Broadband Linearization Technique for mmWave Circuits

Alok Sethi, Jere Rusanen, Janne P. Aikio, Aarno Pärssinen, Timo Rahkonen
University of Oulu, Finland
{firstname.lastname}@oulu.fi

***Abstract* — This paper presents a broadband linearization technique that can be used for mmWave amplifier circuits. It is based on the well-known principle of derivative superposition, where FETs with different operating points are connected in parallel to generate mutually cancelling third order intermodulation distortion (IM3) products. It is demonstrated by measurements in excess of $10\,\mathrm{dB}$ improvement in IM3 obtained from $1\,\mathrm{GHz}$ to $30\,\mathrm{GHz}$, practically free by connecting a NMOS with very low gate bias in parallel of an amplifying NMOS. The reasons and limits of the cancellation are discussed. The inherent broadbandness of the technique makes it extremely suitable to be used in CMOS mmWave circuits.**

***Keywords* — CMOS, SOI, mmWave, DS, broadband, 5G, 3GPP, NR.**

I. Introduction

Fineline CMOS and SOI-CMOS processes are nowadays commonly used for mmWave design as they have sufficient speed and decent linearity but due to the low supply voltage, the dynamic range is often quite limited. In power amplifiers, the output swing and power is often boosted by stacked amplifier structures. However, in the intermediate amplifier stages, buffers, and phase shifters the available dynamic range remains quite narrow. In this paper, we are demonstrating the significant bandwidth of the linearization that can be achieved by combining parallel devices that have different operating points. Given the large bandwidth and the strict linearity requirements for the upcoming 5G standards [1], this technique can be utilized for meeting those constraints without sacrificing power and silicon space.

It is well known that the value and curvature of nonlinear charge (Q–V) and I–V sources depends on the chosen operating point. This finding has led to a concept of non-linearity cancellation by combining devices at different bias points in parallel. This was originally published under the name of derivative superposition, where operating point of parallel devices have been chosen so that their dominant non-linearities cancel as well as possible [2]. In the most basic form, the non-linearity of just the I–V source I_{ds} is considered. The Taylor series coefficients of the output I–V curve are plotted, and bias points of cancelling cubic terms are searched for. The principle has been since employed both in power amplifiers [3] and in LNAs [4], typically improving the intermodulation distortion (IM3) about $10\,\mathrm{dB}$.

In this paper, we study the bandwidth of the linearization technique of non-linearities cancellation using derivative superposition using two parallel transistors. The structure under study is shown in Fig. 1. There are two equal-size NMOS devices with drains shorted together, and gates separately ac-coupled and biased so that one transistor is biased comfortably in the saturation region and acting as a class A amplifying device, while the other one is biased below the threshold, where its contribution to the overall gain and power dissipation is minimal, but it still manages to generate cancelling IM3 distortion.

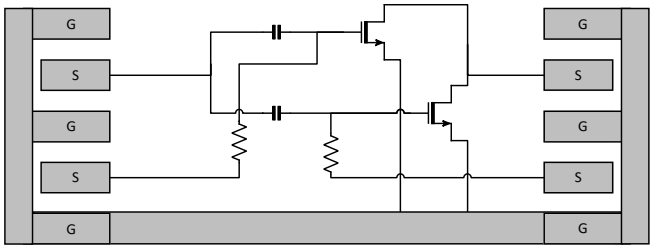

Fig. 1. Circuit diagram.

The organization of the paper is as follows: Section II reviews the mechanism of the distortion cancellation, Section III shows the measured circuit and results, discussion is in Section IV, and conclusions are drawn in Section V.

II. Linearization using parallel devices

The principle of the cancellation can be simply described as below. Let us expand all the non-linear current (I–V) and charge (Q–V) sources to polynomial expansions around the DC operating point, like shown in (1) and (2), respectively, where I_{ds} is the transistor drain current and I_{cgg} is the current corresponding to the rate of change of the total gate charge. Coefficients K_i depend on the I–V or Q–V function shape and operating point, and are often taken as Taylor series coefficients of the current and charge functions. However, running a least-square error fit over the expected signal range usually gives a better estimate with small number of terms [5], [6].

$$I_{ds} = K_0 + K_1 v_{gs} + K_2 v_{gs}^2 + K_3 v_{gs}^3 + ... \quad (1)$$

$$I_{cgg} = \frac{d}{dt}(K_0 + K_1 v_{gs} + K_2 v_{gs}^2 + K_3 v_{gs}^3 + ...) \quad (2)$$

