



Origins and minimization of intermodulation distortion in a pseudo-differential CMOS beamforming receiver

Negar Shabanzadeh¹ · Rehman Akbar¹ · Aarno Pärssinen¹ · Timo Rahkonen¹

Received: 2 April 2021 / Revised: 7 July 2021 / Accepted: 11 July 2021 / Published online: 25 July 2021
© The Author(s) 2021

Abstract

This paper studies how nonlinear distortion is generated in the combination of an inverter-based low-noise amplifier and a passive mixer. The dominant nonlinearity appears to be the quadratic $V_{gs}V_{ds}$ mixing term in the passive mixer that first causes low-frequency IM2 and then upconverts it to IM3. Adding a common-mode feedback (CMFB) cancels the IM2 in a pseudo-differential structure, and hence also reduces the IM3 caused by the cascaded second order nonlinearities significantly. The effect of CMFB gain, bandwidth and linearity were analyzed, and it is concluded that from the linearity point of view, the feedback circuit does not have to be very wideband since the dominant distortion products originate from baseband. Finally, the paper takes a look at the spurious tones rising in the mixing, and how to extend the analysis to include the actual frequency translation effect.

Keywords Nonlinearity analysis · Volterra analysis · Distortion contribution · Common-mode feedback

1 Introduction

The current trend of ever-increasing data rates has imposed the need for millimeter-wave range communication systems, which in turn require parallelism as a means of providing filtering, beam-steering, and most importantly keeping up with the data rate speed and improving the overall throughput. Much interest has been devoted to analyzing the linear properties (gain, noise, beam pattern) of highly parallel systems, but studying their nonlinearity has been less frequent [1, 2]. This paper focuses on the nonlinearity analysis of the circuit blocks in the 3 GHz IF section of a beamforming receiver architecture proposed in [2, 3].

Inverter-based LNAs are quite common in literature. Applying the Cherry–Hooper structure [4] and splitting-load inductive peaking [5] are popular methods for extending the bandwidth of the LNA stage. Some distortion cancellation methodologies have also been reported, e.g. the use of active [6] or passive [7] feedback to reduce IM2 and hence IM3. Also, envelope-dependent adaptive biasing

has been used to minimize power dissipation in ISM applications [8].

This work studies the linearity properties of a 3 GHz receiver IC shown in Fig. 1. The implementation is inductorless due to the multitude of channels and consists of basic inverter-based LNAs, passive mixers, phase control switches, and eventually a summing point in a baseband transimpedance amplifier. The LNA structure is chosen to be as simple as possible to keep the circuit compact and still give sufficient bandwidth. The circuit structures for analysis have been taken from [3], but due to limitations in technology access, the device characteristics and comparisons to simulations have been performed using a different IC process. Hence, the performance is not an exact match to [3].

The linearization of receiver front-ends, particularly LNAs, has been the focus of several papers, from calculative approaches to minimize distortion [9] to newer distortion analysis methodologies [10–13] and circuit-based techniques [14–18]. Chen et al. [9] utilizes the calculation of Volterra kernels to achieve the criteria for IM2 cancellation and prevent it from mixing into IM3 in a broadband inductorless LNA. Deriving the Volterra expressions by hand analysis is still popular, but unfortunately very complicated and time-consuming. Solving the higher-order

✉ Negar Shabanzadeh
Negar.Shabanzadeh@oulu.fi

¹ University of Oulu, Oulu, Finland

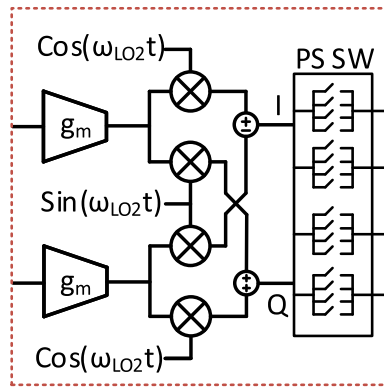


Fig. 1 Block diagram of one slice of the beam-forming receiver under analysis

derivatives needed in the series expansion is also difficult, unless the device models are simplified. Simpler Volterra-based approaches have been proposed, such as splitting the nonlinear transistors to a linear component with a nonlinear drain current source, namely per-distortion nonlinearity analysis [10], combining Volterra and multisine analyses [11], and harmonic distortion analysis in feedback amplifiers [12]. Li and Pileggi [10] manages to decrease complexity by lumping all the contributions from one device together, but cannot separately identify their causes. Boremans et al. [11] gives better insight on the distortion contributions and is able to deal with strong nonlinearities unlike [10] but is still somewhat complex. Palumbo and Pennisi [12] is a block-diagram analysis based on transfer functions, providing analytical insight into the frequency behavior of the related closed-loop distortion factors, but no formulation of $P_{1\text{-dB}}$ and intermodulation distortions [13]. In [13], a general weak nonlinearity model is proposed for LNAs as an effort to facilitate design automation and avoid repeating the analysis for each topology. The model is still specific to LNAs and yields closed-form expressions for the circuit IM2 and IM3 without much discussion on their originating mechanisms. There are also many papers that have come up with circuit innovations to cancel distortion with less emphasis on the analytical part. The most widespread of these include optimizing the overdrive voltage [14], using pre-distortion [15] or post-distortion networks [16, 17], and derivative superposition methods [18]. As for the mixer part, since passive mixers are more linear, fewer studies have been devoted to linearizing them. However, general linearity analysis of passive mixers has been investigated in studies such as [19, 20].

In this paper, the nonlinear current injection method introduced by [21, 22] and extended in [23] is used. To handle the complexity, the results are calculated numerically and visualized as vector contributions. The approach

used in this paper differs from the above in a few ways. First, the products of both capacitive and conductive nonlinear effects are included, and nonlinearities can be separated by origin. Moreover, band-to-band mixing mechanisms are accounted for, specifically the mixing from the second order nonlinearities to cubic terms which is of interest here. Finally, instead of analytical high order derivatives, numerical fitting is used to build the polynomial models. To the authors' knowledge, an analysis like this (building the distortion contributions systematically using a software tool based on an interpretation of Volterra analysis) on a receiver branch has been done for the first time in literature. The same approach has been employed previously on simpler circuits with a smaller number of weakly nonlinear components [23–25]. We are taking it a step further by utilizing it in an LNA-feedback-mixer combination, tackling different levels of nonlinearity at the same time.

The paper is organized as follows: Sect. 2 gives an overview of the analysis technique. Sections 3 and 4 introduce the modeling and analysis of a single transconductance stage respectively, and Sect. 5 includes the passive mixer in the analysis. The effects of the common-mode feedback are studied in parallel in Sects. 4 and 5. CMFB gain and bandwidth considerations are investigated in Sect. 6. Including the frequency translation effects is discussed in Sect. 7, and finally, a discussion is given in Sect. 8.

2 Distortion contribution analysis

The use of polynomial models for calculating frequency translations and employing nonlinear current injection to model the nonlinear effects are well known [21]. In our previous research, we have extended it so that not only the total nonlinearity is given, but any tone phasor can also be plotted as a sum of contributions per device, per device's nonlinear source, and per mixing between the harmonic bands. In this study, we used the MATLAB tool NLSim [24]. To show all the contributions listed above, it needs the device polynomial modeling to be fitted offline based on the DC sweeps of currents and capacitance values.

