Linearization of Active Downconversion Mixers at the IF Using Feedforward Cancellation

Hao Li\textsuperscript{\textcopyright}, Member, IEEE, and Carlos E. Saavedra\textsuperscript{\textcopyright}, Senior Member, IEEE

Abstract—A feedforward linearization technique for third-order intermodulation (IM\textsubscript{3}) distortion cancellation in active downconversion mixers is proposed in this paper. Low-frequency second-order intermodulation (IM\textsubscript{2}) tones are created and multiplied with the mixer’s output to generate low-frequency IM\textsubscript{3} replicas for cancellation. Implemented mostly at the IF band, this technique brings a third-order input intercept point (IIP\textsubscript{3}) improvement independently of the mixer topology and is robust against parasitic parameters. A 2-GHz current commutating mixer linearized by the proposed technique is designed and fabricated using a 130-nm CMOS process to verify the concept. Experimental results show that the mixer with a unit-gain amplifier delivers 8.5 dB of conversion gain and has an IIP\textsubscript{3} of 2.5 dBm before linearization. The linearization technique improves the mixer’s IIP\textsubscript{3} by 12 dB for input signals as large as −15 dBm. The technique has a negligible impact on the mixer’s gain and incurs a noise figure penalty of less than 0.2 dB. The mixer with the unit-gain amplifier consumes a current of 8.4 mA, while the proposed technique circuitry consumes an extra current of 4.2 mA, both using a 1.2-V voltage supply.

Index Terms—Active mixer, blocker, current-bleeding, digital assist, distortion, distortion cancellation, downconverter, dynamic current-injection, feedforward linearization, interferer, intermodulation, linearization, microwave mixer, receiver, MMIC, Gilbert-cell, RFIC, third-order intermodulation, PVT.

I. INTRODUCTION

The investigation of linearization techniques for active mixers is a topic of continued interest because ever-more stringent linearity performance is expected of single-chip mobile downconverter front-ends as the available spectrum is increasingly crowded with interferers.

In derivative superposition (DS) [1]–[8], auxiliary transistors are used which are biased at the edge of threshold to produce IM\textsubscript{3} currents that are out-of-phase with the IM\textsubscript{3} currents produced by transistors in the main path. When the two sets of currents are added at a circuit node, the IM\textsubscript{3} tones are cancelled. DS can improve the mixer's IIP\textsubscript{3} with only minimal chip area and power consumption requirements.

While DS can be sensitive to process, voltage and temperature (PVT) variations, these issues can be mitigated through digital-assist [9].

In a conventional CMOS Gilbert cell mixer, its IIP\textsubscript{3} is proportional to the square root of the bias current flowing in the input transconductance (RF) stage [10]. Since the switching stage of the mixer is stacked above the RF stage, the two stages share the same output current. Therefore, if the bias current is increased in an attempt to improve the linearity of the RF stage, this action will reduce the voltage headroom of the switching stage which ultimately deteriorates the linearity of the mixer. This problem is resolved using a linearization technique known as charge injection, or current bleeding, whereby the current flows in the RF stage and the switching stage are made different [10]–[13]. The benefit of the technique is that several things are accomplished at once: the RF stage is supplied the larger bias current it needs to increase its linearity, the RF transconductance gain is increased as is the conversion gain of the mixer, the switching transistors maintain a lower current flow which allows for softer switching and hence a smaller LO power requirement. Some issues with current-bleeding are increased power consumption and gain roll-off at higher frequencies but there are methods to mitigate these as well [14], [15].

A different linearization technique consists of generating IM\textsubscript{2} tones of the input signals and injecting them at the output of the transconductance stage of the mixer [16], [17]. The IM\textsubscript{2} product mixes with the fundamental input signals to generate an IM\textsubscript{3}, which cancels the intrinsic IM\textsubscript{3} signal in the main RF path. This cancellation technique can suppress IM\textsubscript{3} effectively with negligible extra NF and low additional dc power introduced. However, it can be only applied to transconductors consisting of differential pairs.

The topology reported in [18] adopts negative impedances at the source node of the mixer switches to cancel the intrinsic IM\textsubscript{3} tones generated in the mixer. Due to the dynamic current bleeding provided by these negative impedances, an IIP\textsubscript{3} of 11.8 dBm is achieved. Besides producing a major third-order nonlinearity suppression, this method also brings a significant flicker noise improvement to the current commutating mixer. However, similar to the IM\textsubscript{2} injection technique, this method can only be applied to current commutating mixers.

A feedforward technique is proposed in [19] to improve the IIP\textsubscript{3} of the receiver, which is largely independent of the receiver circuit topology to be linearized. In this technique, a separate receiver is employed as the auxiliary path, in which
IM3 products are generated, down-converted to baseband, digitized, and processed by an adaptive equalizer for canceling the IF band IM3 of the main path. Although the dynamic range of the auxiliary path is much smaller compared to the main path, the adoption of it still increases the power dissipation as well as the complexity of the system drastically.

In this paper, a feedforward IM3 cancellation method for mixers is presented. This method is largely independent of the core mixer’s topology and is robust against parasitic parameters, because the linearization circuitry is mostly in the IF frequency range. This paper is an extended version of [20], a paper from a conference proceedings, that presented the basic principle of the proposed mixer linearization method. Here, we provide a more in-depth analysis of the linearization technique in Section II and we give a significant amount of additional details about the design and implementation of key circuit blocks in Section III. We also give insights on how the scheme can be realized in a practical receiver link and we carry out a study of the impact of PVT variations on IIP3. Experimental test results for a 2-GHz linearized mixer chip using the proposed technique are presented in Section V and Section VI concludes the work.