The idea of using the polynomial models is that it allows easy calculation of spectral regrowth: one only builds the linear equivalent circuit, solves the node voltages, and models the non-linear terms as additional current sources in parallel with

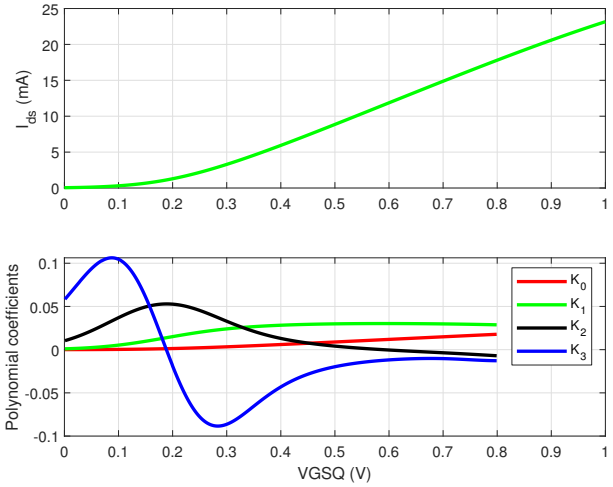

Fig. 2. I_{ds}–V_{gs} and its polynomial model as a function of bias point.

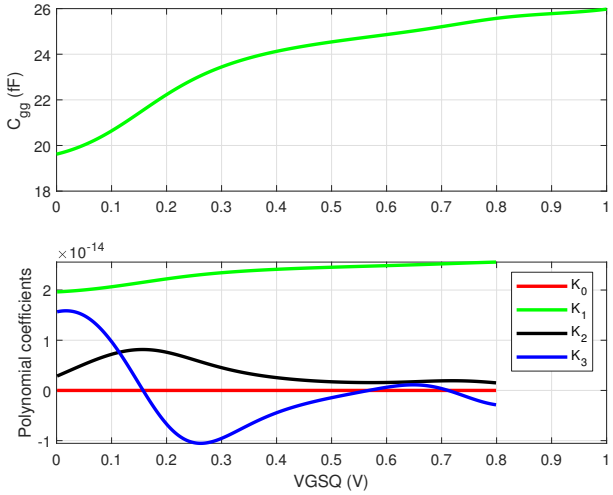

Fig. 3. C_{gg}–V_{gs} and its polynomial model at different bias point.

the linear elements. For equal-amplitude two-tone excitation, the cubic non-linearity causes IM3 currents of magnitude $\frac{3}{4}K_3A^3$ and tone expansion/compression of $\frac{9}{4}K_3A^3$. Here the change in the sign of K_3 causes a $180°$ phase shift, making cancellation possible [7].

Figure 2 shows the simulated I_{ds}–V_{gs} curve at a fixed drain bias, and its polynomial coefficients as a function of the bias point V_{GSQ}. The polynomial is fitted over a $\pm 200\,\mathrm{mV}$ signal range (with 1V maximum supply, V_{GSQ}=800 mV is the highest quiescent bias point where the fit can be made). If we assume that only the cubic non-linearity matters, we notice that the device shows compressing behavior ($K_3 < 0$) at high gate bias but expansive behavior ($K_3 > 0$) near the threshold voltage, where the I–V curve is almost exponential in shape. Hence, it is possible to choose two bias points with opposite phase and thus cancelling K_3 values.

The cubic non-linearity K_3v^3 in the I–V response results directly into IM3 output current, and this effect is broadband and insensitive to passive circuits. Also, the quadratic non-linearity K_2v^2 can result into a third order non-linearity via multiple mixing. A square-law shape generates second harmonics and DC components, but then also operates as a square-law mixer for the fundamental and the generated second order tones. Hence, a second harmonic in the output current can mix with the fundamental in the second degree curvature of output impedance, for example, and result in additional IM3 contribution.

The case is more complicated for non-linear input impedance. Linear term affects only the gain and matching. A K_3v^3 term will generate an IM3 at the input side, from where it propagates to the output. K_2v^2 term generates second harmonics and baseband terms that can further mix to IM3 either in the same quadratic non-linearity in the input impedance, in the second order non-linearity of the I_{ds}–V_{gs} source, or in the non-linearity of the output impedance. In each node, there may also be filtering involved: second order signals are generated as currents and get multiplied by the impedance seen at those frequencies. Hence, filtering the second harmonic or the DC band voltages also impacts the IM3.

Figure 3 shows the expansion of total gate capacitance. This is relatively linear and, moreover, the gate impedance in this setup is low, keeping the current-to-voltage gain small. According to the simulations, any distortion components at the gate were about $20\,\mathrm{dB}$ lower than in the output, making their effect negligible.