The nonlinear current sources need to have several controlling ac voltages. For example, the drain current model of a MOS device is written as (1):

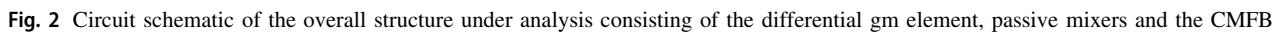
$$i_d = \sum K_{ij} v_{gs}^i v_{ds}^j \quad (1)$$

where i and j are the degree of the nonlinearity taken into account. The coefficients K_{ij} can be found in several ways. Here, a simple least-square error fit has been used over a

Results can be interpreted by keeping in mind that the total distortion in a given node is first plotted as contributions from different devices, and then the results are further plotted as functions of nonlinearity order and harmonic-to-harmonic mixing per device. For example, IM3 may be generated due to the cubic nonlinearity K_{30} , or the quadratic mixing in K_{20} between the fundamental and IM2 in baseband and the second harmonic band [25]. This will be further elaborated in the following sections.

The overall structure under analysis is shown in Fig. 2. M1–M4 form the transconductance stage, M5–M8 are passive mixer transistors, and M9–M10 provide the common-mode feedback. We will start the analysis with the

As an example, Fig. 3 shows the input voltage related coefficients K_{10} , K_{20} , K_{30} of a CMOS inverter-based gm element in terms of the operating point. Curves are calculated for similar-shaped nMOS and pMOSes and at several input amplitude values (increasing the amplitude makes the peaks flatter). It is seen that the nonlinearity is highest around the MOS threshold voltages and minimized in the center. K_{10} (the linear gain) doubles when both devices are on. Summing up the two transistors' currents causes several minima to appear (zero-crossings are marked with o's in the figure), which keeps K_{20} and K_{30} small in the mid-bias range. Therefore, the bias can be chosen anywhere in the range $V_{in} = 0.4 \dots 0.6$, where either a distortion minimum or possible cancellations between two similar but differently biased branches can be achieved.



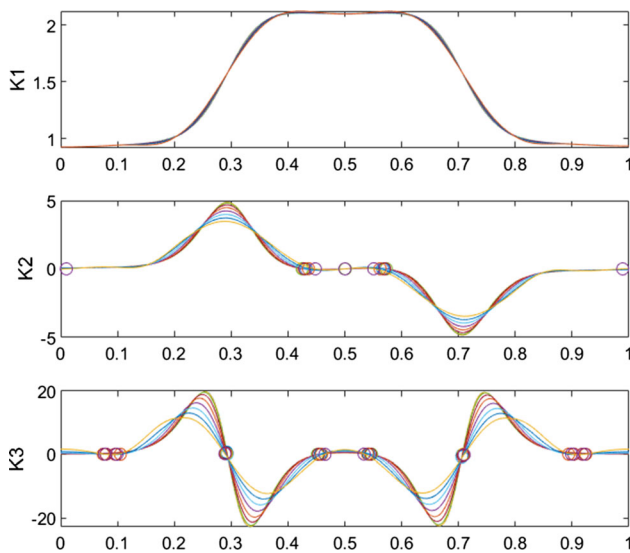


Fig. 3 Input-related nonlinearity coefficients versus input bias point of a CMOS inverter gm stage

4 Transconductance stage nonlinearity analysis

The simulations illustrate that the amplifier itself is very linear. A 10 mV input was given in the two-tone test simulation setup, and the signal magnitudes at the gate and drain were -49 dBV and -32 dBV (all spectral measurements are from a two-sided spectra, so the values should be multiplied by two to get the amplitudes). The IM3 voltage in the output node was -124.2 dBV. The use of common-mode feedback (CMFB) did not affect IM3, but it did reduce the IM2 from -80 to -89.8 dBV. This is due to the fact that even order nonlinearities appear as a common mode signal in differential structures, so the CMFB senses and attenuates them proportionally to the loop gain. In the case of the standalone gm, both IM3 and IM2 are already very low, and the effect of the CMFB will be clearer in the combination of a gm and mixer.

The improvements in nonlinearity are also easily seen by traditional numerical harmonic balance or pss simulation. However, the advantage of polynomial modeling is that it enables us to see which nonlinearity contribution is dominant, and how the inherent cancellation mechanisms are formed. This is illustrated by a vector plot where any higher order IM tone can be plotted, as shown in Fig. 4. Reading the vector plots requires following some notations: The vector “Total” illustrates the magnitude and phase of the chosen IM tone similar to what we see in e.g. harmonic balance simulation, while the other vectors demonstrate how it is built from contributions of different nonlinearities. The contributions are labeled with their source (e.g. gmm3 is the drain current response of transistor M3) and the degree of nonlinearity [K_{ij} refers to term

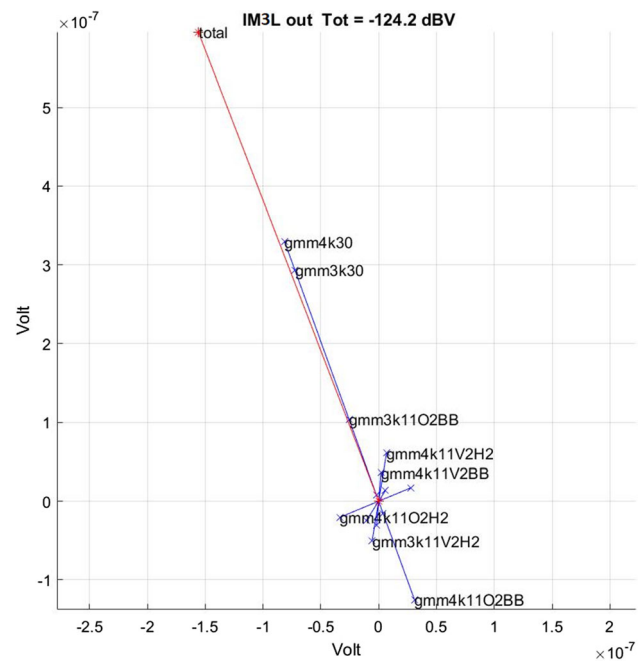


Fig. 4 Vector plots for IM3 voltage in the standalone differential gm stage

$K_{ij}v_{gs}^i v_{ds}^j$ in the I–V Eq. (1)]. Finally, the last label indicates from which harmonic band the mixing is happening, such as baseband or second harmonic. For better visibility, the contributions smaller than a certain value are not plotted.