II. LINEARIZATION TECHNIQUE

The feedforward method is capable of suppressing non-linearities without affecting the circuit topology, since it employs an additional parallel auxiliary path to cancel out the non-linearities on the main path. As illustrated in Fig. 1(a), the nonlinear coefficients that have same values and opposite signs to that of the main path are generated in the auxiliary path. Through a linear combination, the nonlinear signal components of the two paths are canceled out and the linear signal is obtained at the output.

The feedforward linearization technique can be applied to the downconversion mixer to suppress its third-order intermodulation terms, as illustrated in Fig. 1(b) [19]. In the auxiliary path, IM3 around fundamental tones need to be generated for the compensation first. And then, since the IM3 tones to be canceled at the mixer output are at IF frequencies, the generated IM3 tones need to be down-converted by a mixer before the cancellation.

The method described above is typically power-inefficient and susceptible to the parasitics of the analog circuit. Aiming to solve these two problems, a novel scheme is proposed in this work, in which the IM3 for the compensation is generated directly in the IF band, which leads to an interesting method robust against the parasitic parameters.

A. Proposed Feedforward Linearization

A block diagram of the proposed feedforward scheme is given in Fig. 2. As can be seen, an auxiliary path for linearization is added in parallel to the original receiver. In the auxiliary path, a low-frequency IM2 product is generated from fundamental input signal first, and then mixed with mixer output signal to generate an IF IM3 required for the cancellation with the intrinsic IF IM3 signal in the main path. The combining of the mixer output and the cancelling IM3 can be achieved after the mixer using an op-amp IF amplifier or filter. From the system perspective, the combiner used in the proposed technique does not add an extra stage in the main path. A detailed description of the operating principle is provided below using a two-tone signal applied at the input.
Node A: Two closely-spaced tones enter the mixer at the RF input port. In keeping with standard two-tone test analysis of intermodulation distortion [21], the initial amplitude and phase of the signal are assumed to be \( A_0 \) and 0°, respectively. The signal appearing at node A can be expressed as

\[
v_A = A_0 \cos(\omega_1 + \omega_{LO})t + A_0 \cos(\omega_2 + \omega_{LO})t \tag{1}
\]

in which the RF frequencies are expressed as the sum of IF and LO for signals for the simplification of the notation.

Node B: The mixer converts the input RF signal to the IF band by an amount of the LO frequency. Simultaneously, due to the third-order nonlinearities of the mixer, IM3 tones are produced near the fundamental ones in the IF frequencies at the output of the mixer. The signals within the band of interest at node B can be expressed as

\[
v_B = ACGA_0 \cos(\omega_{11} + \Phi_1) + ACGA_0 \cos(\omega_{22} + \Phi_2) + \frac{3}{4} a_3 A_0^3 \cos(\omega_{3L} + \Phi_3) + \frac{3}{4} a_3 A_0^3 \cos(\omega_{3H} + \Phi_4) \tag{2}
\]

where

\[
\omega_{3L} = 2\omega_1 - \omega_2 \\
\omega_{3H} = 2\omega_2 - \omega_1 \tag{3}
\]

and \( ACG \) and \( a_3 \) represent the conversion gain and the third-order nonlinear coefficient of the mixer, respectively; \( \Phi_1-4 \) represents the additional phase introduced to each tone by the mixer.

Node C: In the auxiliary path, a low-frequency IM2 tone of the input signal is generated first at node C, noted as

\[
v_C = a_2 A_0^2 \cos(\omega_2 - \omega_1)t \tag{4}
\]

where \( a_2 \) represents the second-order coefficient of the IM2 generator. As the IM2 tone stays at low frequency, its phase shift due to parasitic capacitors of the circuit can be ignored here without losing accuracy. In fact, this phase shift determines the frequency spacing of the blockers the proposed technique can deal with, which will be discussed in details in next section.

Node D: The baseband multiplier multiplies the baseband signals at Node B by those at Node C, generating four third-order and four fifth-order tones located around the fundamental tones of the mixer output. Since the fifth-order products are small and also not related with IM3 cancellation, they are neglected here and their effects will be discussed later. Here without losing accuracy, only third-order products at node D are listed,

\[
v_D \approx A_{IM3} \cos(\omega_{11} + \Phi_1) + A_{IM3} \cos(\omega_{22} + \Phi_2) + A_{IM3} \cos(\omega_{3L} + \Phi_1) + A_{IM3} \cos(\omega_{3H} + \Phi_2) \tag{5}
\]

where

\[
A_{IM3} = \frac{1}{2} K_m a_2 ACG A_0^3 \tag{6}
\]

and \( K_m \) is the multiplying gain of the baseband multiplier. The phase shift of this operation is ignored too since it is also in the IF band.