III. Measuring Setup and Results

To verify the cancellation operation, the circuit shown in Fig. 1 was fabricated using $45\,\mathrm{nm}$ CMOS SOI technology. Its micrograph is shown in Fig. 4 and the measurement setup for providing common drain and separate gate biases is shown in Fig. 5. The measurement setup consists of a vector network analyzer, probe station, differential probes, bias tee, combiner, and programmable power supply. One signal pad of each probe is used to provide bias to the gate of the transistors. The other signal pad is used for the RF signal. Bias to the common drain is provided via a bias tee. No calibration was applied.

Figure 6 shows the measured IM3 levels of a two tone test for a fixed $-6\,\mathrm{dBm}$ input power. The amplifying device has a gate bias of 0.5, 0.6, and $0.7\,\mathrm{V}$, respectively (red, green, and blue curves), and the bias V_{GSQ} of the cancelling device is swept from $0\,\mathrm{V}$ upwards. Hence, the right end of the curves show a situation corresponding to a $2\times$ size device with the given 0.5, 0.6, or 0.7 V bias, and the left side of the curves depict a case where one transistor is biased to provide amplification and the other transistor is at a much lower bias. All curves show about $10\,\mathrm{dB}$ improvement in the IM3 level, when the bias of the cancelling device is about $150\,\mathrm{mV}$. Thus, this is roughly the amount of linearization achievable.

The most important finding is that when the measurement is repeated at the center frequencies of 1, 10, 20, and 30 GHz, the cancellation performance remains essentially constant (measurement at $30\,\mathrm{GHz}$ is already noisy due to higher attenuation, but the shape is still similar). Hence, the

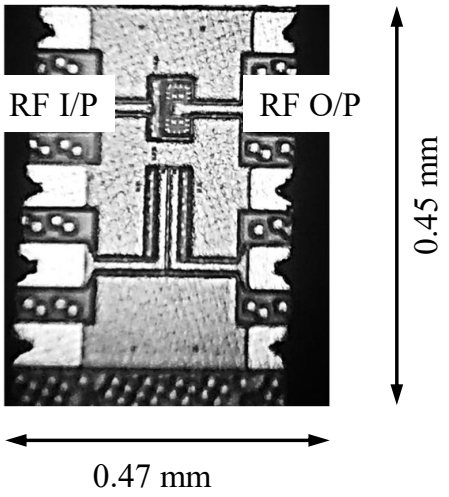

Fig. 4. Chip photograph of the fabricated structure.

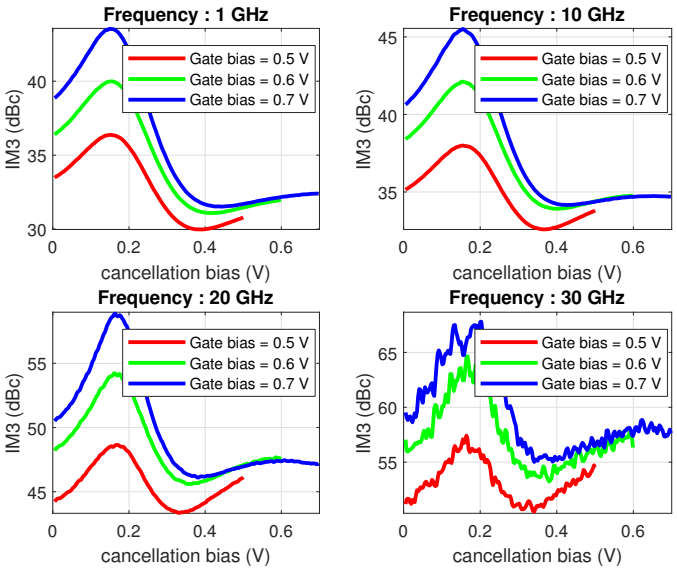

Fig. 6. Measured IM3 at a fixed source power.

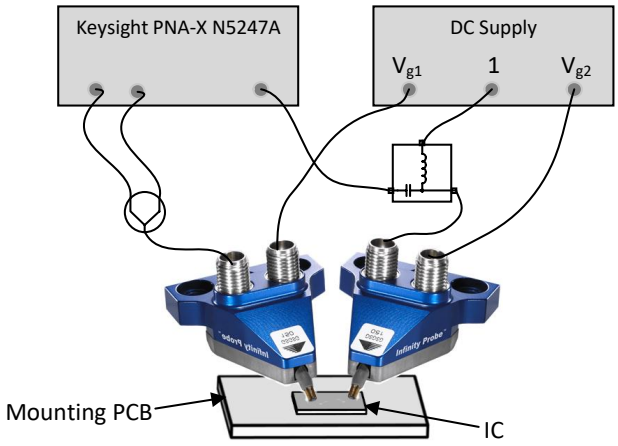

Fig. 5. Measurement Setup.

achieved improvement in linearity is extremely broadband. This is because the overall IM3 is dominantly generated by cubic non-linearity. Hence, it is not affected by intermediate filtering around second harmonic region, for example.