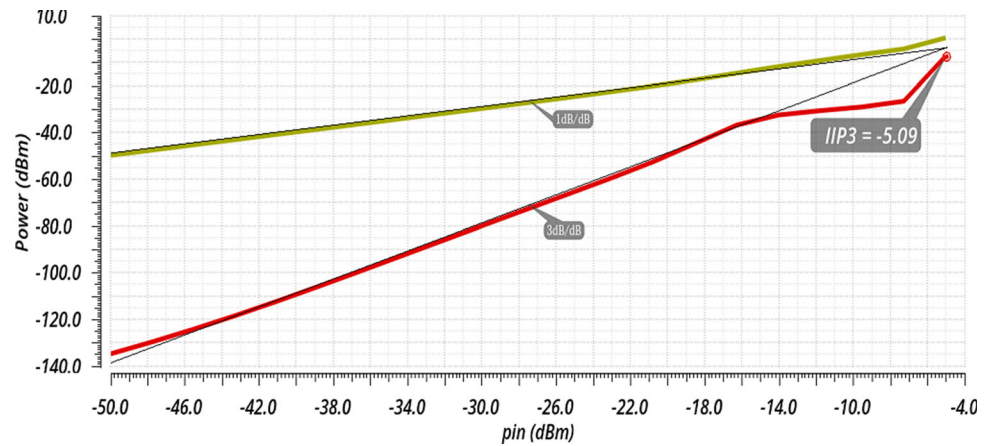
According to Fig. 4, the cubic v_{gs} -related nonlinearities (the K_{30} terms) of transistors M3 and M4 clearly dominate IM3 in the output node. There are also some cascaded second-order mixing results (resulting from baseband (BB) or second harmonic (H2) band signals mixing to IM3). Although they are small in value, an interesting conclusion can be derived from the fact that the effects of nMOS and pMOS (gmm3k11O2BB and gm4k11O2BB) are cancelling each other. Since Early voltage is modeled with the $K_{11}v_{gs}v_{ds}$ term, this cancellation suggests that a low Early voltage does not degrade the linearity of an inverter stage, i.e. the inverter linearity is not very sensitive to Early voltage variations.

For comparison, Fig. 5 shows the results from pss analysis in Cadence Spectre corresponding to Fig. 4. It gives an input-referred third-order intercept point (IIP3) of -5.09 dB, which agrees with the value from the MATLAB simulations.

5 Mixer stage nonlinearity analysis

The next stage to be studied is the mixer block. A passive mixer structure is commonly used due to its high linearity, high voltage headroom, low power consumption and good

Fig. 5 IIP3 value of the gm-cmfb combination from Cadence pss simulation



1/f noise performance. We start the nonlinearity analysis at its simplest with a single transistor biased in the triode region and acting as a pass switch, which is *continuously on* during the analysis. The procedure here resembles the one in [20], but is done numerically using the nonlinear current injection technique instead of closed-form kernel derivations. Also, we are including the effect of the common-mode feedback.

The DC fitting is done through the same procedure as described in Sect. 2, but now the device is in the triode region. Thus, the dominant nonlinearity in the nMOS switch is the fact that the voltage v_{in} directly modulates the gate-source voltage and on resistance. Expanding the classical second order I–V equation $I_D = \beta(V_{GS} - V_T - V_{DS}/2)V_{DS}$, we obtain an ac model of the form (2):

$$i = \beta V_{on} v_{ds} - (\beta/2) v_{ds}^2 - \beta v_{ds} v_{in} = K_{01} v_{ds} + K_{02} v_{ds}^2 + K_{11} v_{gs} v_{ds} \quad (2)$$

where the last K_{11} term is usually dominant. The device also produces another second order mixing (K_{02}), which is not the desired RF mixing but a nonlinear effect that appears while the switch is conducting. The actual mixing effect is studied later in Sect. 7.

The first analysis case consists of a single gm and mixer pass-transistor without the common-mode feedback. As seen in Fig. 2, the simple single-transistor mixer has been extended to a balanced structure now. The input is again a 10 mV two-tone, giving -32 dBV in the output of the gm stage. Now, the IM3 tones are some 40 dB higher than the case of a standalone gm element. To see what causes this, we need to look at the details of Fig. 6, which shows the IM3 contributions in the output of the gm element at node A in Fig. 2.

The vector plot shows all contributions from all the eight nonlinear devices M1–M8, and the dominant contribution seems to be the mixer transistor M8, where both the

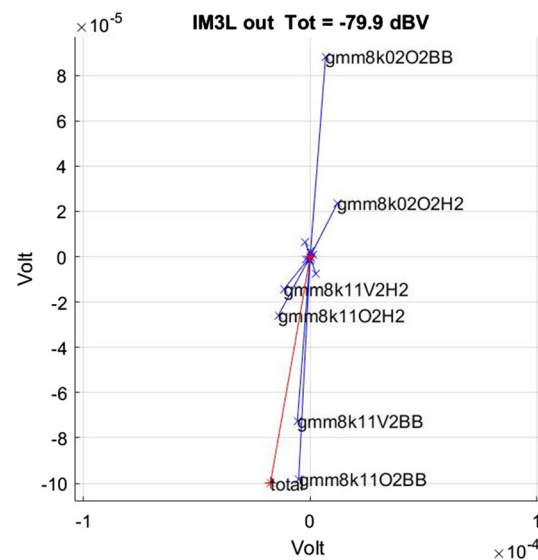


Fig. 6 Vector plot for IM3 output voltage in the gm-mixer combination without common-mode feedback

Ids curvature ($K_{02} v_{ds}^2$) and the modulation of the on-resistance ($K_{11} v_{gs} v_{ds}$) show strong but partially compensating contributions. These effects are again cascaded effects, generating IM2 voltages in the nodes around the mixer, and then mixing it further to IM3. Hence, these contributions can be reduced by filtering away the generated IM2 tones (baseband and second harmonic bands) at the drain and source terminals of the mixer transistor. Alternatively, as filtering consumes too much area, one can reduce the switch on resistance to minimize the v_{ds} magnitude. For instance, Akbar et al. [3] achieves this by employing LO bootstrapping to increase the gate drive.

The effect of the common-mode feedback is illustrated in Fig. 7, neglecting the nonlinearity of the CMFB transistors. Now, the feedback loop senses the generated IM2 and reduces its level from -55 to -79 dBV. Now that there is less signal to mix with the fundamental tones, the IM3 reduces from -80 to -120 dBV. The quadratic

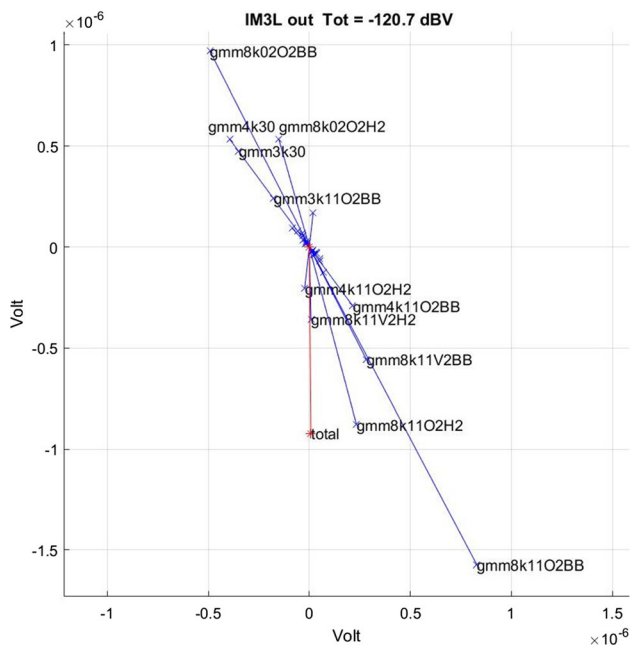
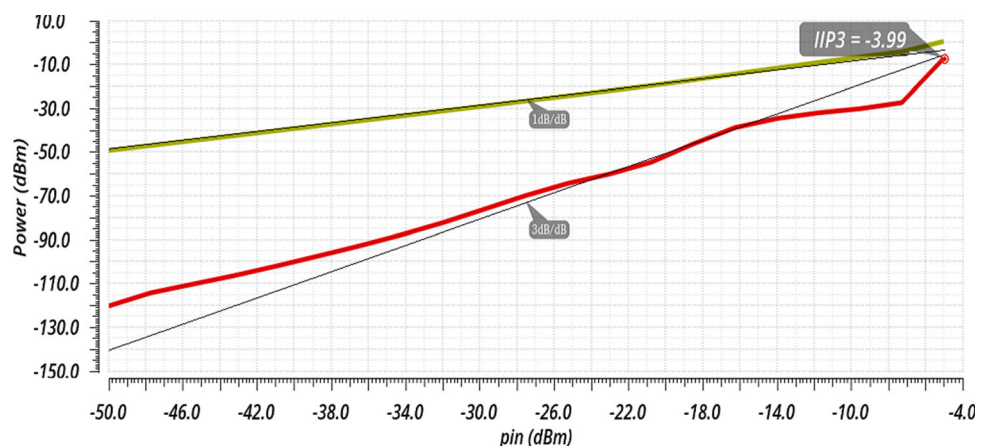


