Noise and Nonlinearity Modeling of Active Mixers for Fast and Accurate Estimation

Wei Cheng, Anne Johan Annema, *Member, IEEE*, Jeroen A. Croon, and Bram Nauta, *Fellow, IEEE*

***Abstract*—This paper presents a model of active mixers for a fast and accurate estimation of noise and nonlinearity. Based on closed-form expressions, this model estimates the noise figure, IIP3, and IIP2 of the time-varying mixer by a limited number of time-invariant circuit calculations. The model shows that the decreasing transistor output resistance, together with the low supply voltage in deep-submicrometer technologies, significantly contributes to the flicker-noise leakage. Design insights for low flicker noise are then presented. The model also shows that the slope of the LO signal has a significant effect on IIP2, while it has a little effect on IIP3. A new IP2 calibration technique using slope tuning is presented.**

***Index Terms*—Active mixer, dc offset, flicker noise, IIP2, IIP3, intermodulation, mismatch, nonlinearity.**

I. Introduction

THE active mixer is a critical building block in the RF front-end. With a higher conversion gain, the active mixer provides a better noise suppression of the subsequent stages than passive mixers. Unfortunately, the CMOS active mixer suffers more from flicker noise and nonlinearity than the passive mixer, which degrades the overall noise and linearity performance in zero-IF and low-IF receivers [1]–[5]. For circuit design insights as well as for design automation and synthesis of the RF circuits where typically iterative dimensioning loops are involved [6], a model that enables a fast and accurate estimation of noise and nonlinearity is desirable. A number of papers present noise and nonlinearity analyses for the Gilbert mixer to provide the design guidelines [7]–[13] or to build a high-level model during the architectural design of the RF front-ends [14]. However, for noise analyses, the transistor output resistance is typically neglected [7]–[9] or oversimplified [10]. For IIP3 calculations, [11], [13], and [14] focus on numerical calculations, while [12] neglects the periodic property of the transistor nonlinearity for IIP2. In [11]–[13], for both IIP2 and IIP3 analyses, transistor nonlinearities other than the transconductance nonlinearity are neglected. In [11], [13], and [14], the effect of the LO slope is not considered.

In this paper, a time-varying small-signal and weakly nonlinear analysis is used, including both the output resistance and capacitances. It is shown that the output resistance effects may significantly contribute to the flicker-noise leakage and hence may make the flicker noise cancellation technique of tuning out the capacitance less effective [15], [16]. The effect of the finite LO slope on IIP3 can be neglected, while neglecting the LO rise/fall time can underestimate IIP2. Aiming for the circuit design guidelines as well as constructing an estimation model for the automatic synthesis of the active mixers, we introduce a closed-form model that properly models the output noise and nonlinearity of the active mixers. These closed-form expressions use linear interpolation between a limited number of time-invariant circuit calculations in one LO period. The noise model derived in this paper requires data from only two ac calculations. The IIP3 model in this paper requires one time-invariant nonlinearity calculation, while the IIP2 model requires data from a few time-invariant nonlinearity calculations. Since the time-varying mixer performance is estimated by time-invariant noise and nonlinearity calculations, this model involves no complex numerical analyses, and it can be easily utilized by circuit designers and fast mixer design automation algorithms.

Section II introduces the fundamentals of the active mixers in deep-submicrometer technologies. Section III presents the time-varying small-signal analysis for the noise model. The impact of the transistor output resistance is investigated, and the design insights for flicker-noise leakage reduction are presented. Section IV uses the time-varying weakly nonlinear analysis to derive the closed-form expressions for IIP3 and IIP2. The impact of the LO slope is analyzed for both IIP3 and IIP2, and a new IIP2 calibration technique is proposed. Section V presents the benchmarking of the accuracy for our model for the mixer operating in different bias conditions and at different frequencies. The conclusion is drawn in Section VI.

II. Active Mixer in Deep-Submicrometer Technologies

A mixer is a periodically time-varying circuit whose periodic steady state is modulated by the periodic LO signal. At any instantaneous time, the (quasi-)dc bias for the mixer is fixed, and therefore, the circuit can be linearized around this (quasi-)dc operating point. As a result, for noise analysis, the transistors within the mixer can be described by periodic small-signal parameters such as periodic transconductances, output resistances, and capacitances. For nonlinearity analyses, the transistors can be modeled by periodic weakly nonlinearities such as periodic nonlinear transconductances, output resistances, and capacitances. Note that this assumes that the transient effects are small, which is a valid simplification for mixers that operate in the low gigahertz region in modern CMOS processes. As a result of the periodic behavior, the transfer functions from the input port to the output port of the mixer can be described by periodic small-signal and weakly nonlinear properties of transistors and by time-invariant properties of passives in the circuit [17]. In the analyses in this paper,

Manuscript received March 07, 2010; revised June 15, 2010; accepted July 07, 2010. Date of publication November 18, 2010; date of current version January 28, 2011. This work was supported by NXP Semiconductors. This paper was recommended by Associate Editor A. Tasic.

W. Cheng, A. J. Annema, and B. Nauta are with the IC-Design Group, Centre for Telematics and Information Technology (CTIT), University of Twente, 7500 AE Enschede, The Netherlands (e-mail: w.cheng@utwente.nl).

J. A. Croon is with the NXP-TSMC Research Center, 5656 AE Eindhoven, The Netherlands (e-mail: jeroen.croon@nxp.com).

Digital Object Identifier 10.1109/TCSI.2010.2072110

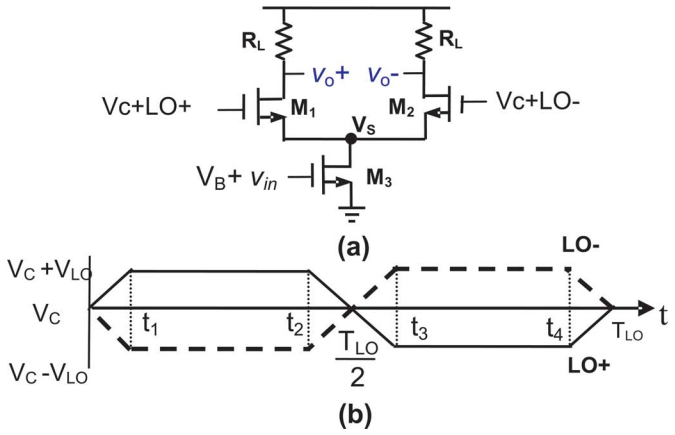

Fig. 1. (a) Schematic of the single-balanced Gilbert mixer. (b) Waveform of the LO signal.

a time-varying small-signal analysis is applied to derive a noise model, while a time-varying weakly nonlinear analysis is applied to derive a nonlinearity model. For simplicity reasons, the single-balanced Gilbert mixer shown in Fig. 1(a) is used for the analyses of noise and nonlinearity. We assume that the LO signal at the gate of the switching pair can be properly modeled by a trapezoid shown in Fig. 1(b).

In deep-submicrometer technologies, two additional issues are considered in this paper compared to previous work [7]–[13].

1) In deep-submicrometer CMOS, the output resistance of the short transistors is relatively low so that the flicker-noise contribution due to the output resistance can be as significant as that from the output capacitance. In this paper, therefore, the influence of the output resistance of M_3 on the flicker-noise leakage is taken into account.
2) At $t=0$, the gate bias of M_1 is equal to the common-mode voltage of LO (V_c). Due to the low supply voltage in the submicrometer technologies, V_c can be so low that M_3 is in the triode region. Since the drain current of M_3 is small, the voltage drop across the load is small, and M_1 is generally in saturation. As the LO+ increases, M_3 gradually enters the saturation region. During $(0, 0.5T_{\mathrm{LO}})$, M_1 stays in the saturation region, and M_3 toggles between the triode and saturation regions. Similarly, during $(0.5T_{\mathrm{LO}}, T_{\mathrm{LO}})$, M_2 stays in the saturation region, and M_3 toggles between the triode and saturation regions. In the triode region, the output conductance nonlinearity and the cross-modulation nonlinearity (e.g., $g_{x21} = (1/2) \times (\partial^3 I_{\mathrm{DS}}/\partial V_{\mathrm{GS}}^2 \partial V_{\mathrm{DS}})$) become significant. Therefore, the analyses in this paper take into account the transconductance nonlinearity as well as the output conductance nonlinearity and the cross-modulation nonlinearity for IIP2 and IIP3 modeling.

III. Time-Varying Small-Signal Noise Analysis

A. Noise Model for Active Mixers

The flicker-noise output of the Gilbert mixer is dominantly contributed by the switch pair $\mathrm{M}_1/\mathrm{M}_2$, while transistor M_3 is causing thermal-noise folding [7]–[10]. The mixer output noise can be approximated by a stationary process [18], and therefore, the output noise voltage contributed by transistor M_1, M_2, and M_3 is given by

$$\begin{aligned} v_{n,\mathrm{out}}^{M_k}[v_{\mathrm{LO}}(t)] &= H^{M_k}[v_{\mathrm{LO}}(t)] \cdot V_{n,\mathrm{in}}^{M_k}[v_{\mathrm{LO}}(t)]\, e^{j\omega_{\mathrm{in}}t} \\ &= F_n^{M_k}[v_{\mathrm{LO}}(t)] e^{j\omega_{\mathrm{in}}t}, \qquad n \in \{th, fl\} \end{aligned} \tag{1}$$

where $V_{n,\mathrm{in}}^{M_k}[v_{\mathrm{LO}}(t)]$ is the equivalent gate-referred root mean square (rms) noise voltage of transistor M_k, with either flicker noise ($n = fl$) or thermal noise ($n = th$); $v_{\mathrm{LO}}(t)$ is the LO signal; and $H^{M_k}[v_{\mathrm{LO}}(t)]$ accounts for the transfer function between the noise source to the output terminals. For an LO period T_{LO} and by assuming that M_1 and M_2 are symmetric, $|H^{M_1}[v_{\mathrm{LO}}(t)]| = |H^{M_2}[v_{\mathrm{LO}}(t + (T_{\mathrm{LO}}/2))]|$. Consequently, it is sufficient to focus on $v_{n,\mathrm{out}}^{M_1}$ and $v_{n,\mathrm{out}}^{M_3}$ in the analyses. Because of its periodic nature, the term $F_n^{M_k}[v_{\mathrm{LO}}(t)]$ in (1) can be replaced by its Fourier series, yielding

$$\begin{aligned} v_{n,\mathrm{out}}^{M_k}[v_{\mathrm{LO}}(t)] &= \sum_{p=-\infty}^{+\infty} f_{p,n}^{M_k} e^{jp\omega_{\mathrm{LO}}t} e^{j\omega_{\mathrm{in}}t} \\ &= \sum_{p=-\infty}^{+\infty} f_{p,n}^{M_k} e^{j(p\omega_{\mathrm{LO}}+\omega_{\mathrm{in}})t}, \qquad n \in \{th, fl\} \end{aligned} \tag{2}$$

where the dc term $f_{0,n}^{M_k}$ accounts for the noise leakage (from the inputs at ω_{in} to the outputs at ω_{in}), and the mth-order Fourier coefficients $f_{-m,n}^{M_k}$ account for the noise folding (from the inputs at $\omega_{\mathrm{in}} = \omega_{IF} + m\omega_{\mathrm{LO}}$ to the outputs at $\omega_{\mathrm{in}} - m\omega_{\mathrm{LO}} = \omega_{IF}$). As a result, the output noise of the down-conversion single-balance Gilbert mixer contributed by the transistors is given by

$$S_{fl,\mathrm{out}} = 2 \times \left| f_{0,fl}^{M_1} \right|^2 + \left| f_{0,fl}^{M_3} \right|^2 \tag{3}$$

$$S_{th,\mathrm{out}} = \sum_{m=-\infty}^{+\infty} \left| f_{-m,th}^{M_3} \right|^2 + 2 \times \sum_{m=-\infty}^{+\infty} \left| f_{-m,th}^{M_1} \right|^2. \tag{4}$$

In these relations, $f_{0,fl}^{M_k}$ is the dc term of $F_{fl}^{M_k}$ in (1) for the gate-referred flicker noise $V_{fl,\mathrm{in}}^{M_k}$, and $f_{-m,th}^{M_k}$ is the Fourier series coefficients of $F_{th}^{M_k}$ in (1) for the gate-referred thermal noise $V_{th,\mathrm{in}}^{M_k}$.