IV. Discussion

Our measurement equipment was limited to 30 GHz and the achieved cancellation looks practically constant throughout the entire measured frequency range. To explain why it is so broadband, whether this is the usual case, and if we can trust our simulations on this phenomenon, we also need to look at the simulation results. Let us discuss the modeling issues first.

The main amplifier is biased in normal class A mode, therefore, its modeling is assumed as good as it can be. The cancelling device, however, is operating at low bias in a region seldom used, and also at a frequency where it can not provide much gain any more. This raises two questions: does the shape of the simulated I–V curve correspond with the reality and how does the device operate at frequency of the order of its f_{MAX}?

The I–V curve shape question apparently is not solved very precisely. The simulations predicted up to 30 dB of improvement in IM3 over a narrow bias range of 20 mV centered around a cancellation bias of 45 mV, multiple cancellation peaks depending on the magnitude of the excitation signal, and change in the magnitude of the required V_{GSQ} for the cancelling transistor with the frequency of the excitation signal. In measurements, we consistently had only one optimum and about 10 dB cancellation as seen from Fig. 6.

Regarding the high-frequency behavior of the cancelling device, let us first look at the simulated phases of the fundamental and IM3 currents generated by the devices. The amplifying device behaves like a normal class A amplifier. The cancelling device shows weak fundamental gain with a phase shift, because a part of the fundamental current is caused by capacitive feed-through via C_{dg} as seen in Fig. 7. The IM3 current generated by the cancelling device is comparable to the IM3 current generated by the amplifying device and the phase difference between them is approximately 180° as seen in Fig. 8.

It is well known that lumped device models tend to fail near the f_{MAX} of the device. Here, the f_{MAX} drops from 196 GHz at V_{GSQ} of 500 mV to 70 GHz at V_{GSQ} of 150 mV. The device model for the cancelling transistor is operating at its maximum potential, thus, some problems can be expected. It is believed that, for example, non-quasi static modeling begins to have big effects here, but according to the measurements it is safe to say that while the cancelling device is not providing gain for the fundamental current, it still generates distortion that is not suffering large phase shifts, which maintains the possibility of cancellation in the summing node.

In addition to the above device properties, the bandwidth of the cancellation is significantly affected by where the distortion is generated and how. The cubic non-linearities generate the IM3 in one step, without any intermediate mixing mechanisms,

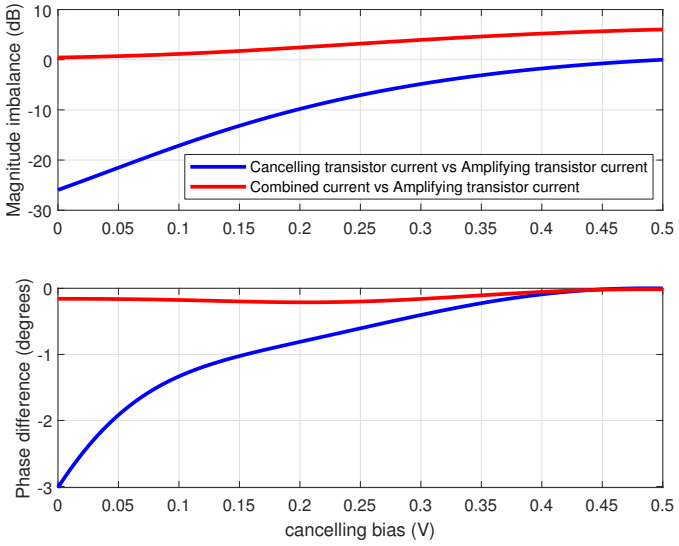

Fig. 7. Simulated amplitude and phase imbalance between the fundamental currents at 1 GHz.V_{GSQ} of the amplifying transistor is set to 500 mV