Fig. 7 Vector plots for IM3 nonlinear voltages in the gm-mixer combination with the common-mode feedback

effects are way smaller in this case, and the cubic input-related nonlinearities of the gm transistors M3 and M4 (*gmm3k30* and *gmm4k30*) are visible again in Fig. 7. Yet, the overall IM3 is still dominated by second-order mixing in M8 (terms *gmm8k11OBB* and *gm8k11OH2*), which is harvesting IM2 tones from both the baseband and second harmonic band.

Figure 8 demonstrates the IIP3 result of the pss simulation corresponding to the gm-mixer combination with the common-mode feedback, which is again consistent with the MATLAB results.

Fig. 8 IIP3 value of the gm-mixer-CMFB combination from Cadence pss simulation



6 The effects of CMFB bandwidth and linearity

The choice of gain and bandwidth of the common-mode feedback circuit is the next thing to be considered. The CM-correcting pMOS devices function as parallel loads for the CM signal, and to be effective, the total common mode impedance they introduce should be noticeably smaller than the existing common-mode output impedance between the two transconductance and mixer stages. Hence, we can either make the pMOSes smaller and compensate that by increasing the gain or vice versa. There is not much difference between these two choices considering the linear signal only, although big devices change the biasing and loading of the inverter stage (which is why keeping them small is preferred). However, the situation is more complex from distortion point of view. In addition to the linear common mode signal, any second order distortion in the amplifier and mixer is also seen as a common mode signal, and attenuating them is beneficial as seen above. Yet, there are two second order bands: the rectified signal near DC, and the second harmonic band. This is where the bandwidth of the CMFB circuit comes into play.

From the linear signal point of view, the most narrow-band solution is to correct just the dc offsets. However, if we aim at correcting amplitude mismatches e.g. due to a driving balun, the CMFB loop should have a bandwidth exceeding the carrier frequency. Therefore, we consider three different bandwidths which include only the beat frequency (100 kHz), the carrier frequency (3 GHz), and the second harmonic band (6 GHz). The beat frequency is chosen arbitrarily, and as the main circuit in [3] supports a baseband bandwidth of over 400 MHz, it can also be pushed further. Results are presented in Table 1 assuming both opposite-phased (Diff) and in-phase (CM) inputs, and a third case where a 10% amplitude mismatch is introduced between the differential inputs. The differential input is how the circuit is actually used, CM input is only shown to

Table 1 The effect of CMFB bandwidth on distortion results

Input	CMFB 3 dB BW	Linear signal (dB)	IM2 (dB)	IM3 (dB)
Diff	100 kHz		− 78	− 118.8
	3 GHz	− 33.4	− 77.7	− 121.3
	6 GHz		− 77.7	− 121.3
CM	100 kHz	− 32.4	− 75.4	− 112
	3 GHz	− 32.5	− 75.7	− 115.7
	6 GHz	− 32.8	− 76.3	− 117.7
Asymm	100 kHz		− 79.9	− 122.2
	3 GHz	− 34.4	− 79.5	− 124.6
	6 GHz		− 79.5	− 124.4

illustrate the common-mode rejection, and the last case illustrates a more practical case, as presuming perfect symmetry in simulations often hides some of the effects seen in reality. The op-amp gain is 25 dB in all examples, but since the gain is interdependent on the pMOS nonlinearities, we will get back to its effects once they are included.

Table 1 suggests that increasing the CMFB bandwidth from 100 kHz to 3 GHz does not have a significant impact on the distortion cancellation, and the 3 GHz and 6 GHz cases are almost similar. To elaborate this, let's take another look at Figs. 6 and 7: as seen in Fig. 6, the dominant cause of IM3 is the up-conversion of the baseband IM2 result *gmm8k11O2BB*, while the second harmonic products (ending with H2) are smaller and already cancelling each other to some extent. The same pattern is seen in Fig. 7 as well. So, even a narrowband CMFB can yield a considerable improvement in IM3. Although further increase in bandwidth helps to reduce the second order tones in the second harmonic band, it does not matter much if the bandwidth covers fundamental or second harmonics; the loop gain is sufficient in either case. This means that the CMFB circuit in this setup does not have to be very complex or capable of working at high frequencies to give satisfying results, as the improvement from Figs. 6 and 7 with the narrow-band op-amp CMFB in Fig. 2 also confirms.

The above discussion emphasizes the linearity of the feedback devices itself, as any nonlinearity originated in the feedback loop input is not attenuated by the loop. We have assumed completely linear devices in the CMFB loop in the previous sections, but it would be interesting to see if the nonlinearity of the controlling pMOSes has a significant impact on the results and how it can be relaxed by the choice of gain and bandwidth. Since the signal amplified by the loop drives the feedback devices, the combination of high gain and small devices (which was our design choice

for the linear response) means that the nonlinearity will also be amplified. So, if this signal is large enough, the feedback loop must be strictly linear to maintain its positive effect. Luckily, the mentioned signal is not the input RF signal, but just the small common-mode error in it, in addition to small IM2 components. Hence, the nonlinearity effect of the CMFB amplifier is mitigated to the point that it does not cause problems. The three cases given in Table 1 are simulated again with nonlinear feedback devices, and the results given in Table 2 are in agreement with this conclusion. As seen from Table 2, IM3 increases by a maximum of 9 dB in the differential input and 14 dB in the mismatched input case, which is not slight, yet low compared to the 30 dB cancellation we still have. The performance degradation is worse at higher frequencies and in the mismatched case, which is also anticipated: the linearity of the feedback path matters if the feedback transistors can mix IM2 to IM3. This is illustrated if we have a mismatched condition where there is also a fundamental common-mode signal, as in the mismatched case. However, a 10% mismatch is still so small that the resulting deterioration in the total IM3 will not be significant (3–5 dB).