Node E: Signals at Node D are fed back to the main path and added with the mixer output through a combining circuit to cancel the IM3 tones. Assuming a unit gain for the combiner, the signals at Node E can be expressed as

\[
v_E = ACGA_0 \cos(\omega_{11} + \Phi_1) + A_{IM3} \cos(\omega_{11} + \Phi_2) + ACGA_0 \cos(\omega_{22} + \Phi_2) + A_{IM3} \cos(\omega_{22} + \Phi_1) + \frac{3}{4} a_3 A_0^3 \cos(\omega_{3L} + \Phi_3) + A_{IM3} \cos(\omega_{3L} + \Phi_1) + \frac{3}{4} a_3 A_0^3 \cos(\omega_{3H} + \Phi_4) + A_{IM3} \cos(\omega_{3H} + \Phi_2) \tag{7}
\]

The condition for complete IM3 cancellation is clear: the corresponding IM3 tones at same frequencies should have the same amplitude and 180° of phase difference. Referring to (7), this condition can be expressed mathematically as

\[
a_3 = -\frac{2}{3} K_m a_2 ACG \tag{8a}
\]

\[
\Phi_1 = \Phi_3 \tag{8b}
\]

\[
\Phi_2 = \Phi_4 \tag{8c}
\]

B. Discussions on Cancellation Condition

The cancellation condition in (8) reveals alignment requirements of two aspects for the IM3 tones between the main and auxiliary paths: the amplitude and the phase.

The amplitude alignment, expressed by (8a), can be fulfilled by adjusting \( a_2 \) of the IM2 generator and \( K_m \) of the multiplier in the auxiliary path, with \( ACG \) and \( a_3 \) regarded as constants once the mixer in the main path is designed. Since the entire circuits are implemented in the differential manner, “...” in (8a) can be easily realized by reversing the positive and negative terminals while combining the signals of the two paths. Note that the original and the cancelling IM3 are both in the third order of input amplitude, and all the parameters in (8a) are device-related and have nothing to do with the input amplitude. In this way, if (8a) is fulfilled, the effective cancellation over wide range of input power can be achieved because the injected signal automatically tracks with the input signal.

The phase alignment requirement consists of two parts. The first part is explicitly expressed by (8b) and (8c). The second part is the validation of the assumption used in the above derivation for many times, i.e. the phase shifts introduced to the signal of interest by the auxiliary path can be ignored.

The first part expressed by (8b) and (8c) originates from the intrinsic working principle of the proposed method. In this method, fundamental outputs of the mixer are used to produce IM3 terms for the compensation, along with which their phases shifts (\( \Phi_1, \Phi_2 \)) are introduced to IM3 terms, too, as explained in (5). Thus, from (8b) and (8c), it appears that only if the phases of generated IM3 need to be equal to that of their counterpart (\( \Phi_3, \Phi_4 \)) from the main path, a complete cancellation can be achieved.

(8b) and (8c) can be met spontaneously if the two input tones are located close to one another. Although \( \Phi_1 \sim \Phi_4 \) of the mixer depends on many factors, such as working frequency, nonlinearities, and the loading effects at the mixer.
output, the four tones of the output located at \( \omega_1, \omega_2, \omega_{IM3,L}, \) and \( \omega_{IM3,H} \) experience approximately equal phase shifts regardless of the circuit topology if the two input tones are located close to one another, i.e. \( \Phi_1 \approx \Phi_2 \approx \Phi_3 \approx \Phi_4 \). With the increase of two-tone space, the phase differences among \( \Phi_1 \sim \Phi_4 \) start to increase and the IM3 cancellation will thus degrade. However, the in-band blockers and even the out-of-band blockers are often relatively close to signal of interest for most of the protocols. Within the required frequency range for these protocols, the proposed technique will still work effectively.

The second part, i.e. the phase shift assumption, is reasonable because the signal of interest in the auxiliary path is in the IF band. If the bandwidth of the auxiliary path is designed properly, the phase shift introduced from the auxiliary circuitry to the signal of interest can be limited to a small range, which does not severely affect the IM3 cancellation. It will be further described by the circuit description in Section III and proved by the design example in Section IV.

C. Effects on Gain and IIP5

While producing IM3 for cancellation, the auxiliary path also generates other tones that can affect performance of the main path, such as gain and IIP5. As can be observed from both Fig. 2 and (5) in Section II-A, two tones located at \( \omega_1 \) and \( \omega_2 \) are generated along the IM3 for the cancellation. What is more, in the description of the signals at Node D, only the third-order products related to IM3 cancellation were included for simplicity of discussion. In fact, four fifth-order terms are generated at the same node simultaneously, which can be expressed as

\[
\begin{align*}
\nu_{D,IM5} &= A_{IM5} \cos(\omega_1 t + \Phi_3) + A_{IM5} \cos(\omega_2 t + \Phi_4) \\
&+ A_{IM5} \cos(\omega_{IM5,L} t + \Phi_3) + A_{IM5} \cos(\omega_{IM5,H} t + \Phi_4)
\end{align*}
\]

where

\[
A_{IM5} = \frac{1}{2} K_{m \alpha^2 \beta \alpha^3} A_0^5
\]

and

\[
\begin{align*}
\omega_{IM5,L} &= 3\omega_1 - 2\omega_2 \\
\omega_{IM5,H} &= 3\omega_2 - 2\omega_1
\end{align*}
\]

The spectrum at Node D and Node E are redrawn in Fig. 3 with all these terms considered. As can be observed, both the gain and IIP3 are affected by the adoption of the auxiliary path.

However, the effects on both the conversion gain and IIP3 can be neglected without losing accuracy. To prove it, a quantitative example is provided here. Suppose a two-tone input signal of \(-20\) dBm is applied at a baseline mixer with an IIP3 of 0 dBm. Then at the mixer output, the IM3 is 40 dBc below the fundamental tone. Since the generated IM5 tones are in the same order as the uncompensated IM3 of the main path, the largest effects of these terms on the fundamental tone is around 0.09 dB. Furthermore, \( A_{IM5} \) is smaller than \( A_{IM3} \) by the same amount of \( A_{IM3} \) to the amplitude of the fundamental tones, which is 40 dBc, too. Therefore, the effect of the IM5 tones is negligible, too.