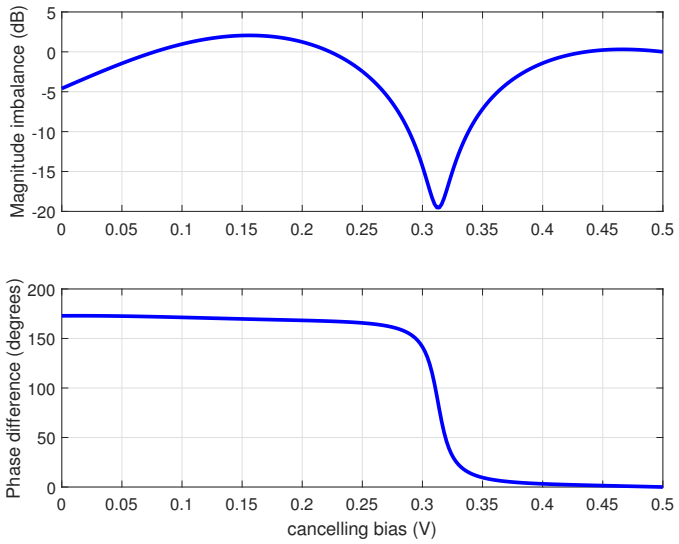

Fig. 8. Simulated amplitude and phase imbalance between the IM3 current of the cancelling and the amplifying transistors. Measured at the individual transistor drains with V_{GSQ} of the amplifying transistor set to 500 mV.

where as, IM3 components from second order non-linearities are generated via multiple out-of-band mixing products and thus, their frequency response will be seen in the resulting IM3, which will yield narrow band cancellation. Here, the multiple second order mixing are not strong enough to be seen. If the gate drive impedance is increased, they can also become visible, and thus, the frequency response of the input match will affect the bandwidth of the cancellation.

The measurements above were made by having two equal-sized devices in parallel. As the dimensioning criterion is to achieve equal K_3v^3 with opposite signs, and since the non-linearity in the I–V curve is the primary contributor to IM3 here, the high positive peak of K_3 in Fig. 2 suggests that the size of the cancelling device can be reduced down to about 25% of the amplifying device. This would noticeably reduce the input and output capacitances.

V. Conclusions

The use of derivative superposition principle in linearization is well known and used in many documented circuits. In this paper, we illustrated the inherent broadbandness of the technique, about 10 dB improvement in IM3 was measured over a frequency range of 1 GHz–30 GHz. The upper frequency limit on the measurement comes from the available measurement setup and not from the circuit. This improvement in linearity was achieved because the output non-linearity was mainly caused by the memoryless cubic shape of I–V characteristics. Multiple mixing of quadratic non-linearities, on the other hand, is prone to the harmonic filtering of the intermediate nodes and will result in narrow band cancellation. Furthermore, as the cancelling transistor is operating in a region which is close to its f_{MAX}, the simulations are not completely reliable and the shape and cancellation may be worse than in simulations.

Acknowledgment

The research leading to these results has received funding from the Nokia Corporation Ltd. The authors would like to acknowledge Global Foundries for the silicon processing and technical support. This research has been also in part financially supported by Academy of Finland 6Genesis Flagship (grant 318927) and UniOGS Oulu doctoral school.

References

[1] 3GPP, "Base Station (BS) radio transmission and reception," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.104, 12 2018, version 15.4.0. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/38_series/38.104/38104-f40.zip

[2] D. Webster, J. Scott, and D. Haigh, "Control of circuit distortion by the derivative superposition method [mmic amplifier]," *IEEE Microwave and Guided Wave Letters*, vol. 6, no. 3, pp. 123–125, March 1996.

[3] S. Ock, K. Han, J.-R. Lee, and B. Kim, "A modified cascode type low noise amplifier using dual common source transistors," in *2002 IEEE MTT-S International Microwave Symposium Digest (Cat. No.02CH37278)*, vol. 3, June 2002, pp. 1423–1426 vol.3.

[4] J. R. Gajadharsing, "Low distortion rf-ldmos power transistor for wireless communications base station applications," in *IEEE MTT-S International Microwave Symposium Digest, 2003*, vol. 3, June 2003, pp. 1563–1566 vol.3.

[5] J. P. Aikio and T. Rahkonen, "Detailed distortion analysis technique based on simulated large-signal voltage and current spectra," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 10, pp. 3057–3066, Oct 2005.

[6] J. P. Aikio, T. Rahkonen, and J. C. Pedro, "Extraction of a multi-dimensional polynomial device model for an improved distortion contribution analysis technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 1, pp. 155–164, Jan 2015.

[7] S. Maas, *Nonlinear Microwave and RF Circuits*. Artech House, 2003. [Online]. Available: https://books.google.fi/books?id=SSw6gWLG-d4C