Finally, the effect of op-amp gain is studied and three cases with 15 dB, 25 dB and 35 dB gain are compared in Table 3 (the bandwidth is kept at 100 kHz). The linear signal and IM2 are fixed at − 33.4 dB and − 78 dB, respectively. It is found that solely increasing the gain improves distortion cancellation, but this changes when the controlling pMOS sizes and nonlinear coefficients are scaled in proportion to the gain. (Note that the 25 dB gain case is taken as a standard, and the pMOSes in the other two cases are scaled with respect to it). In this case, the effect is positive from 15 to 25 dB, but detrimental from 25 to 35 dB. This is due to the fact that, as explained previously, the pMOS nonlinearities are amplified by this gain and could become effective past a certain gain value. We

Table 2 The effect of CMFB nonlinearity on distortion results

Input	CMFB 3 dB BW	Linear signal (dB)	IM2 (dB)	IM3 (dB)
Diff	100 kHz		− 78	− 114.1
	3 GHz	− 33.4	− 77.7	− 112.8
	6 GHz		− 77.7	− 112.7
CM	100 kHz	− 32.4	− 75.4	− 107.8
	3 GHz	− 29.9	− 47.1	− 46.3
	6 GHz	− 25.4	− 44.3	− 42.5
Asymm	100 kHz		− 79.9	− 116.9
	3 GHz	− 34.4	− 80.3	− 112.1
	6 GHz		− 80.2	− 110.3

Table 3 The effect of CMFB op-amp gain on distortion results

Opamp gain (dB)	IM3 (dB)	Scaled PMOSes
15	– 116.1	No
15	– 113.2	Yes
25	– 114.1	–
35	– 111.5	No
35	– 112.6	Yes

can conclude that sizing the pMOSes based on constant overall gm and choosing the gain accordingly is the best option, but increasing the gain further may make the distortion in the feedback loop visible.

7 Modeling the frequency conversion

So far, the analysis has not assumed any mixing effect - instead, the mixer has been considered as a pass transistor with a fixed and high gate bias. Adding the actual mixing could give rise to new and significant spurious tones that may affect the overall distortion. However, the used distortion contribution analysis tool does not easily bode with the inclusion of mixing effects. In this section, we will present analyses on what is expected to happen and discuss the technical modeling problems and some practical findings.

7.1 Ways to include mixing in the distortion contribution analysis

We have explored three alternatives to implement the mixer in our nonlinearity analysis tool. One is to employ the spectral presentation of the time-varying mixer gain, as traditionally done in analyzing mixers. Typically, this demands solving the frequency translating terms and then copying the circuit multiple times for different frequencies, as in [27], which makes the impedance matrix representing the circuit significantly larger. It may also be complicated to fit into the distortion contribution analysis since it might require injecting higher order signals, while our tool builds the distortion products order by order.

The second option is to model the mixer behaviorally. There are several ways to do this, but the most straightforward would be to perform the mixing numerically somewhere outside the device and inject the mixed signal externally. This makes the device a pass transistor again, limiting its function to a non-mixing nonlinear device. Fitting would not be troublesome in this scenario, since it is fitted in exactly the same way as before. Nevertheless, modeling the external injection could cause complications,

such as having to break down the circuit in a somewhat similar way to time-variant fitting.

Lastly and most simply, we could model the mixer in the same way as its circuit implementation, feeding LO to the device gate and RF to its source. This imposes a large-signal fit that covers two operating regions as opposed to the small signal fit done in Sect. 5. Implementing the mixer as a large-signal nonlinearity would be appealing in the sense that it would cause no changes in the current NLSim analysis flow nor any added circuit complexity, and would easily lend itself for modeling the effects of the nonlinear capacitances in the mixer as well. The idea is also backed up in [28], which shows that a time-invariant nonlinearity fitted on large enough range has almost the same modeling strength as a time-varying one. Hence, we have proceeded with this method in the remainder of the paper.

7.2 The spectral effects of the mixing

Before diving into the modeling details, it would be helpful to give a circuit analysis of the spectral effects a real mixer could add and discuss the changes it would bring to our previous analysis. This will give us insight on what is really occurring in the circuit and clarify what to expect from the large-signal modeling of the mixer. As the passive mixer is notoriously leaky between the RF and IF ports (described in e.g. [26]), additional spectral components will sum up into RF. If strong enough, they may easily affect the total amount of distortion.

A passive mixer can be understood as a switch that is either on or off with the frequency of the LO signal and a duty cycle D . The periodic LO waveform can be written as its Fourier components, and if $D = 50\%$, the dominant spectral components of a two-sided spectrum of the 0/1 gating function are $X(-f_{LO}) = 1/\pi$, $X(0) = 0.5$, and $X(f_{LO}) = 1/\pi$, where $X(-f_{LO})$ and $X(f_{LO})$ are the negative and positive LO spectral components. In the balanced mixer, we would have components with the same magnitude, but a phase shift of 180 degrees in the fundamentals.

Assuming pass transistor on-conductance of g_{on} , the down-conversion gain will be $X(f_{LO}) \cdot g_{on}$, while leakage from both RF to output and IF to input is $X(0) \cdot g_{on}$. This causes the IF to appear in the RF port of the mixer, where it will be an additional signal in the distortion contribution analysis and a relatively strong interferer in a single-ended mixer. However, if the mixer is balanced as shown in Fig. 9, the $X(f_{LO})$ gain terms of the two mixers will be in opposite phases, causing the back-propagating IF terms to cancel each other in the mixer input port. This is how IF leakage cancels out in the input node of a balanced mixer, as illustrated in the figure.

The mixer is still nonlinear. The IM3 it generates on the input side will be mixed to the output side, and the second

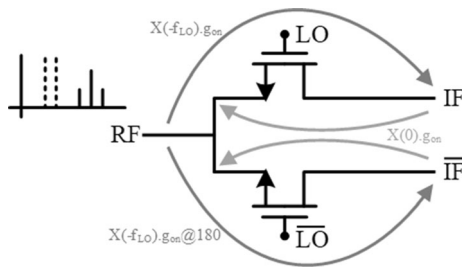


Fig. 9 Mixing from RF to IF and leakage back to the RF port

harmonic of the IF will leak back to the RF side and sum up there coherently. Hence, though it is not cancelled, it is smaller than the IF leakage in power and appears as a common-mode signal in the pseudo-differential structure, in which case the common mode feedback helps in cancelling it. Therefore, a dramatic change to the previous analysis is not to be expected, although including the real mixing effect would make it more comprehensive and may also reveal some unpredicted phenomena.

7.3 Large-signal fitting results and vector contribution analysis

The large-signal fit of the mixing transistor was made over a v_{gs} range that covered the high-amplitude LO signal and a v_{ds} range of 0.1 V. The obtained set of coefficients was placed into NLSim with a single FET device as the test setup. The resulted spectrum around the baseband at the device output side was compared against the actual circuit simulation. The down-conversion gain was closely matching within 1 dB, but the IM3 predicted by NLSim was more than 10 dB higher than in the circuit simulations. Initially, the model order was reduced to the smallest with which it could represent the I–V curve, having only v_{gs} , v_{gs}^2 , v_{ds} , v_{ds}^3 and the corresponding cross-terms. However, increasing the order of the model did not improve the results. To get to the bottom of this, we will first take a look at how the nonlinearity analysis evolves order by order with the fitted values, and which coefficients are the most effective at each order. Then, we will have a clearer viewpoint on whether the modeling scheme needs to be rethought.