III. KEY BUILDING BLOCKS

The proposed technique can be readily applied to the mixer in CMOS process, as all the operations required in the scheme can be implemented either by the circuit topologies reported in the former CMOS state-of-the-art [17], [22], [23], or by reusing the existing blocks in the receiver.

A. IM2 Generator

Based on (4), the IM2 generator needs to produce a low-frequency IM2 tone that is in-phase with the envelope of the input RF signals and proportional to the square of the input magnitude. Its signal flow diagram is provided in Fig. 4(a). As can be seen, the input signal first goes through a 2nd-order distorser to produce all the second-order harmonics and intermodulations, and then through a tone selector to pick out the low-frequency IM2 tone among them.

Although appearing as a separate stage in Fig. 4(a), the tone selector can be implemented within the distorser circuit.
To explain this, suppose the two-tone RF input signals, $\omega_1$ and $\omega_2$, are located at 2.01 and 2.02 GHz. Consequently, the desired IM2 tone $\omega_2 - \omega_1$ is located at 10 MHz. Consequently, other second-order tones, including $2\omega_1$, $\omega_1 + \omega_2$, $2\omega_2$ and DC, are located at 4.02, 4.03, 4.04 and 0 GHz, which can all be well distinguished from 10 MHz. Obviously, the high frequency second-order tones can be easily filtered out by the low-pass filter integrated at the load of the distorter, and the DC tone can be eliminated by the inter-stage bias circuit.

The above numerical example also reveals that the high frequency second-order tones are typically far apart from the desired IM2 tone. As a result, they can be easily filtered out without introducing significant phase shift to the tone of interest if proper pole position of the low-pass filter is selected.

"\[ i_n = g_1(v_g - v_s) + g_2(v_g - v_s)^2 + g_3(v_g - v_s)^3 + \ldots \]"

where $g_i$ represents the $i$th-order transconductance coefficient of the device, which are given by

\[ g_1 = \frac{\partial I_{DS}}{\partial V_{GS}}, \]
\[ g_2 = \frac{1}{2!} \frac{\partial^2 I_{DS}}{\partial V_{GS}^2}, \]
\[ g_3 = \frac{1}{3!} \frac{\partial^3 I_{DS}}{\partial V_{GS}^3}. \]

If the differential two-tone input signal, as in (1), is applied to the squaring circuit, the fundamental tones will be canceled at the output while the common-mode second-order tones remain and can be expressed as

\[ v_{out,2nd} = (g_{MI1} \cos\omega_1 + g_{MI2} \cos\omega_2) \times (-2g_{M1}R_1) \times [R_1/(\alpha C_1)^{-1}] \]

where $g_{MIi}$ represents the $i$th-order transconductance coefficient of $M_i$. As already explained above, with proper design of $R_1$ and $C_1$ as well as the DC block $C_4$, the desired low-frequency IM2 tone can be selected without phase shift, expressed as

\[ v_{out,2nd}|_{\omega_2 - \omega_1} = -2g_{M1}R_1A_0^2 \cos(\omega_2 - \omega_1)T. \]

Since the squaring circuit produces a single-ended output, an active balun is used to recast the low-frequency IM2 tone to the differential manner, which is composed of $M_4$, $R_2$, $R_3$ $C_2$ and $C_3$ in Fig. 4(b). As can be seen, to further attenuate the unwanted second-order tones at RF frequency, $RC$ networks are used at both outputs of the balun, too. The transfer functions of the balun for the two paths are approximately

\[ V_{dd} \]

\[ V_{ss} \]

\[ V_0^- \]

\[ V_0^+ \]

\[ R_1 \]

\[ R_2 \]

\[ R_3 \]

\[ M_1 \]

\[ M_2 \]

\[ M_3 \]

\[ M_4 \]

\[ V_6 \]

\[ M_5 \]

\[ M_6 \]

\[ M_7 \]

\[ v_{x+} \]

\[ v_{x-} \]

\[ v_{y+} \]

\[ v_{y-} \]

\[ v_{y} \]

\[ v_o \]

\[ v_{o+} \]

\[ v_{o-} \]

\[ K_m = -\mu n C_{ox} \left( \frac{W}{L} \right) M_i R_j \]
in which \((\frac{W}{L})_{M_i}\) represents the dimension of squaring transistors \(M_1 \sim M_4\) and \(R_i\) represents the load resistance \(R_1\) and \(R_2\).

Similar to \(IM_2\) generator, the phase shift introduced to the output \(IM_3\) tone by the multiplier can be neglected, too. For the proposed topology, the dominant pole is located at the output, which consists of the load resistors (\(R_1\) and \(R_2\)) and the parasitic capacitor at the same node. With sufficient current assigned to the multiplier, dominant pole frequency can be far from the that of output signal, and thus brings negligible phase shift. This analysis will be further proven by the simulation results provided in Section IV-B.

**C. Signal Combiner**

Based on (7), the combiner adds output signal of the mixer and that of the auxiliary path to realize the \(IM_3\) cancellation. Because the combiner is the subsequent stage to the mixer, its linearity requirement is stringent to prevent more \(IM_3\) terms from being generated during the signal combination. To quantitatively acquire the linearity requirement, suppose the proposed technique is applied to a mixer with 10-dB gain and 0-dBm \(IIP_3\). Even if the generated \(IM_3\) tones from the auxiliary path are identical in amplitude and 180° out of phase to the \(IM_3\) of the mixer, the combiner still requires an \(IIP_3\) of as large as 20 dBm to achieve an 10-dBm \(IIP_3\) improvement for the mixer.