Assuming a symmetric LO signal, for the given rise/fall time, the driving signal $v_{\mathrm{LO}}(t)$ is determined by three time instants: t_1, $0.5T_{\mathrm{LO}}$, and t_4. The waveforms $F_n^{M_1}[v_{\mathrm{LO}}(t)]$ and $F_n^{M_3}[v_{\mathrm{LO}}(t)]$ at t_1, $0.5T_{\mathrm{LO}}$, and t_4 can be easily obtained from the time-varying small-signal models shown in Fig. 2. To avoid solving differential equations, we simplify the analysis by modeling the time-varying small-signal capacitance in transistor M_k with a time-varying admittance $j\omega_{\mathrm{LO}}C(t)$. In this paper, we focus on demonstrating this time-varying noise analysis. Therefore, we only include the transconductance of M_1, M_2, and M_3 and the output impedance of M_3 ($Z_{\mathrm{ds}}^{M_3}(t) = r_{\mathrm{ds}}^{M_3}(t)/(1 + j\omega_{\mathrm{LO}} C_{\mathrm{ds}}^{M_3}(t) r_{\mathrm{ds}}^{M_3}(t))$), as shown in Fig. 2(a). Nevertheless, for highly accurate noise modeling used in the mixer design automation, all capacitances and conductances of the transistor are taken into account. This yields

$$F_n^{M_1}\big|_{t_1} = \left. \frac{-g_m^{M_1} R_L v_{n,\mathrm{in}}^{M_1}}{1 + g_m^{M_1} Z_{\mathrm{ds}}^{M_3}} \right|_{t_1} \approx \left. \frac{-R_L v_{n,\mathrm{in}}^{M_1}}{Z_{\mathrm{ds}}^{M_3}} \right|_{t_1} \tag{5}$$

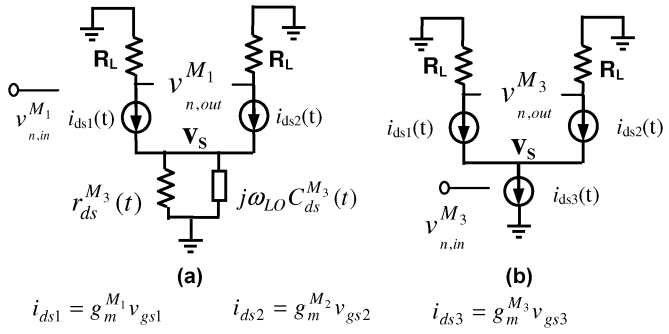

Fig. 2. Time-varying small-signal model for calculating (a) $F_n^{M_1}[v_{\rm LO}(t)]$ and (b) $F_n^{M_3}[v_{\rm LO}(t)]$.

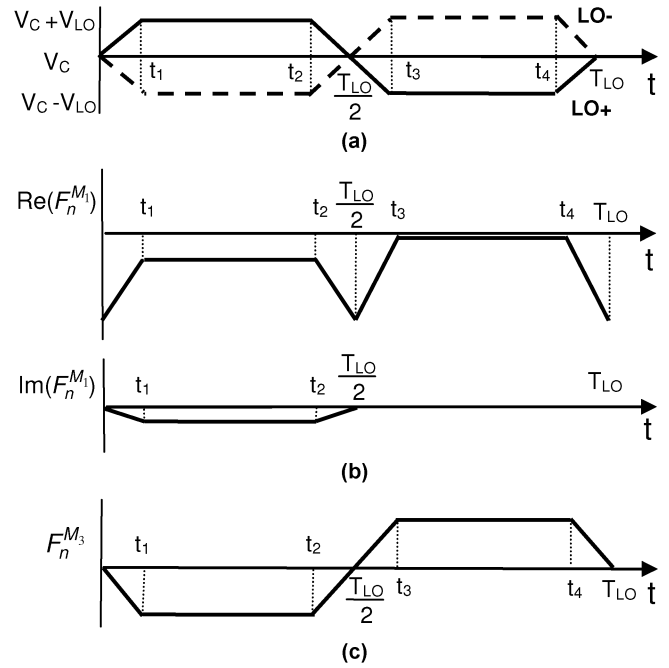

Fig. 3. (a) Waveform of the LO signal $v_{\rm LO}(t)$. (b) Approximation of $F_n^{M_1}$. (c) Approximation of $F_n^{M_3}$.

$$F_n^{M_1}\Big|_{\frac{T_{\rm LO}}{2}} = -\,g_m^{M_1}R_L v_{n,\rm in}^{M_1}\Big|_{\frac{T_{\rm LO}}{2}} \tag{6}$$

$$F_n^{M_1}\Big|_{t_3} \approx 0 \tag{7}$$

$$F_n^{M_3}\Big|_{t_1} = \frac{-g_m^{M_1}g_m^{M_3}R_L Z_{\rm ds}^{M_3}v_{n,\rm in}^{M_3}}{1+g_m^{M_1}Z_{\rm ds}^{M_3}}\Big|_{t_1} \approx -g_m^{M_3}R_L v_{n,\rm in}^{M_3}\Big|_{t_1} \tag{8}$$

$$F_n^{M_3}\Big|_{\frac{T_{\rm LO}}{2}} = 0 \tag{9}$$

$$F_n^{M_3}\Big|_{t_3} = \frac{g_m^{M_1}g_m^{M_3}R_L Z_{\rm ds}^{M_3}v_{n,\rm in}^{M_3}}{1+g_m^{M_1}Z_{\rm ds}^{M_3}}\Big|_{t_3} = -\,F_n^{M_3}\Big|_{t_1}. \tag{10}$$

For the mixers in the modern deep-submicrometer CMOS and by operating at frequencies of up to the lower gigahertz range, the transient effects can be neglected. Then, $F_n^{M_k}[v_{\rm LO}(t)]$ can be sufficiently accurately approximated by interpolating between $F_n^{M_k}[v_{\rm LO}(t_1)]$, $F_n^{M_k}[v_{\rm LO}(0.5T_{\rm LO})]$, and $F_n^{M_k}[v_{\rm LO}(t_3)]$. These approximations are shown in Fig. 3(b) and (c).

At (t_1, t_2), $\rm M_1$ and $\rm M_3$ form a cascode amplifier, and in this period, the drive voltage of $\rm M_1$ ($v_{\rm LO}$) has its maximum value. Then, $F_n^{M_3}|_{t_1}$ is equal to the output noise voltage due to the (equivalent) input referred noise of $\rm M_3$. Because of the finite output impedance of $\rm M_3$ in deep-submicrometer CMOS, the noise contribution from the cascode transistor $\rm M_1$, given by (5), cannot be neglected.

At $0.5T_{\rm LO}$, both $\rm M_1$ and $\rm M_2$ are on, and they form a balanced differential pair. Then, the output impedance of $\rm M_3$ has a negligibly small effect on $F_n^{M_1}$, as shown by (6).

At (t_3, t_4), $\rm M_2$ and $\rm M_3$ act as a cascode amplifier, and $\rm M_1$ is off. Thus, $F_n^{M_1}$ is close to zero, and $F_n^{M_3}$ is at its positive maximum. Being an odd function, $F_n^{M_3}[v_{\rm LO}(t)]$ has no even Fourier series coefficients, and thus, for transistor $\rm M_3$, the flicker noise only up-converts to the sidebands around the odd harmonics of the LO. The thermal noise at the sidebands around the odd harmonics of the LO frequency folds back to the IF band. As for $\rm M_1$, the dc term of $F_n^{M_1}[v_{\rm LO}(t)]$ accounts for the noise at the output without frequency translation, which corresponds to the flicker-noise leakage. The thermal noise at the sidebands around the harmonics of the LO frequency folds back to the IF band.

Assuming a symmetrical LO signal, with the rise/fall time equal to $\alpha \cdot T_{\rm LO}$, the time instants t_1 to t_4 can be rewritten as $t_1 = 0.5{\cdot}\alpha T_{\rm LO}, t_2 = 0.5{\times}(1{-}\alpha){\cdot}T_{\rm LO}, t_3 = 0.5{\times}(1{+}\alpha){\cdot}T_{\rm LO}$, and $t_4 = (1-0.5\alpha)\cdot T_{\rm LO}$. Again, under the assumption of negligibly small transient effects, this enables rewriting (3) and (4) into

$$\begin{aligned} S_{fl,\rm out} &= 2\times\left|f_{0,fl}^{M_1}\right|^2 \\ &= \frac{2}{T_{\rm LO}^2}\times\Bigg[(t_2-t_1)\cdot F_{fl}^{M_1}\Big|_{t_1} \\ &\qquad + t_1\cdot\left(\left|F_{fl}^{M_1}{}_{\frac{T_{\rm LO}}{2}} - F_{fl}^{M_1}\right|_{t_1}\right) \\ &\qquad + \left(t_4-\frac{T_{\rm LO}}{2}\right)\cdot F_{fl}^{M_1}\Big|_{\frac{T_{\rm LO}}{2}}\Bigg]^2 \\ &= \frac{1}{2}\left(F_{fl}^{M_1}\Big|_{t_1} + 2\cdot\alpha\cdot F_{fl}^{M_1}\Big|_{\frac{T_{\rm LO}}{2}} \right. \\ &\qquad \left. - 3\cdot\alpha\cdot F_{fl}^{M_1}\Big|_{t_1}\right)^2 \end{aligned} \tag{11}$$

$$\begin{aligned} S_{th,\rm out} &= \sum_{m=-\infty}^{+\infty}\left|f_{-m,th}^{M_3}\right|^2 + \sum_{m=-\infty}^{+\infty}\left|f_{-m,th}^{M_1}\right|^2 \\ &= \frac{1}{T_{\rm LO}}\int_{T_{\rm LO}}\left|F_{th}^{M_3}\right|^2 dt + \frac{1}{T_{\rm LO}}\int_{T_{\rm LO}}\left|F_{th}^{M_1}\right|^2 dt \\ &= \frac{3-4\alpha}{3}\left(F_{th}^{M_3}\Big|_{t_1}\right)^2 + \frac{3-4\alpha}{6}\cdot\left(F_{th}^{M_1}\Big|_{t_1}\right)^2 \\ &\quad + \frac{\alpha}{6}\cdot\Bigg[2F_{th}^{M_1}\Big|_{t_1}\cdot F_{th}^{M_1}\Big|_{\frac{T_{\rm LO}}{2}} \\ &\qquad +4\cdot\left(F_{th}^{M_1}\Big|_{\frac{T_{\rm LO}}{2}}\right)^2\Bigg]. \end{aligned} \tag{12}$$