As mentioned previously, NLSim takes the node voltages, converts it to distortion currents using the nonlinear coefficients, and then solves the distortion voltages by multiplying the nonlinear currents by the node impedances. So, looking at the IF for instance, we will have (assuming just a resistive load R_{load} at the drain of a single mixer):

$$V1_{drain} = -R_{load}(K_{01}v_{gs} + K_{01}v_{ds}) \quad (3)$$

$$V2_{drain} = -R_{load}(K_{20}V_{gs}^2 + K_{11}V_{gs}V_{ds}) \quad (4)$$

where $V1_{drain}$ and $V2_{drain}$ are first and second order voltages at the device output. Multiplication in time-domain translates to convolution in frequency domain. Here, $V1$ is the combination of LO and RF as shown in Fig. 10, hence NLSim sees (4) at the IF frequency as:

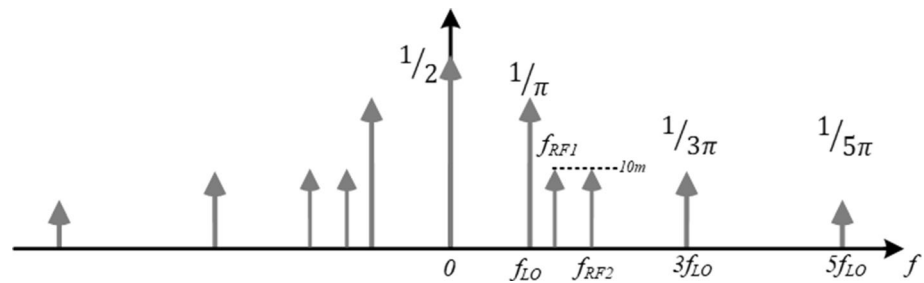
$$V2_{drain} = -R_{load} \times [2K_{20} \cdot V_{gs}(f_{RF}) \cdot V_{gs}(-f_{LO}) + K_{11} \cdot V_{gs}(f_{RF}) \cdot V_{ds}(-f_{LO}) + K_{11} \cdot V_{ds}(f_{RF}) \cdot V_{gs}(-f_{LO})] \quad (5)$$

The same procedure is continued up to the fifth order. As the hand analysis gets too lengthy for the scope of this paper, we will continue investigating the buildup of nonlinearities using the vector plots, which are showing essentially similar results. As expected, $V2$ is the largest voltage generated and the higher order terms tend to be smaller and cancel each other out. Thus, K_{11} is the most influential coefficient since it models the mixing of LO and RF. Even though the fit produces rather large higher order cross-terms, K_{11} particularly determines the conversion gain as the others do not yet come into play. Nevertheless, the situation is different with IM3s, which are non-existent in $V3$ (the beat frequency and IF are both present in $V2$ but not in $V1$, thus $RF \pm IF$ disappears in their convolution) and emerge in the fourth and fifth order only. Figure 11 illustrates the orientation of the vector plots in $V4$ and $V5$. In the former, K_{21} and less dominantly K_{13} are trying to cancel K_{11} . In the latter, the effect of K_{11} is not seen at all, hence no cancelling terms and a bigger contribution to IM3s. The overall result is a sum of all orders, so a smaller IM3 means either a smaller K_{11} or higher K_{21} and K_{13} for $V4$ and $V5$ to be cancelling. The model has predicted K_{11} correctly as discussed above, so the problem must lie within the higher order cross-term coefficients. We have limited the model order so far to make the analysis easier to follow, but adding more cross-terms would make them appear here as well, building up in $V5$ and canceling K_{11} in $V4$. The reason for this apparently inherent problem in the fitting will be explained in the next section.

7.4 Problem of large-signal fitting

Intuitively, one might speculate that the fitting problem originates from trying to fit two different operating regions into one set of coefficients. Fitting such a large scale nonlinearity is unfortunately a compromise and quite sensitive to the voltage range used for the fitting. Figure 12 illustrates the spectral performance for a clean two-tone RF test versus the DC I–V fitting range. The numbers on the axes indicate the ratio of the fitting range to the actual signal range (i.e. point (1,1) is the actual operating point). It

Fig. 10 First order voltage in the large signal modeling



is seen that the conversion gain (left) matches quite well, while the IM3 of the down-converted result (right) is more sensitive. However, the IM3 error given in NLSim is not the same as this result, which leads us to the most interesting finding of the fitting: a set of coefficients giving 1 dB IM3 match when driven by a clean zero-impedance two-tone still generates 10 dB error when simulated with proper terminal impedances.

This means that the terminating impedances are further converting second order current to voltage and causing it to mix into IM3, which is in line with the vector plots in Sect. 7.3 (K_{21} transferring second and third order output-related components to IM3). The interpretation for this could be that, to achieve the on-off large-signal switching, high even order nonlinearities are needed to model the off region. This generates high IM2 currents that mix further into IM3. In other words, the strength needed for modeling the on-off switching overestimates and outweighs the curvature in the on region. In essence, with a 16-term

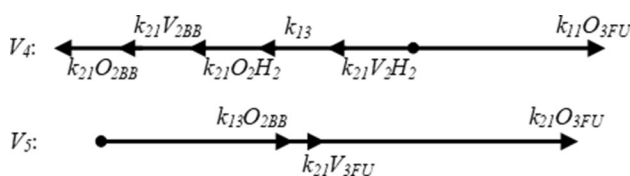
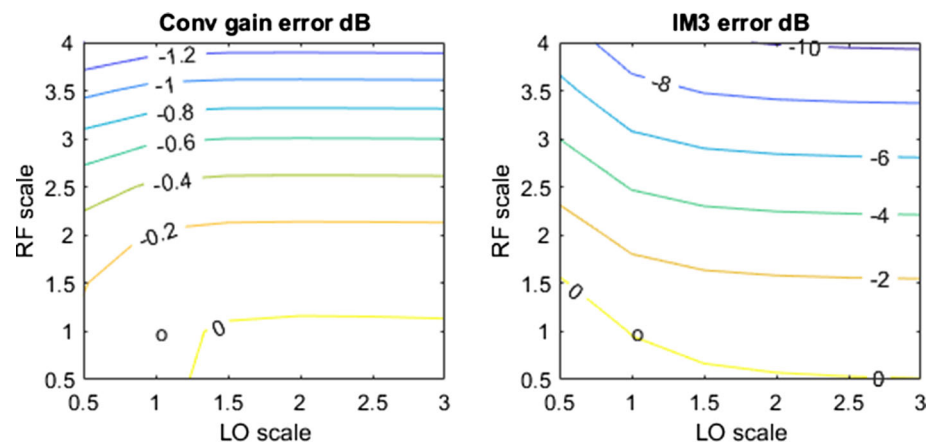


Fig. 11 Vector plots of the nonlinear voltages resulting in IM3

Fig. 12 Error of conversion gain (left) and IM3 (right) of a passive mixer versus the fitting voltage range of the polynomial model



polynomial, we can achieve a -35 dB NMSE fit for the large-signal I–V response and very precise conversion gain. Nevertheless, the on region (which is quite linear in practice) is now modeled by a set of mutually cancelling high-order polynomial terms that cause the contribution analysis to give unrealistic results.

In conclusion, the large-signal nonlinear modeling seems to be appropriate for large-signal gain estimation, but cannot be used simultaneously for small-scale distortion analysis. Hence, it seems necessary to proceed towards time-varying or behavioral modeling in future works.

8 Discussion

This paper has investigated the mechanisms that generate IM3 in the combination of an inverter-type amplifier and a passive mixer. The utilized distortion contribution analysis technique allows to show the relative importance of all the contributions of polynomially modelled nonlinear devices.