Op-amp-based adder is adopted here to function as the signal combiner, because it demonstrates superior linearity performance resulting from the use of strong feedback. As shown in Fig. 6, suppose two sets of differential signals, denoted as \(v_x\) and \(v_y\), are applied at the input, the output of the combiner can be expressed as

\[
v_{out} = -R_f(\frac{1}{R_1}v_x + \frac{1}{R_2}v_y).
\]

If \(R_1 = R_1 = R_2\) is fulfilled, the operation expressed in (7) is obtained. Obviously, the weights of \(v_x\) and \(v_y\) can be adjusted during combination by controlling the ratio of \(R_1\) to \(R_2\).

Although the adoption of an op-amp-based adder may seem more power-intensive compared to a current-mode transconductance-based adder, it is actually power-efficient from a system perspective, since the op-amp-based adder can be realized through reusing the subsequent stage of the mixer in a practical receiver design. In a typical receiver architecture [24], the mixer is usually followed by IF amplifiers or filters made of op-amp-based circuits, which can be readily modified to realize the signal combination without affecting their original functions.

**IV. CHIP DESIGN DETAILS**

As a proof of concept, the proposed linearization technique is applied to a 2-GHz Gilbert mixer to improve its \(IIP_3\) performance. The detailed schematic of the proposed circuit is shown Fig. 8.

**A. Mixer Design**

The mixer to be linearized adopts a Gilbert cell configuration with its tail current source omitted, as shown in Fig. 8. Although losing some common mode rejection capability due to the absence of the tail current source, this mixer can work under lower supply voltage, and is thus widely used in low voltage applications [18], [25], [26].

This mixer topology is adopted as it is a good example to show the versatility of the proposed technique. The implementation of many linearization techniques relies on the configuration of the mixer. For instance, the \(IM_2\) injection method [16], [17] needs to inject \(IM_2\) distortion into the tail current source, and thus can be hardly applied to the mixer used in this work. However, the proposed scheme can be used to linearize this mixer without a problem, as its application does not depend on the mixer configuration.
The mixer simulates a conversion gain of 8.7 dB and an IIP3 of 2.9 dBm. Translating these performance specifications to the parameters used in the derivation of Section II-A, it can be found that $A_{CG}$ and $a_3$ are equal to 2.7 and 18.5, respectively.

**B. Auxiliary Path Design**

1) **IM$_2$ Generator:** In addition to the circuit configuration proposed in Section III-A, a variable gain amplifier (VGA) is supplemented to adjust the amplitude of IM$_2$ tone, forming the IM$_2$ generator used in the chip, as can be seen in Fig. 8. Two goals are mainly considered during the design. First, the second-order coefficient of IM$_2$ Generator ($a_2$ in Section II-A) with large tuning range is desired. Second, phase shift introduced to the IM$_2$ tone should be negligible over a reasonable two-tone spacing frequency.

Before demonstrating the design details, to provide an intuitive illustration of IM$_2$ Generator’s working principle, the transient waveforms of its input and output are provided in Fig. 9. In this plot, two-tone signals are applied at the input with the tones located at 2.25 and 2.35 GHz, respectively, both with a strength of $-22$ dBm. Comparing the input and output waveforms, it can be observed that the generated IM$_2$ tone is in-phase with the envelop of input signals.

The second-order coefficient $a_2$ can be adjusted through tuning the bias of squaring devices (M7, M8) and controlling the gain of the VGA. With M7 and M8 designed to be 24-$\mu$m/0.12-$\mu$m, $g_2$ versus $V_{GS}$ are plotted in Fig. 10. To obtain a wide tuning range for $g_2$, the transistors are biased at around the middle of slope (0.4 V), at which $g_2$ of around $0.018 A/V^2$ is acquired. The VGA, which is realized by a differential common-source amplifier with resistive source degeneration, provides 8 gain steps from $-4.9$ dB to 2.8 dB with a step size of approximately 1.1 dB. Based on these settings, $a_2$ is equal to 4.80 with the gain of VGA set to $-1.6$ dB. Taking the tuning capability of the VGA into consideration, $a_2$ ranges from 3.16 to 8.15.
The phase shift introduced to the desired IM2 tone mainly results from filters used to attenuate the unwanted IM2 tones at other frequencies. Along with the signal path of IM2 tone in the IM2 generator, there are three first-order low-pass filters located at the load of each stage (consisting of \(R_7\) and \(C_7\), \(R_8,9\) and \(C_8,9\), \(R_{17-20}\) and the parasitic capacitors at the same nodes); and three high-pass filters at the biasing circuit of each stage (consisting of \(R_5\) and \(C_5\), \(R_{10}\) and \(C_{10}\), \(R_{11,12}\) and \(C_{11,12}\), \(R_{13,14}\) and \(C_{13,14}\)). In order to get negligible IF IM2 tones over a wide frequency range, the poles of low- and high-pass filters are chosen to be 1.5 GHz and 30 kHz respectively. Consequently, the phase shift on the IM2 of interest is less than 10° when the two-tone spacing is in the range of 1 to 50 MHz, as indicated in Fig. 11. At the same time, an attenuation of 9 dB on the unwanted second-order harmonics at around 4 GHz can be realized. The pole position can be adjusted to meet the different two-tone bandwidth requirements, which in turn causes changes in out-of-band suppression and power consumption as a trade off.