Given that the total output noise mainly consists of the flicker noise from $\rm M_1$ and $\rm M_2$, the thermal noise from $\rm M_1$, $\rm M_2$, and $\rm M_3$,

the load R_L, and the input source impedance R_s, the single-side band noise figure (NF) is given by

$$NF_{SSB} = \frac{S_{fl,\text{out}} + S_{th,\text{out}} + S_{Rs,\text{out}} + S_{RL,\text{out}}}{0.5S_{Rs,\text{out}}} \quad (13)$$

where $S_{R_{\text{load}},\text{out}} \approx 8kTR_{\text{load}}$. While the input source resistance Rs contributes to the output noise in the same way as the thermal noise from M_3, (8) and (12) yield $S_{Rs,\text{out}} \approx (1-(4\alpha/3)) \cdot (g_m^{M_3} R_L|_{t_1})^2 kTR_s$, with $v_{n,\text{in}}^{M_3}$ in (8) replaced by $\sqrt{kTR_s}$ for perfect input matching. Note that (11) and (12) use only the transistor properties, bias conditions, and component values at two distinct time instances: at t_1 and at $0.5T_{\text{LO}}$. As a result, the presented estimation for the active mixer NF can be realized by two ac calculations for the trapezoid LO signals with finite rise and fall times.

B. Impact of the Transistor Output Impedance on the Flicker Noise

In the previous section, (11) indicates that the impact of the output impedance on the flicker-noise leakages is described by the dc term of $F_n^{M_1}$ $(n = fl)$. By using (5) and (6), (11) is simplified to

$$S_{fl,\text{out}} \approx 0.5 \times \left[\left. \frac{(3\alpha-1) \cdot g_m^{M_1} R_L v_{fl,\text{in}}^{M_1}}{1 + g_m^{M_1} Z_{\text{ds}}^{M_3}} \right|_{t_1} - 2\alpha \, g_m^{M_1} R_L v_{fl,\text{in}}^{M_1} \Big|_{\frac{T_{\text{LO}}}{2}} \right]^2 \quad (14)$$

where the former term is the integral (or area) of $F_n^{M_1}$ in (t_1, t_2), shown in Fig. 2(b), and the latter term is the integral of $F_n^{M_1}$ in (t_2, t_3), $(0, t_1)$, and (t_4, T_{LO}). Then, (13) can be simplified to the following:

1) for low IF (flicker noise dominant)

$$\begin{aligned} NF_{SSB} &\approx \frac{S_{fl,\text{out}}}{0.5S_{Rs,\text{out}}} \\ &= \frac{1}{\left(1-\frac{4\alpha}{3}\right) KTR_s} \\ &\quad \times \left[\left. \frac{(3\alpha-1) \cdot R_L v_{fl,\text{in}}^{M_1}}{g_m^{M_3} Z_{\text{ds}}^{M_3}} \right|_{t_1} - \frac{2\alpha \, g_m^{M_1} v_{fl,\text{in}}^{M_1} \Big|_{\frac{T_{\text{LO}}}{2}}}{g_m^{M_3}\Big|_{t_1}} \right]^2 \end{aligned} \quad (15)$$

2) for high IF (thermal noise dominant)

$$\begin{aligned} NF_{SSB} &\approx \frac{S_{th,\text{out}} + S_{Rs,\text{out}} + S_{RL,\text{out}}}{0.5S_{Rs,\text{out}}} \\ &= 2 \cdot \left[\frac{8}{\left(1-\frac{4\alpha}{3}\right) \cdot {g_m^{M_3}}^2 R_L R_s} + \frac{4\gamma}{g_m^{M_3} R_s} \right] \Bigg|_{t_1}. \end{aligned} \quad (16)$$

Two flicker-noise leakage mechanisms are represented by (14). At $(0, t_1)$, (t_2, t_3), and (t_4, T_{LO}), the mixer acts as a differential pair. The flicker noise of the switch pair is transferred to the output just like the signal amplified by the differential amplifier. For this mechanism, the output impedance of M_3 has no effect on the flicker-noise leakage, which is shown by the second term in (14). At (t_1, t_2) and (t_3, t_4), one transistor in the switch pair is off, and the mixer acts as a cascode amplifier. Due to the finite output impedance of M_3, the flicker noise of the switch pair leaks to the output.

In summary, (14) suggests that the slope of the LO, the gain of the differential pair, the input-referred flicker-noise voltage of the switch M_1/M_2, and the output impedance of M_3 all determine the flicker-noise leakage. The following approaches can be followed to reduce the flicker-noise leakage.

1) Reducing the rise/fall time of the LO signal (smaller α) decreases the area of the spike in (t_2, t_3), $(0, t_1)$, and (t_4, T_{LO}).
2) Choosing a wider switch pair (smaller $v_{n,fl}^{M_1}$), which comes at the price of higher LO power.
3) Reducing $g_m^{M_1}|_{T_{\text{LO}}/2}$ so that the height of the spike in (t_2, t_3) decreases. This can be realized by reducing the bias current of the switch pair at $0.5T_{\text{LO}}$ (smaller $g_m^{M_1}$) [15], [16] and [19] and by choosing a low common-mode voltage V_c for the LO. At $0.5T_{\text{LO}}$, a low V_c can force M_3 into the triode region, which reduces the dc current of M_1 and M_2, resulting in a decrease of $g_m^{M_1}|_{T_{\text{LO}}/2}$.
4) Increasing the output impedance of M_3 (larger $Z_{\text{ds}}^{M_3}$).

In technologies with long-channel transistors where the output capacitance of M_3 is dominant in $Z_{\text{ds}}^{M_3}$, the output resistance $r_{\text{ds}}^{M_3}$ can be neglected [7]–[9]. To increase $Z_{\text{ds}}^{M_3}$, then the inductors can be used to tune out the output capacitance for flicker-noise reduction [15], [16]. However, nowadays, the technology scaling offers f_T figures well above 100 GHz, while it also brings a lower output resistance and a lower supply voltage [20]. Neglecting the effect of $r_{\text{ds}}^{M_3}$ in deep-submicrometer technologies can yield a significant underestimation of the output flicker noise. Fig. 4 shows an illustration: with an ideal square-wave LO signal at 2 GHz and IF@10 kHz (flicker noise dominant), the calculated NF is compared with the simulated results for a different bias current. The NF is calculated using (13), including both $r_{\text{ds}}^{M_3}$ and $C_{\text{ds}}^{M_3}$ (cross symbol) and including only $C_{\text{ds}}^{M_3}$ (square symbol). The simulation is performed in Spectre for a standard 90-nm CMOS process.[1] Driven by an ideal square-wave LO, the flicker-noise leakage is only caused by the finite output impedance of M_3. Including only $C_{\text{ds}}^{M_3}$, as done in [7]–[9], underestimates the flicker leakage by over 7 dB compared with the model taking into account both $r_{\text{ds}}^{M_3}$ and $C_{\text{ds}}^{M_3}$. This suggests that, for deep-submicrometer CMOS technologies with low supply voltage and low output resistance, $r_{\text{ds}}^{M_3}$ is dominant in the flicker-noise leakage rather than $C_{\text{ds}}^{M_3}$. Consequently, flicker-noise cancellation by tuning out the output capacitance is less effective in modern deep-submicrometer processes.

C. Optimum Transistor Length for Low Flicker-Noise Leakage

As discussed in Section III-B, a larger r_{ds} of M_3 reduces the flicker-noise leakage. In order to keep the same power consumption and the same transconductance g_m for M_3, both the width and length of M_3 can be increased in proportion to keep the same W/L ratio, which results in a larger $r_{\text{ds}}^{M_3}$. At the optimum, $r_{\text{ds}}^{M_3}$ is equal to $1/j\omega_{\text{LO}} C_{\text{ds}}^{M_3}$, and further increasing W_3 and

[1]This process is used for all simulations in this paper. The PSP compact MOSFET model [30] is used for all simulations.

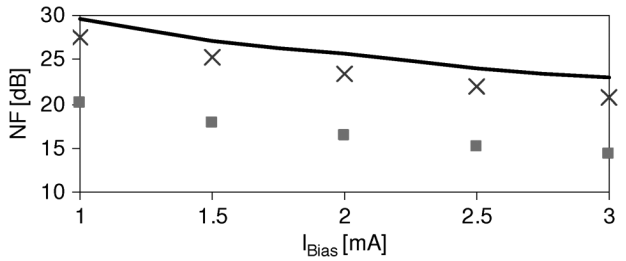

Fig. 4. Gilbert mixer's NF_{SSB} (IF@10 kHz) for ideal square-wave LO. (a) Simulated results (line). (b) Calculated with r_{ds} and C_{ds} (cross). (c) Calculated only with C_{ds} (square) as a function of the bias current.

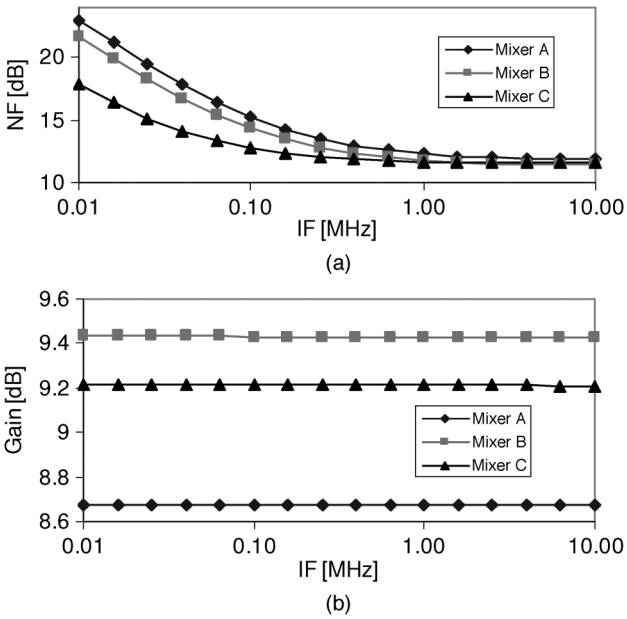

Fig. 5. (a) NF and (b) gain of three mixer designs with various channel length.

L_3 cannot reduce the flicker-noise leakage since the increasing $C_{ds}^{M_3}$ decreases $Z_{ds}^{M_3}$.