The first finding was that, even though the inverter-type gm is quite linear, its linearity can still be further improved. Thus, a polynomial model of its input-output response was extracted at different bias points, and several zero-crossing points of the quadratic and cubic coefficients were observed in the operating band. Consequently, it was found that it is possible to bias the circuit in a sweet spot or near it, and even apply the parallelism of the structure so that if

one gm slice is biased expansively, its neighbor would be compressive.

The second finding was that—as expected—the common-mode feedback loop attenuates the even-order nonlinearities that appear as common-mode interference in the output of the pseudo-differential amplifier. The amount of this attenuation depends on the available loop gain and bandwidth, which are also investigated in the paper.

The third finding was that, in the passive mixer, the input voltage directly modulates the V_{GS} of the switch. This generates a strong second-order nonlinearity that first generates IM2 products, and then mixes them again with fundamental tones and generates IM3. The common-mode feedback is very useful here, as it reduces the IM2 level by 24 dB and IM3 level by 40 dB, to a level where it can be said to match the performance of the gm amplifier again.

The fourth finding was that, luckily the bandwidth of the common-mode feedback does not need to be dramatically wide. With the baseband IM products being the most dominant cause of distortion, cancelling the low-frequency IM2 is already quite helpful. Also, as the first and second harmonic bands are just an octave away from each other, it does not matter much if the 3dB bandwidth ends in the fundamental or the second harmonic band.

The fifth finding was that, the nonlinearity of the CM-correcting pMOSes (which is assumed to be the dominant reason for nonlinearity regarding the CMFB) does not prove very destructive as the signal present in that node is small compared to the input RF signal. Although it causes 9–14 dB degradation in nonlinearity performance in the worst case scenario, there is still about 30 dB improvement over the no-feedback case. However, if the op-amp gain is too high, this effect may become more adverse.

Finally, the amount of spurious components generated by the mixer leakage was evaluated. The IF leakage to the RF side generates a strong spurious tone in a single mixer, but gets cancelled in a balanced mixer structure. The common-mode feedback helps further by reducing its second harmonic. In attempts to modify the previous analysis to handle the mixing effects, it was found that large-signal fitting cannot model mixing and distortion simultaneously, predicting the conversion gain very well but failing to predict the amount of small-signal distortion correctly. This happens because IM3s are dictated by higher order nonlinear voltages and hence higher order cross-terms, while the fitting model cannot circumvent all the cross-terms due to trying to make a trade-off between two different device regions.

The study was made using a MATLAB-based distortion contribution analysis tool NLSim [24]. Device I–V curves were extracted in Cadence, fitted to polynomial models in MATLAB, and calculated to vectors in NLSim. Results were checked against pss simulations in Cadence.

Acknowledgements This project was funded by Finnish Academy Project 325928 and Academy of Finland 6Genesis Flagship under Grant 318927.

Funding Open access funding provided by University of Oulu including Oulu University Hospital.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study. License restrictions may apply to sharing the simulations that support the findings of this paper.

References

1. Tervo, N., Aikio, J., Tuovinen, T., Rahkonen, T., & Parssinen, A. (2017). Digital predistortion of amplitude varying phased array utilising over-the-air combining. In *IEEE MTT-S international microwave symposium (IMS), Honolulu, HI* (pp. 1165–1168). <https://doi.org/10.1109/MWSYM.2017.8058809>.
2. Soer, M. C. M., Klumperink, E. A. M., van den Broek, D., Nauta, B., & van Vliet, F. E. (2017). Beamformer with constant-GM vector modulators and its spatial intermodulation distortion. *IEEE Journal of Solid-State Circuits*, 52(3), 735–746. <https://doi.org/10.1109/JSSC.2016.2639545>.
3. Akbar, R., et al. (2019). A wideband IF receiver module for flexibly scalable mmWave beamforming combining and interference cancellation. In *ESSCIRC, IEEE 45th European solid state circuits conference (ESSCIRC), Cracow, Poland, 2019* (pp. 213–216). <https://doi.org/10.1109/ESSCIRC.2019.8902869>.
4. Nakajima, T., Amakawa, S., Ishihara, N., & Masu, K. (2009). A scalable wideband low-noise amplifier consisting of CMOS inverter circuits for multi-standard RF receivers. In *2009 3rd international conference on signals, circuits and systems (SCS)* (pp. 1–4). <https://doi.org/10.1109/ICSCS.2009.5412585>.
5. Chao, S., Kuo, J., Lin, C., Tsai, M., & Wang, H. (2008). A DC–11.5 GHz low-power, wideband amplifier using splitting-load inductive peaking technique. *IEEE Microwave and Wireless Components Letters*, 18(7), 482–484. <https://doi.org/10.1109/LMWC.2008.925099>.
6. Liraneto Torres Costa, A., Klimach, H., & Bampi, S. (2016). Ultra-low voltage wideband inductorless balun LNA with high gain and high IP2 for sub-GHz applications. In *2016 IEEE international symposium on circuits and systems (ISCAS)* (pp. 289–292). <https://doi.org/10.1109/ISCAS.2016.7527227>.
7. Martins, M. A., Mak, P., & Martins, R. P. (2012). A 0.02-to-6GHz SDR Balun-LNA using a triple-stage inverter-based amplifier. In *2012 IEEE international symposium on circuits and systems (ISCAS)* (pp. 472–475). <https://doi.org/10.1109/ISCAS.2012.6272067>.

