12)$$

$$i_{IM3condmod} = \frac{i_{IM3cond}}{1 - A_v\beta} \quad (13)$$

The IM3 output current is calculated using these simple equations and is compared with simulation for various load impedances up to 6 GHz. The calculated result shows that G_{m3} and G_{d3} are the dominant nonlinear source up to the frequency. At a low frequency, G_{d3} dominates device nonlinearity, but at a high frequency, the transconductance becomes more important. For 0.25um process, the transconductance nonlinearity becomes comparable to the output conductance nonlinearity at about 4 GHz and above that, the transconductance nonlinearity dominates.

IV. Substrate effects on nonlinearity

The conductive substrate has some effects on CMOS linearity and the effects have been studied using the model proposed in reference [4]. In the substrate network, the conductance is considered as a constant value but the source-substrate and drain-substrate diodes are nonlinear components. These diodes are always reverse biased and function as nonlinear capacitances. The substrate model used for our calculation is shown in figure 6.

We modified equation (6) - (13) to consider the substrate effect. Calculated results are compared with harmonic balance simulation results. As shown in figure 7, there are good agreements between the calculated and simulated results. We also include the load dependent nonlinearity contributions from the transconductance and output conductance in figure 7 and the OIP3 and gain curves in figure 8. At a high frequency, Z_{gain} is reduced because of the capacitances, but Z_{lin} remains constant.

The C_{db} adds output capacitance and device gain is reduced. Therefore, the output voltage swing is reduced and the effect of output conductance becomes less important. When the substrate network is included, the cross over frequency where the transconductance nonlinearity becomes larger than the output conductance nonlinearity is reduced to about 2GHz, and OIP3 is significantly reduced. Therefore, at a high frequency,the optimum load impedance for the linearity can be even higher than the power matching impedance.

V. Conclusions

CMOS linearity has been studied and we found that it is not improved with gate length reduction. Only the increases of current density and drain bias voltage improve the output linearity. But short gate length CMOS can operate at a higher current density and so can deliver a higher linearity there.

G_{m3}, and G_{d3} are the dominant nonlinear sources for the frequency band we have studied (up to 6 GHz). And their effects are calculated with a simple equations we have derived. Internal capacitances and substrate network are also included in the calculation. We found that those components are not the sources of nonlinearity but they reduce the high frequency gain and the optimum impedance for low distortions. Therefore, at a high frequency, the transconductance nonlinearity becomes a dominant source. In this case, the OIP3 is reduced significantly. The crossover frequency for a 0.25um MOSFET is around 2 GHz.

Acknowledgments

The authors would like to thank Hynix for supplying device samples. This work is supported in part by Brain Korea 21 project and Hynix.

References

[1] Pierre H. Woerlee, et al., "RF-CMOS Performance Trends," *IEEE Trans. on Electron Devices*, vol.48, no.8, pp.1776-1782, Aug. 2001.

[2] Chengzhou Wang, et al., "A Nonlinear Capacitance Cancellation Technique and its Application to a CMOS Class AB Power Amplifier," *IEEE RFIC Symposium*, pp.39-42, 2001.

[3] Tae Moon Roh, Youngsik Kim, Youngsuk Suh and Bumman Kim, "A new simple extraction method for higher order components of channel current in GaAs MESFET," *27'th European Microwave Conf. Dig.*, Jerusalem, Israel, Sept 1997.

[4] TSMC Co. LTD, *"TSMC 0.25um Mixed-Signal Salicide RF Spice Models,"* Sept. 2000.

[5] Suet Fong Tin, Ashraf A. Osman, Kartikeya Mayaram, and Chenming Hu, "A Simple Subcircuit Extension of the BSIM3v3 Model for CMOS RF Design," *IEEE J. Solid-State Circuits*, vol.35, no.4, pp.612–624, April 2000.