The length of the switching pair M_1/M_2 can also be increased to reduce the flicker-noise source at the cost of larger gate-source capacitance. Note that, for short transistors, the input capacitance is composed of the intrinsic gate-source capacitance and two relatively large overlap capacitances. Since the overlap capacitance is hardly affected by the transistor length, the LO power consumption will increase less than proportional to L.

For demonstration purposes, Fig. 5 shows the simulation results for three differently dimensioned mixers.

1) Mixer A: Using a minimum length for M_3 and M_1/M_2 ($W_3/L_3 = 60/0.1$, $W_1/L_1 = W_2/L_2 = 106/0.1$, and $V_{GT}^{M_3} = 0.13$).
2) Mixer B: The same as mixer A except that the width and length of M_3 are doubled with respect to the mixer A implementation ($W_3/L_3 = 120/0.2$, $W_1/L_1 = W_2/L_2 = 106/0.1$, and $V_{GT}^{M_3} = 0.124$).
3) Mixer C: The same as mixer B except that the length of transistors M_1 and M_2 is tripled with respect to the mixer B implementation ($W_3/L_3 = 120/0.2$, $W_1/L_1 = W_2/L_2 = 106/0.3$, and $V_{GT}^{M_3} = 0.125$).

For the three designs, the same LO driver ($f_{LO} = 2.01$ GHz, $V_{LO} = 1$ V, and $\alpha = 0.06$) is used, and the power consumption for all three mixers is set to 1.96 mW. This constant power consumption implies that the biasing conditions of M_3 are slightly different for all three implementations, which results in a small change in the transconductance of M_3. This minor change of M_3 transconductance results in a different gain of the mixer [see Fig. 5(b)]. Compared to the mixer using transistors with minimum length (mixer A), using only a longer channel length in M_3 (mixer B) can decrease NF@10 kHz by 1.3 dB since the larger $r_{ds}^{M_3}$ reduces the flicker-noise leakage. NF@10 MHz decreases by 0.4 dB due to a gain increase with 0.7 dB. Using longer M_3 and M_1/M_2 (mixer C) decreases NF@10 kHz by 5.2 dB since the flicker noise of the switching pair decreases for a longer transistor. NF@10 MHz decreases by 0.3 dB. Note that the LO driver power consumption increases by 25%, although we triple the length of the switch pair M_1/M_2.

Overall, it can be concluded that the minimum length is not optimum for transistors in the active mixer to reduce flicker noise. By properly increasing W and L of M_3 and by using longer M_1/M_2, the noise performance can be improved without any gain penalty but at the cost of a small increase in LO power. As a side effect, the transistor mismatches are also reduced for larger transistors, and flicker-noise cancellation by tuning out $C_{ds}^{M_3}$ [15], [16] becomes effective again.

IV. Time-Varying Weakly Nonlinear Analysis

For the analysis of the active mixer nonlinearity (IIP3 and IIP2), typically, a number of simplifications are made in literature [11], [13].

1) Only taking the transconductance nonlinearity into account. This is acceptable for older CMOS technologies but is certainly not acceptable for modern short-channel RF CMOS.
2) Neglecting the effect of finite LO slopes.
3) Calculating the switching pair and input stage nonlinearities separately.
4) Assuming constant, bias-independent, linear and nonlinear transconductances.

In this section, we explicitly use time-varying weakly nonlinear analyses, including the following.

1) The effect on IIP3 and IIP2 of all transistors' nonlinear conductances. This includes the resistive nonlinearities (transconductance, output conductance, and cross-modulation conductive terms describing the fact that the drain–source current is controlled by both v_{gs} and v_{ds}) and all capacitive nonlinearities (capacitance, transcapacitance, and cross-modulation capacitive terms).
2) The effect of the finite LO signal slope on IIP2 and IIP3.
3) Taking the switching pair and input stage in one circuit model so that that the mutual effects are included in the analyses.
4) The periodic MOS transistor nonlinearities.

A. IIP3 Estimation

For nonlinearity analysis, the mixer is considered as a weakly nonlinear circuit with respect to the input RF signal, where the dc bias is periodically changed by the LO signal. Then, with a two-tone input signal at ω_{RF_1} and ω_{RF_2}, the fundamental signal and intermodulation distortions at the output of the circuit shown in Fig. 1 are functions of the periodic LO, and thus,

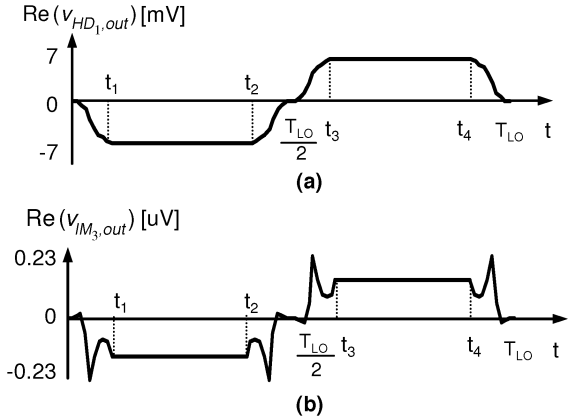

Fig. 6. Waveforms of (a) $v_{HD_1,\text{out}}[v_{\text{LO}}(t)]$ and (b) $v_{\text{IM}_3,\text{out}}[v_{\text{LO}}(t)]$.

they can be extended to a Fourier series as

$$v_{HD_1,\text{out}}\left[v_{\text{LO}}(t)\right] = \sum_{p=-\infty}^{+\infty} f_p^{HD_1} e^{jp\omega_{\text{LO}}t} \cdot e^{j\omega_{RF_1}t} \quad (17)$$

$$v_{\text{IM}_3,\text{out}}\left[v_{\text{LO}}(t)\right] = \sum_{p=-\infty}^{+\infty} f_p^{\text{IM}_3} e^{jp\omega_{\text{LO}}t} \cdot e^{j\left(2\omega_{RF_1}-\omega_{RF_2}\right)t} \quad (18)$$

where $f_p^{HD_1}$ and $f_p^{\text{IM}_3}$ are the Fourier coefficients. For low-side injection, the fundamental signal and IM3 are located in the IF band at $\omega_{RF_1}-\omega_{\text{LO}}$ and $2\omega_{RF_1}-\omega_{RF_2}-\omega_{\text{LO}}$, respectively. The magnitude of these signals is accounted for by the first-order Fourier series coefficients $f_{-1}^{HD_1}$ and $f_{-1}^{\text{IM}_3}$

$$V_{HD_1,IF} = f_{-1}^{HD_1} \quad (19)$$

$$V_{\text{IM}_3,\text{out}} = f_{-1}^{\text{IM}_3}. \quad (20)$$

As a result, the IIP3 can be obtained by using linear extrapolation for a small input power P_{in}

$$\begin{aligned} IIP3\ [\text{dBm}] &= \frac{1}{2}\times \text{dB}\left(\frac{V_{HD_1,IF}}{V_{\text{IM}_3,\text{out}}}\right) + P_{\text{in}}\ [\text{dBm}] \\ &= \frac{1}{2}\times \text{dB}\left(\frac{f_{-1}^{HD_1}}{f_{-1}^{\text{IM}_3}}\right) + P_{\text{in}}\ [\text{dBm}]. \end{aligned} \quad (21)$$

The exact waveforms of $v_{HD_1,\text{out}}[v_{\text{LO}}(t)]$ and $v_{\text{IM}_3,\text{out}}[v_{\text{LO}}(t)]$ can be obtained from the simulations. Fig. 6 shows the real part of the simulated waveforms for the mixer, with the gate bias of M_1 and M_2 changed according to the waveform of the LO signal shown in Fig. 3(a). The two-tone RF signals at the gate of M_3 are at 1.01 and 1.014 GHz, with a 1-mV amplitude. Note that the waveform of $v_{HD_1,\text{out}}[v_{\text{LO}}(t)]$ is the same as $F_n^{M_3}[v_{\text{LO}}(t)]$, which is approximated in Fig. 3(c). Therefore, (8)–(10) are reused for the approximation of $v_{HD_1,\text{out}}[v_{\text{LO}}(t)]$.

At (t_1,t_2), M_1 and M_3 act as a cascode amplifier, while at (t_3,t_4), M_2 and M_3 act as a cascode amplifier, and thus, $v_{HD_1,\text{out}}[v_{\text{LO}}(t)]$ reaches the negative and positive maxima, respectively [see Fig. 6(a)]. At $(0,t_1)$ and $(t_2,0.5T_{\text{LO}})$, M_2 is assumed to be off, and M_1 stays in the saturation region, while M_3 may toggle between the triode and saturation regions. Fig. 7 shows the ratio between the third-order transconductance nonlinearity g_{m3}, output conductance $g_{\text{ds}3}$, and cross-modulation nonlinearity ($g_{x21} = (1/2)\times(\partial^3 I_{\text{DS}}/\partial V_{\text{GS}}^2\partial V_{\text{DS}})$ and $g_{x12} = (1/2)\times(\partial^3 I_{\text{DS}}/\partial V_{\text{GS}}\partial V_{\text{DS}}^2)$). In the triode region, the cross-modulation nonlinearity and output conductance nonlinearity are dominant, while in the saturation region, the transconductance nonlinearity is dominant. Therefore, an IM3 peak occurs at $(0,t_1)$ and $(t_2,0.5T_{\text{LO}})$ when the transistor is well in the triode region [see Fig. 6(b)]. The same waveform can be seen in $(0.5T_{\text{LO}},t_3)$ and (t_4,T_{LO}) when M_1 is assumed to be off, and M_3 gradually changes from the triode to the saturation region.

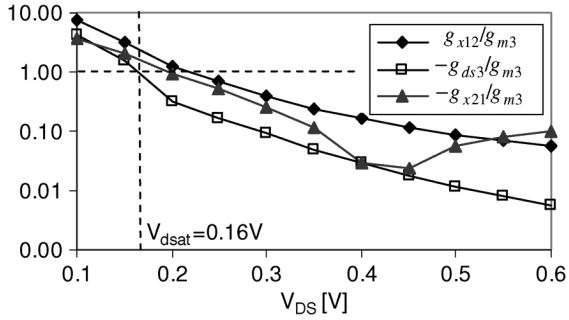

Fig. 7. $-g_{\text{ds}3}/g_{m3}$, $-g_{x21}/g_{m3}$, and g_{x12}/g_{m3} for an NMOS transistor in logarithm scale, $W/L = 60/0.1\ \mu$m, and $V_{\text{GT}} = 0.19$ V as a function of the drain–source voltage V_{DS}.