8. Ye, D., Xu, R., & Richard Shi, C. (2019). 26.4 A 2.4 GHz 65nm CMOS mixer-first receiver using 4-stage cascaded inverter-based envelope-biased LNAs achieving 66 dB in-band interference tolerance and -83 dBm sensitivity. In *2019 IEEE international solid-state circuits conference—(ISSCC)* (pp. 414–416). <https://doi.org/10.1109/ISSCC.2019.8662451>.
9. Chen, W., Liu, G., Zdravko, B., & Niknejad, A. M. (2008). A highly linear broadband CMOS LNA employing noise and distortion cancellation. *IEEE Journal of Solid-State Circuits*, 43(5), 1164–1176. <https://doi.org/10.1109/JSSC.2008.920335>.
10. Li, P., & Pileggi, L. T. (2003). Efficient per-nonlinear distortion analysis for analog and RF circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(10), 1297–1309. <https://doi.org/10.1109/TCAD.2003.818130>.
11. Borremans, J., de Locht, L., Wambacq, P., & Rolain, Y. (2007). Nonlinearity analysis of analog, RF circuits using combined multisine and volterra analysis. In *Design, automation and test in Europe conference and exhibition, 2007* (pp. 1–6). <https://doi.org/10.1109/DATE.2007.364601>.
12. Palumbo, G., & Pennisi, S. (2003). High-frequency harmonic distortion in feedback amplifiers: Analysis and applications. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 50(3), 328–340. <https://doi.org/10.1109/TCSI.2003.808835>.
13. Cheng, W., Annema, A. J., Croon, J. A., Klaassen, D. B. M., & Nauta, B. (2008). A general weak nonlinearity model for LNAs. In *IEEE custom integrated circuits conference, 2008* (pp. 221–224). <https://doi.org/10.1109/CICC.2008.4672063>.
14. Aparin, V., Brown, G., & Larson, L. E. (2004). Linearization of CMOS LNA's via optimum gate biasing. In *2004 IEEE international symposium on circuits and systems (IEEE Cat. No.04CH37512)* (pp. IV–748). <https://doi.org/10.1109/ISCAS.2004.1329112>.
15. Jafarnejad, R., Jannesari, A., & Sobhi, J. (2017). Pre-distortion technique to improve linearity of low noise amplifier. *Microelectronics Journal*, 61, 95–105. <https://doi.org/10.1016/j.mejo.2017.01.006>.
16. Amirabadi, A., Zokaei, A., Bagheri, M., & Alirezazadeh, F. (2015). Highly linear wide-band differential LNA using active feedback as post distortion. In *IEEE international symposium on circuits and systems (ISCAS), 2015* (pp. 654–657). <https://doi.org/10.1109/ISCAS.2015.7168718>.
17. Cheng, C., Li, X., Li, Y., Zeng, H., Wang, Z., & Yasir, U. (2021). A triple-cascode X-band LNA design with modified post-distortion network. *Electronics*, 10(5), 546. <https://doi.org/10.3390/electronics10050546>.
18. Kim, T. W., Kim, B., & Lee, K. (2004). Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors. *IEEE Journal of Solid-State Circuits*, 39(1), 223–229. <https://doi.org/10.1109/JSSC.2003.820843>.
19. Liu, T. T. & Rabaey, J. M. (2011). Linearity analysis of CMOS passive mixer. In *IEEE international symposium of circuits and systems (ISCAS), Rio de Janeiro, 2011* (pp. 2833–2836). <https://doi.org/10.1109/ISCAS.2011.5938195>.
20. Khatri, H., Gudem, P. S., & Larson, L. E. (2009). Distortion in current commutating passive CMOS downconversion mixers. *IEEE Transactions on Microwave Theory and Techniques*, 57(11), 2671–2681. <https://doi.org/10.1109/TMTT.2009.2031930>.
21. Maas, S. A. (2003). *Nonlinear microwave and Rf circuits* (2nd ed.). Artech House.
22. Wambacq, P., & Sansen, W. (2013). *Distortion analysis of analog integrated circuits*. Springer. <https://doi.org/10.1007/978-1-4757-5003-4>.
23. Aikio, J. P., & Rahkonen, T. (2005). Detailed distortion analysis technique based on simulated large-signal voltage and current spectra. *IEEE Transactions on Microwave Theory and Techniques*, 53(10), 3057–3066. <https://doi.org/10.1109/TMTT.2005.855132>.
24. Heiskanen, A., & Rahkonen, T. (2002). 5th order multi-tone Volterra simulator with component-level output. In *2002 IEEE international symposium on circuits and systems. Proceedings (Cat. No.02CH37353)* (pp. III–III). <https://doi.org/10.1109/ISCAS.2002.1010293>.
25. Aikio, J., Rahkonen, T., & Korkala, T. (2014). Analysis of band-to-band mixing distortion contributions in some usual circuit topologies. *Analog Integrated Circuits and Signal Processing*, 80, 417–426. <https://doi.org/10.1007/s10470-014-0320-2>.
26. Andrews, C., & Molnar, A. C. (2010). A passive mixer-first receiver with digitally controlled and widely tunable RF interface. *IEEE Journal of Solid-State Circuits*, 45(12), 2696–2708. <https://doi.org/10.1109/JSSC.2010.2077151>.
27. Sosio, M., Liscidini, A., & Castello, R. (2013). An intuitive current-driven passive mixer model based on switched-capacitor theory. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 60(2), 66–70. <https://doi.org/10.1109/TCSII.2012.2234993>.
28. Rahkonen, T., & Aikio, J. (2013). Comparison of ordinary and time-varying Volterra analysis for finding distortion contributions. *Analog Integrated Circuits and Signal Processing*, 74, 3–9. <https://doi.org/10.1007/s10470-012-9904-x>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Negar Shabanzadeh was born in Babol, Iran in 1993. She received her B.Sc. in Electrical Engineering from Babol Noshirvani University of Technology in 2013, and her M.Sc. in the same field from Tarbiat Modares University in 2016. She is now pursuing her Ph.D. in University of Oulu, doing research on the nonlinearity of electrical circuits and distortion contribution analysis.



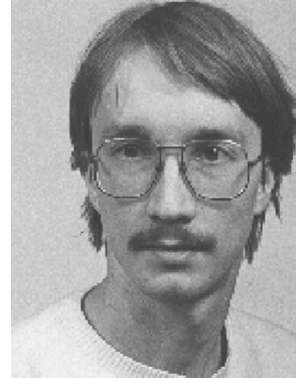
Rehman Akbar received the M.Sc. degree in electrical engineering from Tampere University of Technology (TUT), Tampere, Finland, in 2014. From 2012 to July 2015, he was a member of the RFIC Design Lab at TUT, where he worked on VHF range DC-DC Converters for mobile devices. He is currently pursuing a Ph.D. degree in electrical engineering at the University of Oulu, Oulu, Finland. His current research interests include wideband RF/mmWave/BB integrated circuits and next-generation Wireless systems.



Aarno Pärssinen (Senior Member, IEEE) received the M.Sc., Licentiate in Technology, and Doctor of Science degrees in electrical engineering from the Helsinki University of Technology, Finland, in 1995, 1997, and 2000, respectively. From 1994 to 2000, he was with the Electronic Circuit Design Laboratory, Helsinki University of Technology, Espoo, Finland, working on direct conversion receivers and subsampling mixers for wireless communica-

tions. In 1996, he was a Research Visitor with the University of California at Santa Barbara, Santa Barbara, CA, USA. From 2000 to 2011, he was with the Nokia Research Center, Helsinki, Finland. From 2009 to 2011, he was a member of the Nokia CEO Technology Council. From 2011 to 2013, he was with Renesas Mobile Corporation, Helsinki, working as a Distinguished Researcher and an RF Research Manager. From October 2013 to September 2014, he was an Associate Technical Director with Broadcom Corporation, Helsinki. Since September 2014, he has been with the Centre for Wireless Communications, University of Oulu, Oulu, Finland, where he is currently a Professor. He is also leading devices and circuits research area in 6G flagship program financed by the Academy of Finland. His research interests include wireless systems and transceiver architectures for wireless communications with special emphasis on the radio

frequency (RF) and analog integrated circuit and system design. He has authored or coauthored one book, two book chapters, more than 150 international journal and conference papers, and holds several patents. He is also one of the original contributors to Bluetooth low-energy extension, now called BT LE. Dr. Pärssinen served as a member for the Technical Program Committee of the International Solid-State Circuits Conference from 2007 to 2017, where he was the Chair of the European Regional Committee from 2012 to 2013 and the Chair of the Wireless Sub-Committee from 2014 to 2017. He has been serving as a Solid-State Circuits Society Representative for the IEEE 5G Initiative from 2015 to 2019.



Timo Rahkonen was born in Jyväskylä, Finland, 1962. He received his Diploma Engineer, Licentiate and Doctor of Technology degrees in 1986, 1991 and 1994, respectively, in University of Oulu, Finland. Currently he is a professor in circuit theory and circuit design in University of Oulu, Oulu, Finland, conducting research on linearization and error correction techniques for RF power amplifiers and A/D and D/A converters.