In addition, the balance of the active balun in IM2 generator is easy to achieve, because IM2 tone of interest is located at relatively low frequency. With \(R_8\) and \(R_9\) designed to be equal, and \(C_9\) to be 30 fF larger than \(C_8\), the positive and negative output of the active balun are well balanced. The amplitude and phase imbalance of the two outputs are plotted in Fig. 12. As can be seen, the amplitude and phase imbalances are limited into 0.005 dB and 0.15° over the two-tone spacing frequency 0 to 50 MHz. It can be safely assumed that the two branch are perfectly balanced within the operating bandwidth.

2) Baseband Multiplier: The multiplier used in the chip adopts the topology proposed in Section III-B, as seen in Fig. 8. Two goals are mainly considered during the design. First, good linearity is desired to be guaranteed, so that the amplitude of output IM3 tone can be proportional to the cube of the input RF signal’s amplitude. Second, the phase shift introduced by the multiplier should be small.

Two measures are taken to enhance the linearity of multiplier. First, proper transistor sizes are selected to suppress the source of nonlinearity. For the adopted topology, the main contributor of the nonlinearity is gate-to-source voltage variation of the source followers when the current of squaring transistors varies. Thus, the transistor sizes of the source followers (\(M_{17-18}\)) are chosen to be 60-\(\mu\)m/0.12-\(\mu\)m, which are much larger than that of the squaring circuits (\(M_{19-22}\), 30-\(\mu\)m/0.12-\(\mu\)m). In this way, the gate-to-source voltage variation of the source followers is less affected by the the current through the squaring transistor. Second, from systematic perspective, the mixer output signal is attenuated by 12 dB attenuation through a voltage divider before arriving at the multiplier, as illustrated in Fig. 8, which prevents the multiplier from being saturated by large blockers at the input.

As a result, the generated IM3 tone is proportional to the cube of the input RF signal’s amplitude in a large power range. Fig. 13 plots the power of IM3 at multiplier output versus RF input of the chip. As can be observed, the output power increases at a slope of 3dB/dB when the input power is in a reasonable range. As the input power keeps increasing, the curve starts to compress and reach 1 dB compression point eventually when the input power is up to \(-13\) dBm. However this is already sufficient large for the in-band blockers in most of the wireless receiver applications.

To guarantee the negligible phase shift, the dominant pole of the multiplier is designed to be as large as 1050 MHz. The phase shift introduced by the dominant pole of this value is plotted in Fig. 14. As can be seen, the phase shift is less
D. Simulated IIP3 Improvement on PVT Variation

\[ \Delta \text{IIP}_3 \text{ (dB) w/o calibration} = 12.0 \begin{pmatrix} 
TT \\
SS \\
FF \\
FS \\
SF 
\end{pmatrix} \]

\[ \Delta \text{IIP}_3 \text{ (dB) w/ calibration} = 12.0 \begin{pmatrix} 
011 \\
100 \\
010 \\
010 \\
100 
\end{pmatrix} \]

\[ \Delta \text{IIP}_3 \text{ (dB) w/ calibration} = 12.0 \begin{pmatrix} 
3.2 \\
12.3 \\
12.0 \\
14.1 \\
19.0 
\end{pmatrix} \]

than 6°, even when the output frequency is up to 200 MHz, and only less phase error is thus introduced when the frequency of output tone is smaller. Thus, phase error from multiplier can barely degrade the cancellation. Taking 12 dB of attenuation into consideration, the equivalent \( K_m \) gotten in this design is around 2.3.

C. Combiner Design

In this work, IIP3 based on two-tone test is the main indicator to demonstrate the effectiveness of the proposed linearization technique. Thus, to get an accurate two-tone test result, the combiner needs to demonstrate a flat gain over the frequency that covers all four output tones. Here, two-tone test result are expected to be observed for a two-tone space from 1 to 20 MHz. With the IF set at 25 MHz for the mixer, a flat gain of the combiner from 1 MHz to 65 MHz is needed. To meet this requirement, a fully differential, two-stage op amp configuration is adopted. A DC gain of 55 dB and a GBW of 430 MHz is designed for the op amp. With \( R_{27,32} \) set to 4 kΩ, a unit gain is achieved with less than 0.2-dB drop up to 80 MHz.

D. Simulated IIP3 Improvement on PVT Variation

While a considerable IIP3 improvement is achieved by the design described above, the amount of the improvement can be affected by the PVT variations.

Simulations are conducted to examine the effects of PVT variations, and the results are summarized in Tables I and II. In each table, IIP3 improvements with the VGA configured for the optimal performance of the nominal condition (TT corner and 27 °C) are listed first. Then the results with VGA tuned to calibrate the chip are demonstrated to explore the best possible performance under different conditions. In these simulations, current mirrors with the temperature-independent constant reference current sources are used to bias all the building blocks in the chip.

As revealed by the amplitude alignment condition expressed in (8a) as well as (18), the proposed technique relies on the

\( I-I-V \) characteristics of the active devices and their derivatives \( g_m, g_2 \) and \( g_3 \). For SS and FF corners, \( V_{th} \) drifts around ±20 mV from its typical value of 420 mV, which causes inaccuracies of the values for the corresponding \( I, g_m, g_2 \) and \( g_3 \). As a consequence, the amplitude alignment and the IIP3 improvement degraded, as shown in Tables I. However, at least 6.7 dB of IIP3 improvement can still be achieved under different corners. What is more, with the assistance of the VGA, the IIP3 improvement can be brought back to above 10 dB.