Equations (19) and (20) indicate that the first-order Fourier components $f_{-1}^{HD_1}$ and $f_{-1}^{\text{IM}_3}$ of $v_{HD_1,\text{out}}[v_{\text{LO}}(t)]$ and $v_{\text{IM}_3,\text{out}}[v_{\text{LO}}(t)]$ determine the IIP3. Note that sharp details of a signal are mainly caused by its high-order Fourier series coefficients [21]: for the estimation of only the first-order Fourier coefficient, a sufficiently accurate approximation of the waveforms in Fig. 6 is the trapezoidal waveform in Fig. 8.[2] Now, using (8) and (19)–(21), the voltage conversion gain is

$$V_{\text{gain}} = \frac{f_{-1}^{HD_1}}{V_{\text{IN}}} = \frac{2\sin(\alpha\pi)}{\pi^2\alpha}\frac{g_m^{M_1}g_m^{M_3}R_L Z_{\text{ds}}^{M_3}}{1+g_m^{M_1}Z_{\text{ds}}^{M_3}}\bigg|_{t_1} \quad (22)$$

and the IIP3 can be written as

$$\begin{aligned} IIP3\ [\text{dBm}] &= \frac{1}{2}\times \text{dB}\left(\frac{f_{-1}^{HD_1}}{f_{-1}^{\text{IM}_3}}\right) + P_{\text{in}}\ [\text{dBm}] \\ &\approx \frac{1}{2}\times \text{dB}\left(\frac{\frac{2\sin[\pi\cdot\alpha]\cdot v_{HD_1,\text{out}}|_{t_1}}{\pi^2\cdot\alpha}}{\frac{2\sin[\pi\cdot\alpha]\cdot v_{\text{IM}_3,\text{out}}|_{t_1}}{\pi^2\cdot\alpha}}\right) \\ &\quad + P_{\text{in}}\ [\text{dBm}] \\ &= \frac{1}{2}\times \text{dB}\left(\frac{v_{HD_1,\text{out}}}{v_{\text{IM}_3,\text{out}}}\right)\bigg|_{t_1} + P_{\text{in}}\ [\text{dBm}] \\ &= IIP3|_{t_1} \end{aligned} \quad (23)$$

where (see the Appendix for a first-order derivation)

$$\begin{aligned} v_{\text{IM}_3,\text{out}} \approx & \frac{-3\times V_{\text{IN}}^3 R_L}{4\times\left(1+g_m^{M_1}Z_{\text{ds}}^{M_3}\right)} \\ & \times\Big[Z_{\text{ds}}^{M_3}g_m^{M_1}\left(g_{m3}^{M_3}-g_{x21}^{M_3}\right)+g_{m3}^{M_1}-g_{\text{ds}3}^{M_1}\left(g_{m1}^{M_3}R_L\right)^3 \\ & -g_{x21}^{M_1}g_{m1}^{M_3}R_L+g_{x12}^{M_1}\cdot\left(g_{m1}^{M_3}R_L\right)^2\Big]. \end{aligned} \quad (24)$$

[2]The simulations show that the difference between the first-order Fourier series component of the original waveform and that of the approximated waveform typically is smaller than 2%.

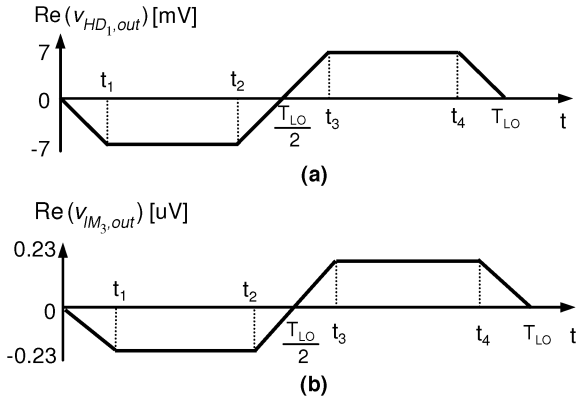

Fig. 8. Approximation of (a) $v_{HD_1,\mathrm{out}}[v_{\mathrm{LO}}(t)]$ and (b) $v_{\mathrm{IM}_3,\mathrm{out}}[v_{\mathrm{LO}}(t)]$.

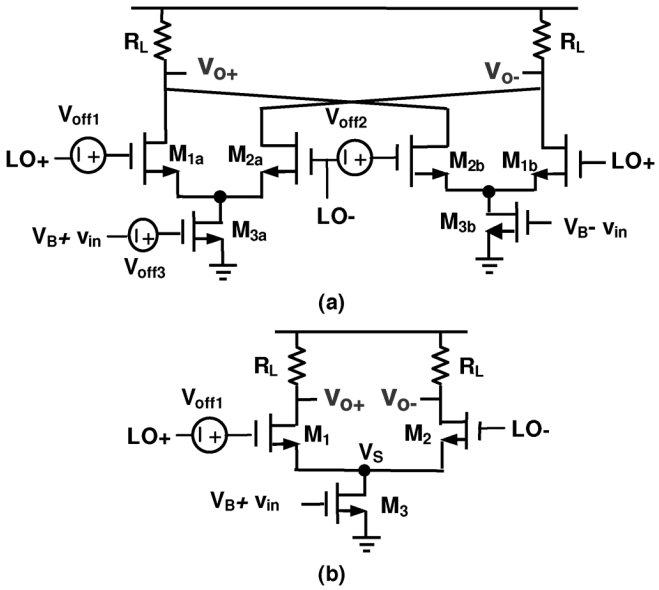

Fig. 9. (a) Double-balanced mixer with offset voltages for modeling the transistor mismatches. (b) Single-balanced mixer with offset voltages for modeling the transistor mismatches.

Equation (24) indicates that other third-order nonlinearities besides g_{m3} can significantly contribute to the output IM3. Fig. 7 shows that the nonlinearities g_{m3} and g_{x12} have the same sign (positive), while $g_{\mathrm{ds}3}$ and g_{x12} have the opposite sign (negative). Thus, all nonlinearity terms, with their weighting factor in (24), have a positive value, which shows that the contributions of each third-order nonlinearity to the IM3 add up. For low supply voltages, the LO signals with a large swing can easily drive M_1 out of the saturation region at t_1. Then, the output conductance nonlinearity $g_{\mathrm{ds}3}$ and the cross-modulation nonlinearities g_{x21} and g_{x12} of M_1 dramatically increase, which increases the $v_{\mathrm{IM}_3,\mathrm{out}}|_{t_1}$, thus decreasing IIP3, as indicated by (23) and (24).

In summary, the IIP3 of the time-varying mixer can be estimated by one time-invariant IIP3 calculation at the maximum of the LO signal. The effect of the slope of the LO signal on IIP3 can be neglected. In the low supply voltage processes, for high IIP3, an LO signal with a large swing is not desirable because the switching transistor enters into the triode region when the LO reaches its maximum.

B. IIP2 Estimation

Mismatches in transistors and load resistors, self-mixing, and transistor nonlinearity together cause finite IIP2 for the balanced mixer [12]. The effect of self-mixing and mismatches in load resistors can be made negligible using layout counter measures [22]. Then, the remaining dominant factors for IIP2 are transistor mismatches and nonlinearities. For the double-balanced mixer, transistor mismatch can be modeled by the three dc offset voltages shown in Fig. 9(a) ($V_{\mathrm{off},1}$ for the mismatch of $\mathrm{M}_{1a}/\mathrm{M}_{2a}$, $V_{\mathrm{off},2}$ for the mismatch of $\mathrm{M}_{1b}/\mathrm{M}_{2b}$, and $V_{\mathrm{off},3}$ for the mismatch of $\mathrm{M}_{3a}/\mathrm{M}_{3b}$). Since the effect of the switch pair mismatch is typically much larger than that of the transconductors [12], we will neglect $V_{\mathrm{off},3}$ and will use the single-balanced mixer shown in Fig. 9(b) in this analysis.

As explored in the previous section, the mixer is considered as a nonlinear circuit with respect to the input RF signal, where the dc bias is periodically changed by the LO signal. With dc offset, the LO is not symmetric, and therefore, the single-ended IM2s at the positive and negative outputs are not equal. Hence, they will not cancel. As a function of the asymmetric LO signal, for a two-tone input signal at ω_{RF_1} and ω_{RF_2}, the IM2 at the differential output can be extended to a Fourier series as

$$\begin{aligned} v_{\mathrm{IM}_2,\mathrm{out}}\left[v_{\mathrm{LO}}(t)\right] &= v_{\mathrm{IM}_2+}\left[v_{\mathrm{LO}}(t)\right] - v_{\mathrm{IM}_2-}\left[v_{\mathrm{LO}}(t)\right] \\ &= \left(\sum_{p=-\infty}^{+\infty} f_{p,v+}^{\mathrm{IM}_2} e^{jp\omega_{\mathrm{LO}}t} \right. \\ &\quad \left. - \sum_{p=-\infty}^{+\infty} f_{p,v-}^{\mathrm{IM}_2} e^{jp\omega_{\mathrm{LO}}t} \right) \\ &\quad \cdot e^{j(\omega_{RF_1}-\omega_{RF_2})t}. \end{aligned} \tag{25}$$

For low-side injection, the IM2 distortion is located at $\omega_{RF_1} - \omega_{RF_2}$ in the IF band, which is accounted by the zeroth-order Fourier series components $f_{0,v+}^{\mathrm{IM}_2}$ and $f_{0,v-}^{\mathrm{IM}_2}$

$$v_{\mathrm{IM}_2,\mathrm{out}} = f_{0,v+}^{\mathrm{IM}_2} - f_{0,v-}^{\mathrm{IM}_2}. \tag{26}$$

Then, the IIP2 is

$$\begin{aligned} \mathrm{IIP2}\,[\mathrm{dBm}] &= \mathrm{dB}\left(\frac{v_{HD_1,IF}}{v_{\mathrm{IM}_2,\mathrm{out}}}\right) + P_{\mathrm{in}}\,[\mathrm{dBm}] \\ &= \mathrm{dB}\left(\frac{f_{-1}^{HD_1}}{f_{0,v+}^{\mathrm{IM}_2} - f_{0,v-}^{\mathrm{IM}_2}}\right) + P_{\mathrm{in}}\,[\mathrm{dBm}]. \end{aligned} \tag{27}$$

Using the same approach as that for IM3, the real parts of the simulated waveforms of the single-ended IM2 ($v_{\mathrm{IM}_2,+}[v_{\mathrm{LO}}(t)]$ and $v_{\mathrm{IM}_2,-}[v_{\mathrm{LO}}(t)]$) are shown in Fig. 10. Note that the single-ended IM2s at the positive and negative outputs ($f_{0,v+}^{\mathrm{IM}_2}$ and $f_{0,v-}^{\mathrm{IM}_2}$) correspond to the dc term of $v_{\mathrm{IM}_2+}[v_{\mathrm{LO}}(t)]$ and $v_{\mathrm{IM}_2-}[v_{\mathrm{LO}}(t)]$. These dc terms, in turn, correspond to the integral of (or area below) the waveform. In a perfectly symmetric mixer, the single-ended IM2s at the positive and negative outputs ($f_{0,v+}^{\mathrm{IM}_2}$ and $f_{0,v-}^{\mathrm{IM}_2}$) are equal, and therefore, they exactly cancel each other, leading to an infinite IIP2. However, any dc offset introduces an effectively asymmetric LO and an asymmetric bias modulation in $(0, 0.5T_{\mathrm{LO}})$ and $(0.5T_{\mathrm{LO}}, T_{\mathrm{LO}})$, hence resulting in waveform differences between $v_{\mathrm{IM}_2+}[v_{\mathrm{LO}}(t)]$ and $v_{\mathrm{IM}_2-}[v_{\mathrm{LO}}(t)]$. The single-ended

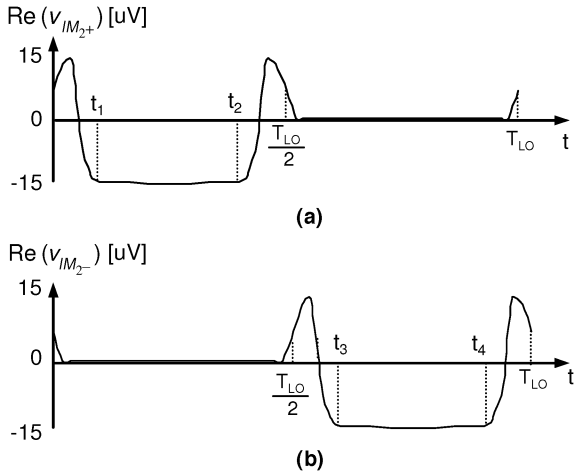

Fig. 10. Waveforms of (a) $v_{\mathrm{IM}_2+}[v_{\mathrm{LO}}(t)]$ and (b) $v_{\mathrm{IM}_2-}[v_{\mathrm{LO}}(t)]$.

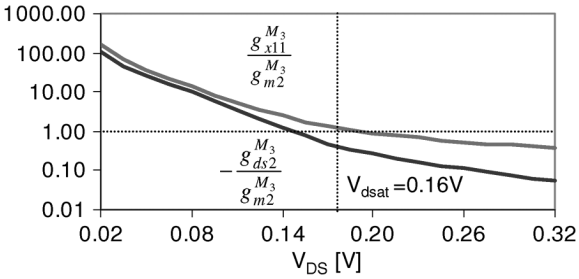

Fig. 11. $-g_{\mathrm{ds2}}^{M_3}/g_{m2}^{M_3}$ and $g_{x11}^{M_3}/g_{m2}^{M_3}$ in logarithmic scale for the NMOS transistor in the triode and saturation regions.

IM2s at the positive and negative outputs do not exactly cancel, which results in a finite IIP2.

Due to a high similarity between $v_{\mathrm{IM}_2+}[v_{\mathrm{LO}}(t)]$ and $v_{\mathrm{IM}_2-}[v_{\mathrm{LO}}(t)]$, we choose to show the details on the estimation of the single-ended IM2 at the positive output $f_{0,v+}^{\mathrm{IM}_2}$. As discussed in Section II, at $(0, 0.5T_{\mathrm{LO}})$, M_1 and M_3 act as a cascode amplifier. Due to the low supply voltage in deep-submicrometer technologies, M_1 stays in the saturation region, and M_3 may toggle between the triode and saturation regions. As derived in the Appendix, transistor M_3 dominantly contributes to the IM2 of the cascode amplifier

$$v_{\mathrm{IM2}+} \approx \frac{r_{\mathrm{ds}}^{M_3} g_{m1}^{M_3} R_L}{1 + r_{\mathrm{ds}}^{M_3} g_{m1}^{M_3}} \cdot V_{\mathrm{IN}}^2 \cdot \left[-g_{m2}^{M_3} - g_{\mathrm{ds2}}^{M_3} \cdot \left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}} \right)^2 + g_{x11}^{M_3} \cdot \frac{g_{m1}^{M_3}}{g_{m1}^{M_1}} \right] \quad (28)$$

where $r_{\mathrm{ds}}^{M_3}$ and $g_{m1}^{M_3}$ are the linear output resistance and transconductance of M_3, $g_{m2}^{M_3} = (1/2) \times (\partial^2 I_{\mathrm{DS}}/\partial V_{\mathrm{GS}}^2)$ is the derivative of the transconductance, $g_{\mathrm{ds2}}^{M_3} = (1/2) \times (\partial^2 I_{\mathrm{DS}}/\partial V_{\mathrm{DS}}^2)$ is the derivative of the output conductance, and $g_{x11}^{M_3} = \partial^2 I_{\mathrm{DS}}/\partial V_{\mathrm{GS}} \partial V_{\mathrm{DS}}$ is the second-order cross-modulation nonlinearity. Equation (28) indicates that the sign of $v_{\mathrm{IM2}+}$ for a different bias is determined by $g_{m2}^{M_3}$, $g_{\mathrm{ds2}}^{M_3}$, and $g_{x11}^{M_3}$. As an example, an NMOS transistor ($W/L = 60/0.1$ μm), with V_{GS} fixed, is simulated by sweeping V_{DS} from 0.02 to 0.32 V. Fig. 11 shows that, in the saturation region, the transconductance nonlinearity ($g_{m2}^{M_3}$) is dominant, while in the triode region, the cross-modulation nonlinearity $g_{x11}^{M_3}$ and the output conductance nonlinearity $g_{\mathrm{ds2}}^{M_3}$ become dominant. As LO rises and falls, M_3 may enter in the triode region where the cross-modulation nonlinearity and output conductance nonlinearity ($g_{x11}^{M_3}$ and $g_{\mathrm{ds2}}^{M_3}$) are dominant. Then, the term $g_{x11}^{M_3} \cdot (g_{m1}^{M_3}/g_{m1}^{M_1}) - g_{\mathrm{ds2}}^{M_3} \cdot (g_{m1}^{M_3}/g_{m1}^{M_1})^2$ in (28) is dominant, and $v_{\mathrm{IM2}+}$ is positive, as shown in Fig. 10(a). As the LO increases, M_3 enters into the saturation region, where $g_{m2}^{M_3}$ becomes dominant. Since $g_{m1}^{M_1}$ and $g_{m1}^{M_3}$ are on the same order of magnitude, the term $-g_{m2}^{M_3}$ in (28) is dominant, and $v_{\mathrm{IM2}+}$ turns negative.

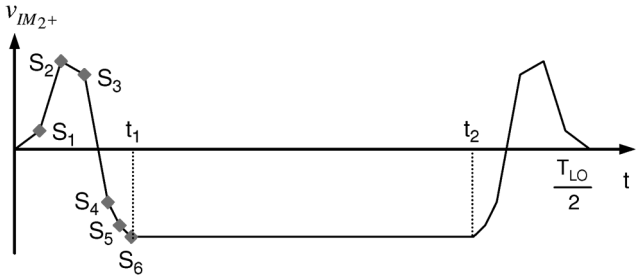

Fig. 12. Estimation of $v_{\mathrm{IM}_2+}[v_{\mathrm{LO}}(t)]$.

Note that the change of $v_{\mathrm{IM2}+}$ between positive and negative values during the LO rise/fall time have not been considered in [12] due to the following simplifications.

1) Only the transconductance nonlinearity of the transistor is considered. As a result, the fact that the nonlinearity of transistor M_3 is also modulated by its drain–source voltage is neglected.
2) The effect of the finite LO slope on IIP2 is neglected. As a result, the fact that, due to a low supply voltage in the deep-submicrometer technologies, transistor M_3 typically toggles between the triode and saturation regions during the LO rise and fall times is neglected.

Since the positive single-ended IM2 $f_{0,v+}^{\mathrm{IM}_2}$ is equal to the integral of the waveform of $v_{\mathrm{IM2}+}$ shown in Fig. 10(a), neglecting the positive area in $(0, t_1)$ and $(t_2, 0.5T_{\mathrm{LO}})$ can overestimate the positive single-ended IM2 $f_{0,v+}^{\mathrm{IM}_2}$. The same conclusion applies to the negative single-ended IM2 $f_{0,v-}^{\mathrm{IM}_2}$. In summary, the single-ended IM2 can be overestimated by neglecting the LO slope, cross-modulation nonlinearity, and output conductance nonlinearity of the transistor. As a result, the differential IM2 $(f_{0,v+}^{\mathrm{IM}_2} - f_{0,v-}^{\mathrm{IM}_2})$ can be significantly misestimated.

In order to give an accurate estimation of the positive single-ended IM2 $f_{0,v+}^{\mathrm{IM}_2}$, a good capture of the waveform, especially in $(0, t_1)$ and $(t_2, 0.5T_{\mathrm{LO}})$, is essential. Fig. 12 shows the estimation of $f_{0,v+}^{\mathrm{IM}_2}$ by six equal-distant samples (S_1 to S_6). Assuming that the rise/fall time of the LO that are equal to $\alpha \cdot T_{\mathrm{LO}}$ yield $t_1 = 0.5\alpha \cdot T_{\mathrm{LO}}$ and $t_2 = 0.5 \times (1 - \alpha) \cdot T_{\mathrm{LO}}$, the area of $v_{\mathrm{IM}_2+}[v_{\mathrm{LO}}(t)]$ is then given by

$$\begin{aligned} f_{0,v+}^{\mathrm{IM}_2} &= \frac{t_1}{3T_{\mathrm{LO}}} \cdot \left[S_1 + S_2 + S_3 + S_4 + S_5 + \frac{S_6}{2} \right] + \frac{t_2 - t_1}{T_{\mathrm{LO}}} \cdot S_6 \\ &= \frac{\alpha}{6} \cdot \Big[v_{\mathrm{IM}_2+}|_{v_{LO+}^1} + v_{\mathrm{IM}_2+}|_{v_{LO+}^2} + v_{\mathrm{IM}_2+}|_{v_{LO+}^3} \\ &\qquad + v_{\mathrm{IM}_2+}|_{v_{LO+}^4} + v_{\mathrm{IM}_2+}|_{v_{LO+}^5} + 0.5 v_{\mathrm{IM}_2+}|_{v_{LO+}^6} \Big] \\ &\quad + (0.5 - \alpha) \cdot v_{\mathrm{IM}_2+}|_{v_{LO+}^6} \end{aligned} \quad (29)$$

where $v_{LO+}^k = V_c + (k/6)V_{\mathrm{LO}} + V_{\mathrm{off}}$.

Similarly, the area of $v_{\mathrm{IM}_2-}[v_{\mathrm{LO}}(t)]$ can be estimated as

$$f_{0,v-}^{\mathrm{IM}_2} = \frac{\alpha}{6}\cdot\Big[v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^1} + v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^2} + v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^3} + v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^4} + v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^5} + 0.5v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^6}\Big] + (0.5-\alpha)\cdot v_{\mathrm{IM}_2-}|_{v_{\mathrm{LO}-}^6} \quad (30)$$

where $v_{LO-}^k = V_c + (k/6)V_{\mathrm{LO}}$.