The variations in the resistor and the capacitor would inevitably cause both phase and amplitude variations on the IM3 tones and thus degrade the IIP3 improvements. First, the amplitude alignment is affected by the variations in the load resistors of each building block. Simulations are conducted with the resistors adjusted to +3σ, which corresponds to 11% variation from their typical values. The results shows that with the two-tone spacing located at the 4 MHz, i.e. the middle band of the auxiliary path (effects of the resistors to the phase alignment can be neglected), the IIP3 improvements degrades from 12 dB to 11 dB. Small degradation is observed since the resistors are adopted as the load for the circuits in both the main and the auxiliary path, and the effects from the resistor variations can cancel to each other to some degree.

Second, the phase alignment is also affected by the variations of both the resistors and the capacitors. Simulation shows that with both the resistors and MIM capacitors adjusted to +3σ (corresponding to about 11% variation from their typical value), the simulated optimal two-tone spacing for the IIP3 improvement shifts from 4 MHz to 2 MHz. Additionally, at the two-tone spacing of 1 MHz, the IIP3 improvement increases by 9.7 to 11.6 dB, while at 20 MHz, it decreases from 10.5 dB to 9.7 dB.

A Monte Carlo simulation using a realistic production variation model for device mismatches and process spread was also conducted. The IIP3 performance of the mixer after 200 simulation runs is presented in Fig. 15. As can be seen, a mean IIP3 of 12.52 dBm (nominally 14.5 dBm) is obtained, which is about 10 dB larger than the mixer without linearization. Furthermore, more than 80% of the results achieve IIP3 of more than 10 dBm.

In the 0.13μm process used in this design, the temperature coefficient of the poly resistors is around \(-1.17 \times 10^{-3}\), and \( |V_{th}| \) for the active MOSFET changes from 420 mV to 480 mV from 27 °C to 80 °C. The IIP3 improvements as a function of the temperature are listed in Tables II. As can be observed, the IIP3 improvements degrades as the temperature moves to extremes. However, with the calibrated of the VGA, the IIP3 improvements degradation can be brought back.
Fig. 15. Monte-Carlo simulation results of mixer $IIP_3$ after 200 runs.

Fig. 16. Micrograph of the mixer die.

corresponding control bits of the VGA used under different temperatures are listed.

V. MEASUREMENT RESULTS

The mixer chip was fabricated using GlobalFoundries’s standard 130 nm CMOS process. The chip occupies an area of $1.2 \times 2 \text{ mm}^2$, in which the active area is $0.3 \times 1.4 \text{ mm}^2$. The micrograph of the chip is shown in Fig. 16. Testing was carried out on-chip using differential GSGSG coplanar probes at the RF, IF and LO ports. For conversion gain, noise figure and intermodulation distortion measurements, a Keysight PSA E4446A spectrum analyzer was used along with Anritsu MG3694A sources.

For comparison, the auxiliary path is turned on and off to enable and disable the cancellation. In both on and off mode, the IF amplifier (combiner) is in series to the mixer. As the IF amplifier is a unit-gain amplifier with a simulated $IIP_3$ of above 35 dBm, its effects to the gain and $IIP_3$ of the mixer in both cases are negligible during the measurement.

The mixer is measured with an LO of $-2 \text{ dBm}$ located at 2 GHz. The mixer and the IF amplifier consume 4 mA and 4.4 mA under 1.2V voltage, separately, and the auxiliary path consumes 4.2 mA when it is turned on.

With two RF signals located at 2.025 GHz and 2.035 GHz applied, a conversion gain of 8.5 dB is obtained and almost unchanged when the linearization part turns on and off, as can be seen in Fig. 17. Furthermore, $IIP_3$ of the baseline mixer is 2.5 dBm and the proposed linearization technique can improve it by 12 dB. The simulation results are provided in the same plot, indicating a good match between the simulation and measurement results.

In the derivation presented in Sec. II, all the operations are considered linear, which indicates that the proposed technique is seemingly able to function in all input power levels. However, when the input power is beyond certain level, the circuit blocks in the linearization technique can be saturated, which leads to inaccurate IM3 cancellation and thus $IIP_3$ degradation. As can be observed in Fig. 17, when the input power is larger than $-15 \text{ dBm}$, the IM3 suppression becomes less effective.

Fig. 18 provides the output spectrum of the mixer before and after linearization, clearly showing the IM3 suppression. Additionally, the spectrum also shows that the gain is not affected by the linearization technique. With the aforementioned frequency setting, the IM2 tone of the mixer output lies left to down-converted IF by 5 MHz, as shown at the leftmost
Table III