Now, by using (27), (29), and (30), the IIP2 of the time-varying mixer can be estimated by a few time-invariant IM2 calculations. Note that the estimation of the single-ended IM2 by samples at different instants includes the periodic property of the transistor nonlinearity. For each instant sample, the IM2 is calculated by (31), as given in the Appendix

$$v_{\mathrm{IM}_2,+} = \frac{R_L V_{\mathrm{IN}}^2}{1+g_m^{M_1} r_{\mathrm{ds}}^{M_3}} \cdot \Bigg\{ r_{\mathrm{ds}}^{M_3} g_m^{M_1} \cdot \Bigg[-g_{m2}^{M_3} - g_{\mathrm{ds}2}^{M_3} \cdot \left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)^2 + g_{x11}^{M_3} \cdot \left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)\Bigg] + \Bigg[-g_{m2}^{M_1} - g_{\mathrm{ds}2}^{M_1} \cdot \left(g_{m1}^{M_3} R_L\right)^2 + g_{x11}^{M_1} \cdot g_{m1}^{M_3} R_L \Bigg]\Bigg\}. \quad (31)$$

First, it shows that, for the narrow-band IM2, the nonlinear capacitance can be neglected. It also shows that the terms with the cross-modulation nonlinearity and output conductance nonlinearity (g_{x11} and $g_{\mathrm{ds}2}$) can cancel the terms with transconductance nonlinearity g_{m2}. Due to the low supply voltage in deep-submicrometer technologies, LO signals with a large swing can easily drive M_1 out of the saturation region at t_1. Then, at (t_1, t_2), M_1 may stay in between the triode and saturation regions, where g_{x11} and $g_{\mathrm{ds}2}$ becomes larger, while M_3 stays in the saturation region. In that case, (31) can be simplified to

$$v_{\mathrm{IM}_2,+} = \frac{R_L V_{\mathrm{IN}}^2}{1+g_m^{M_1} r_{\mathrm{ds}}^{M_3}} \cdot \Bigg[-g_{m2}^{M_3} r_{\mathrm{ds}}^{M_3} g_m^{M_1} - g_{m2}^{M_1} - g_{\mathrm{ds}2}^{M_1} \cdot \left(g_{m1}^{M_3} R_L\right)^2 + g_{x11}^{M_1} \cdot g_{m1}^{M_3} R_L \Bigg] \quad (32)$$

as given in the Appendix. With the scaling factor $(g_{m1}^{M_3} R_L)^2$ and $g_{m1}^{M_3} R_L$ for $g_{\mathrm{ds}2}^{M_1}$ and $g_{x11}^{M_1}$, the single-ended IM2 can be very small. However, due to the high sensitivity of g_{x11} and $g_{\mathrm{ds}2}$ to dc offset voltages (mismatches), the low single-ended IM2 does not guarantee a high differential IIP2. Fig. 13 shows that the single-balanced mixer shown in Fig. 9(b) is simulated for varying gate bias of M_3 with fixed dimensions. The single-ended IIP2 was derived for a situation without mismatch, while the differential IIP2 is the minimum IIP2 from a Monte Carlo mismatch simulation. As the gate bias of M_3 increases, more current flows through R_L. Thus, in (t_1, t_2), M_1 may enter the triode region, where the distortion from g_{x11} and $g_{\mathrm{ds}2}$ of M_1 increases and cancels a larger part of the distortion from g_{m2} of

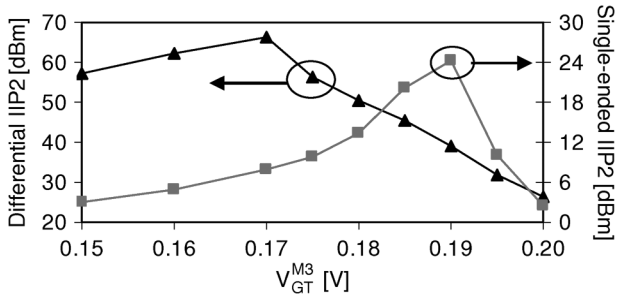

Fig. 13. Single-ended IIP2 and minimum differential IIP2 of the single-balanced mixer for various V_{GT} of the transistor M_3.

M_3. This results in a smaller negative area of $v_{\mathrm{IM}_2,+}$, shown in Fig. 12, and yields a high single-ended IIP2. However, the high differential IIP2 is achieved for a smaller $V_{GT}^{M_3}$. At such bias, M_1 and M_3 are all in the saturation region in (t_1, t_2), where g_{m2} of M_3 is dominant for the single-ended IM2 and less sensitive to the dc offset voltages.

C. Impact of the LO Signal on Mixer Nonlinearity

The LO signal of the mixer practically has finite rise/fall time. However, the influence of the LO slope on the mixer nonlinearity has not yet been investigated in previous literature [11]–[14]. The analysis in Sections IV-A and IV-B shows that, for a low supply voltage, as the LO rises or falls, transistor M_3 may experience a deep triode region operation where the cross-modulation nonlinearity and output conductance nonlinearity become dominant. As discussed in Section IV-A, the IM3 output at the IF band is equal to the first-order Fourier coefficient $f_{-1}^{\mathrm{IM}_3}$ of $v_{\mathrm{IM}_3,\mathrm{out}}[v_{\mathrm{LO}}(t)]$, which is under little influence of the LO slope. Therefore, we conclude that the LO slope effect on the mixer IIP3 can be neglected. The IIP3 of the time-varying system can be estimated by one time-invariant nonlinearity calculation.

As for the IIP2, during the rise/fall time, M_3 toggles between the triode and saturation regions. In the triode region, the cross-modulation nonlinearity and output conductance nonlinearity of the transistor are dominant, and they will result in a positive single-ended IM2. In the saturation region, the transconductance nonlinearity is dominant, and the single-ended IM2 changes to a negative value. Since the overall single-ended IM2 is the sum of the positive and negative contributions in one period, neglecting the LO slope can overestimate the single-ended IM2 and may misestimate the differential IM2. Fig. 14 shows an illustration: the mixer shown in Fig. 9(b) is simulated by sweeping the width and V_{GT} of M_3 with fixed $P_{\mathrm{dc}} = 2$ mW at 2 GHz. A fixed 5-mV dc offset is used to model the mismatch of the switch pair. The simulated IIP2s for a square-wave LO and a LO with finite slopes ($t_{\mathrm{rise}} = t_{\mathrm{fall}} = 0.08T_{\mathrm{LO}}$) are compared. The difference of the IIP2 between using a square-wave LO and using LO with a finite slope can be as large as 30 dB, which demonstrates the importance of including the LO slope in the IIP2 estimation.

D. LO Slope Tuning for IIP2 Calibration

Including the LO slope in the analysis not only provides a more accurate estimation on the IIP2 but also shows one new possibility of introducing intentional mismatch that can be used for IIP2 calibration. In order to achieve a high IIP2, typically,

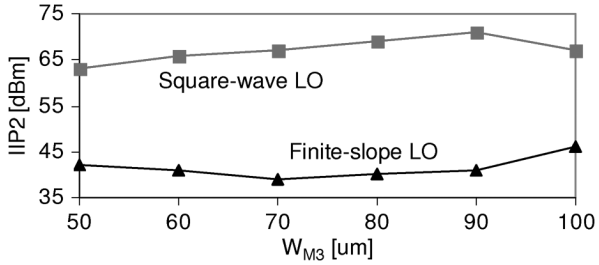

Fig. 14. Simulated IIP2 for (a) a square-wave LO (square symbol) and (b) an LO with a finite slope ($t_{\rm rise} = t_{\rm fall} = 0.08T_{\rm LO}$) (triangle symbol).

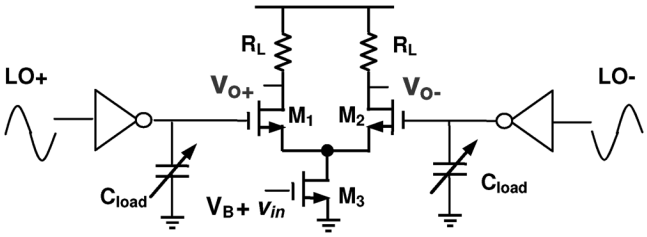

Fig. 15. Schematic of the mixer using LO slope tuning for IP2 calibration.

mismatches are introduced to the mixer to neutralize the differential IM2 output caused by the intrinsic mismatches.

Currently, the possibilities for introducing intentional mismatches are the following:

1) controlling the mismatch between the loads by resistor trimming [23], [24] or by tuning the pMOS load with NF degradation by 1–2 dB due to the noise introduced by the extra pMOS current sources [25];
2) tuning the current sources within the common-mode feedback (CMFB) section for the current mode output mixers [26];
3) tuning the dc level of the LO signal [27].

The discussion in Section IV-B suggests that the IIP2 can also be calibrated by tuning the LO slope. For demonstration purposes, the mixer shown in Fig. 15 is simulated for $f_{\rm LO}$ at 3.01 GHz and two-tone signals at 3.02 and 3.024 GHz with a -25-dBm input power. The mixer is driven by two inverters with load capacitors. By changing these load capacitors, we can tune the LO slope for the mixer. Note that a very small LO slope change is sufficient due to the high sensitivity of the differential IM2 to the LO slope. In this case, the slope change is between -0.75% and $+0.75\%$, shown in Fig. 16(c), while the inverter power dissipation is changing between 2.2 and 2.7 mW. Fig. 16 shows that a high IIP2 can be achieved by tuning the LO slope, while the gain, IIP3, and NF are not affected.

E. Summary

It can be concluded that, by using the time-varying weakly nonlinear analysis, the IIP3 and IIP2 of the mixer can be estimated by a few time-invariant weakly nonlinearity calculation, where the effect of the LO slope is included. Note that, in the time-invariant nonlinearity calculations, the contribution of the switching pair $(\rm M_1/M_2)$ and the input stage $\rm M_3$ is evaluated as a whole circuit but not separately as in [12]. As a result, the nonlinearity of the time-varying circuit can be estimated by time-invariant nonlinearity calculations, which is straightforward by using the Volterra series approach [28] or the general weak nonlinearity model for amplifiers [29].

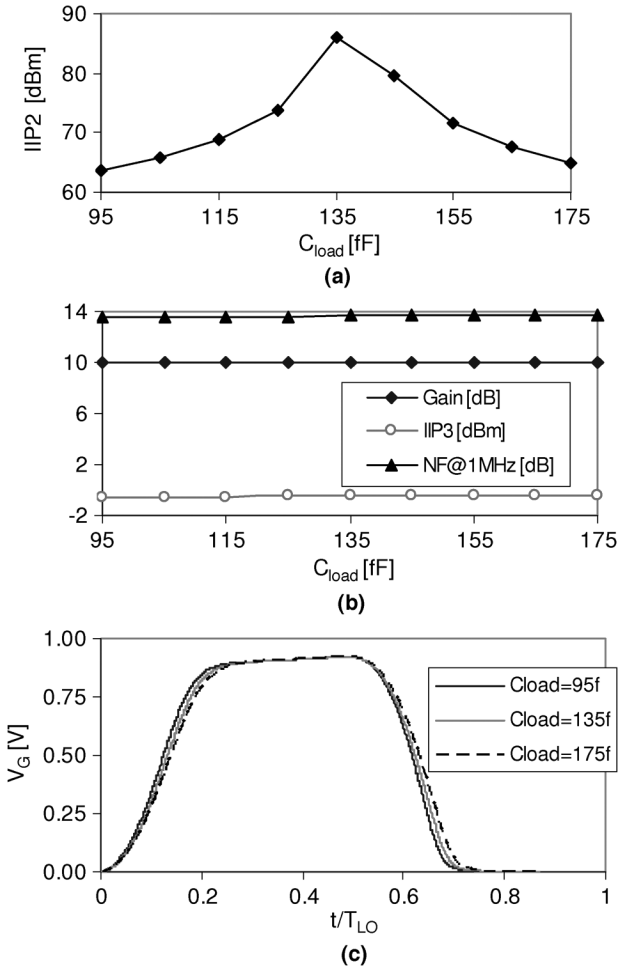

Fig. 16. (a) IIP2, (b) gain, IIP3, and SSB NF at 1 MHz versus the inverter load capacitor tuning and (c) waveform of LO- at the gate of $\rm M_2$ for different tuning capacitors.