<table>
<thead>
<tr>
<th></th>
<th>This Work</th>
<th>[1]</th>
<th>[6]</th>
<th>[14]</th>
<th>[16]</th>
<th>[18]</th>
<th>[27]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF frequency (GHz)</td>
<td>2.0</td>
<td>23−25</td>
<td>1.0</td>
<td>0.5−6.5</td>
<td>2.1</td>
<td>0.9</td>
<td>20−50</td>
</tr>
<tr>
<td>Gain/ΔGain (dB/dB)</td>
<td>8.5/0.0</td>
<td>−4</td>
<td>11/0.7</td>
<td>10/−</td>
<td>15/0.0</td>
<td>17.6/5.0</td>
<td>0/−</td>
</tr>
<tr>
<td>IIP3/ΔIIP3 (dBm/dB)</td>
<td>14.5/+12.0</td>
<td>23/+9</td>
<td>5/+14</td>
<td>9.52/−</td>
<td>15.0/+10.5</td>
<td>11.8/+10.0</td>
<td>9.5/+5.6</td>
</tr>
<tr>
<td>NF/ΔNF (dB/dB)</td>
<td>17.9/+0.2</td>
<td>N/A</td>
<td>12.9/0.0</td>
<td>13/−</td>
<td>14.0/0.0</td>
<td>10.1/+0.1</td>
<td>16/−</td>
</tr>
<tr>
<td>Supply voltage (V)</td>
<td>1.2</td>
<td>2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.8</td>
<td>1.8</td>
<td>1.2</td>
</tr>
<tr>
<td>I_{dc}/ΔI_{dc} (mA/mA)</td>
<td>12.6/+/4.2</td>
<td>8/+</td>
<td>17/+2</td>
<td>4.5/+</td>
<td>4.5/+0.5</td>
<td>10.9/+1.7</td>
<td>5/+1.5</td>
</tr>
<tr>
<td>Chip area (mm²)</td>
<td>0.42</td>
<td>0.72</td>
<td>0.1</td>
<td>0.015</td>
<td>N/A</td>
<td>0.488/0.446</td>
<td>0.485</td>
</tr>
</tbody>
</table>

1 simulation results only
2 full front-end, including an LNA and a mixer
3 NF of the mixer plus IF amplifier
4 the mixer, IF amplifier and linearization circuits consume 4 mA, 4.4 mA and 4.2 mA, respectively
5 includes probe pads

Fig. 19. IIP3 improvement versus two-tone spacing for the mixer.

of the spectrum in Fig. 18. As can be seen, the proposed method does not affect the amplitude of IM2 tone. Thus, it does not affect the IIP2 performance. What is more, it does not interfere other linearization technique to suppress the IM2 tone either.

Fig. 19 shows the IIP3 of the linearized mixer as a function of two-tone spacing with the input power at −25 dBm. The simulation results are also provided as comparison in the same plot. As can be seen, the optimal point is shifted from 4 MHz in simulation to 7 MHz in measurement. After the optimal point the IIP3 improvement slowly degrades as the two-tone spacing increases. Nevertheless, over a large two-tone space range from 1 MHz to 22 MHz, the IIP3 improvement is above 10 dB.

The proposed linearization technique only deteriorates NF of the system by a small amount. When the linearization circuitry is turned off, NF of the mixer in series to combiner is measured to be 17.7 dB at the IF of 25 MHz. When the auxiliary path is turned on, the noise figure only increases by 0.2 dB.

Table III summarizes the measured performance of the mixer in both unlinearized and linearized cases. A performance comparison with other state-of-the-art linearization techniques are also demonstrated in the same table. Several observations can be made to prove the benefits of the proposed method. First, a competitive IIP3 is achieved with lowest supply voltage compared to former state-of-art. For active mixers, a key limitation to the linearity is the supply voltage, which limits the output voltage headroom. Second, although the largest extra current (4.2 mA) seems to be consumed by the proposed scheme, the current consumption is actually determined by the bandwidth of the auxiliary path, i.e. the two-tone space range in which the IIP3 can be improved. If less two-tone space range is needed in a practical scenario, the current consumption can be significantly reduced. Third, while realizations of most prior art relies on the topology of the mixers, the proposed scheme can actually be applied to any active mixers. Furthermore, with some modifications the technique can be applied to upconverter mixers. For such a case, a bandpass filter would be needed to select the upper sideband of the auxiliary multiplier output and the op amp combiner would be replaced by a current-summing network.

VI. Conclusion

A feedforward linearization scheme to cancel the IM3 terms of the mixer is proposed, which is indifferent to mixer topology at the cost of small amount of extra DC power and NF deterioration. Unlike the linearization technique in prior art, where an extra full receiver is used to generate IM3, IM3 for cancellation is generated through the multiplication of low-frequency IM2 signals generated from a squaring circuit and the IF fundamental signal of the mixer output. Consequently, the power consumption of the circuitry in the auxiliary path is low and the cancellation through auxiliary path are immune to the parasitics of the circuit. More than 10 dB IIP3 improvement is demonstrated over a large two-tone space with negligible noise, gain degradation and small extra current.

Acknowledgment

The authors acknowledge CMC Microsystems, Kingston, Ontario, Canada, for IC fabrication services and CAD tool support.
REFERENCES


Carlos E. Saavedra (S’92–M’98–SM’05) received the B.Sc. degree from the University of Virginia and the M.Sc. and Ph.D. degrees from Cornell University, Ithaca, NY, USA, all in electrical engineering. From 1998 to 2000, he was a Senior Engineer at Millitech Corporation, Northampton, MA, USA, where he was a Lead Designer for 380-GHz point-to-point broadband wireless transceivers. In 2000, he joined the Department of Electrical and Computer Engineering, Queen’s University, Kingston, ON, Canada, where he currently holds the rank of Professor. During sabbatical leaves, he has held visiting appointments at the University of Navarra, Spain, the Universidade Federal do Rio Grande do Sul, Brazil, and the University of Twente, Holland. He served as the ECE Chair of Graduate Studies at Queen’s University from 2007 to 2010.

Hao Li (M’14) received the B.Sc. degree in micro-electronics and the M.Sc. (Eng.) degree in integrated circuit design from Xi’an Jiaotong University, Xi’an, China, in 2010 and 2012, respectively, and the Ph.D. degree in electrical engineering from Queen’s University, Kingston, ON, Canada, in 2018. His research interests are in the field of RF CMOS front-end circuits. He served as the Vice Chair for the AP-S/MTT-S Joint Chapter of the IEEE Kingston Branch from 2014 to 2015.