V. Benchmarking Accuracy

To evaluate the accuracy of the model for noise and IIP3 calculation, the single-balanced mixer in Fig. 1 is simulated. For the model of the IIP2, the double-balanced mixer shown in Fig. 9(a) is simulated. The simulation results in Spectre and the calculation results using our model are presented in this section. We implemented the noise and nonlinearity model within a mixer P-cell, which is similar to what we did for a low-noise amplifier [6], where all small-signal parameters and nonlinearities of the transistor are included. This mixer P-cell dimensions the given circuit topology for a given set of specifications. The time-invariant nonlinearity calculation for IIP2 and IIP3 estimation is performed by using the circuit nonlinearity model [29], where all of the resistive and capacitive nonlinearities are included.

For Fig. 17, the Gilbert mixer was dimensioned at 3-mW power consumption, with $f_{\rm LO}$ at 2 GHz and IF at 10 kHz. At this low IF, the flicker noise is dominant. For the LO signal, $V_C = 0.6$ V, $V_{\rm LO} = 0.5$ V, and $\alpha = 0.1$. Fig. 17 shows the SSB NF and the conversion gain as a function of the gate-overdrive voltage of $\rm M_3$. It is shown in Fig. 17(a) that the noise model with output resistance and capacitance (square symbols) has an estimation error that is smaller than 0.9 dB, while the error is 3 dB for the noise model with output capacitance but without output

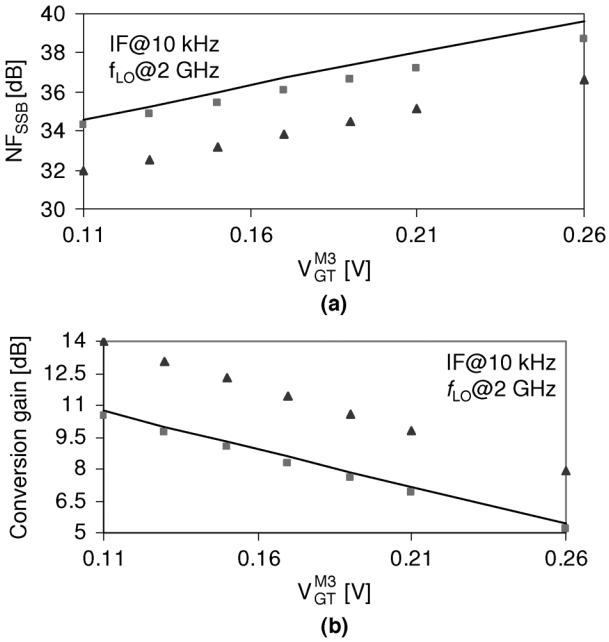

Fig. 17. Gilbert mixer's (a) $\mathrm{NF_{SSB}}$ and (b) conversion gain. Simulated results (line) and modeled results with r_{ds} (squares) and without r_{ds} (triangles) as a function of the overdrive voltage of $\mathrm{M_3}$.

resistance (triangular symbol). Fig. 17(b) shows that the conversion gain resulting from our model with an output resistance has an estimation error that is smaller than 0.3 dB, while the error is more than 2.5 dB if the output resistance of $\mathrm{M_3}$ is neglected. The analyses in Section III-A suggest that, with the scaling of the CMOS technology (having a lower supply voltage, a higher f_T, and a lower output resistance), the flicker-noise leakage caused by the finite output resistance of $\mathrm{M_3}$ becomes significant, and it cannot be neglected. Fig. 18 shows the simulated and calculated SSB NF as a function of the IF frequency (for f_{LO} at 2 GHz, $P_{\mathrm{dc}} = 3$ mW, $V_{\mathrm{LO}} = 0.5$ V, and $\alpha = 0.1$ for $V_C = 0.4$ V, $V_C = 0.5$ V, and $V_C = 0.6$ V). The estimation error of our noise model is smaller than 0.3 dB. As the mixer acts as a balanced differential pair at $0.5T_{\mathrm{LO}}$, a lower common-mode level of the LO signal V_C causes a lower gain for the differential pair, and thus, a smaller noise spikes, as shown in Fig. 2(b). As a result, the flicker-noise leakage is smaller for a lower V_C. Fig. 19 shows a comparison of the simulated and calculated $\mathrm{NF_{SSB}}$ both at IF $= 10$ kHz and IF $= 10$ MHz as a function of the LO frequency (for an 11-dB conversion gain, $P_{\mathrm{dc}} = 3$ mW, $V_{\mathrm{LO}} = 1$ V, and $\alpha = 0.1$). The figure illustrates that the estimation error of our noise model is lower than 1 dB for an f_{LO} below 5 GHz in a 90-nm CMOS technology.

Fig. 20 shows the simulated and calculated IIP3 as a function of the overdrive voltage of $\mathrm{M_3}$, with 3-mW power consumption. For the LO signal, $f_{\mathrm{LO}} = 2$ GHz, $V_C = 0.3$ V, $V_{\mathrm{LO}} = 0.6$ V, and $\alpha = 0.1$. The two-tone signals are at 2.01 and 2.014 GHz, and the IIP3 is extrapolated by sweeping the input power from -25 to -15 dBm. The estimation error is within 0.5 dB for our model, while the error is larger than 6 dB for the model that only includes the g_m nonlinearity. Fig. 21 shows a comparison of the simulated and calculated IIP3s as a function of the LO frequency from 1 to 5 GHz. For the LO signal, $V_C = 0.3$ V, $V_{\mathrm{LO}} = 0.6$ V, and $\alpha = 0.1$. The two-tone signals are at $f_{\mathrm{LO}} + 10$ MHz and $f_{\mathrm{LO}} + 14$ MHz, and the IIP3 is extrapolated by sweeping the

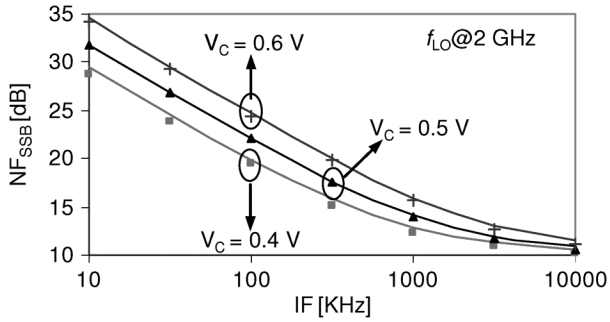

Fig. 18. Gilbert mixer's $\mathrm{NF_{SSB}}$ as a function of the IF frequency for three values of V_C. Simulated $\mathrm{NF_{SSB}}$ (line) and modeled $\mathrm{NF_{SSB}}$ (symbol) for $V_C = 0.4$ V, $V_C = 0.5$ V, and $V_C = 0.6$ V.

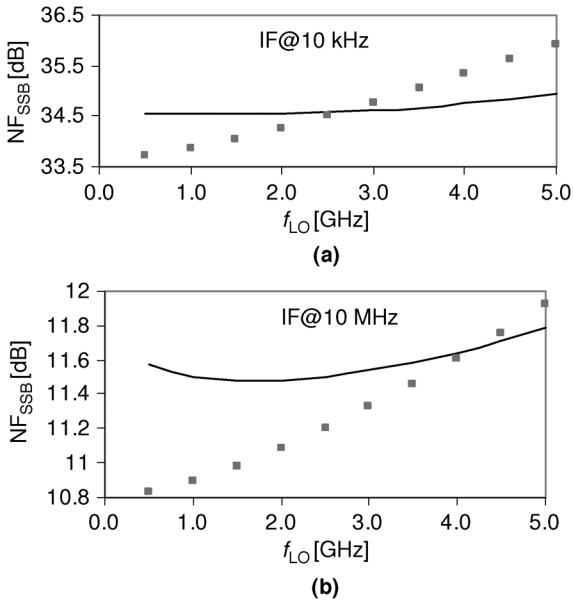

Fig. 19. Gilbert mixer's (a) $\mathrm{NF_{SSB}}$ for IF $= 10$ kHz and (b) $\mathrm{NF_{SSB}}$ for IF $= 10$ MHz as a function of the LO frequency. Simulated $\mathrm{NF_{SSB}}$ (line) and modeled $\mathrm{NF_{SSB}}$ (square symbols) for $f_{\mathrm{LO}} = 2$ GHz, $P_{\mathrm{dc}} = 3$ mW, $V_C = 0.6$ V, $V_{\mathrm{LO}} = 0.5$ V, and $\alpha = 0.1$.

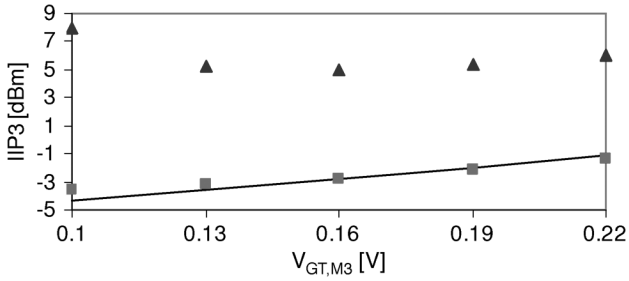

Fig. 20. Gilbert mixer's IIP3 as a function of the overdrive voltage of $\mathrm{M_3}$. Simulated IIP3 (line), modeled IIP3 that includes all nonlinearity (squares), and modeled IIP3 that includes only the g_m nonlinearity (triangular) for $P_{\mathrm{dc}} = 3$ mW, $V_C = 0.3$ V, $V_{\mathrm{LO}} = 0.6$ V, and $\alpha = 0.1$.

input power from -25 to -15 dBm. The estimation error of our IIP3 model is lower than 1 dB, while the error is larger than 5 dB for the model that only includes the g_m nonlinearity. The estimation error in our model increases with frequency because the effect of capacitances on the exact waveform of the drain of $\mathrm{M_3}$ is neglected.

For IIP2, the double-balanced mixer shown in Fig. 9(a) is used. Three dc offset voltage sources model the switch pair mismatch and input stage mismatch, where V_{off1} is 5 mV, V_{off2} is 3

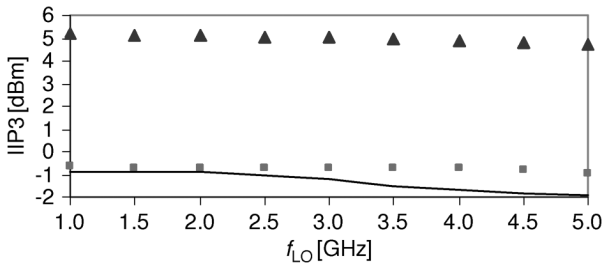

Fig. 21. Gilbert mixer's IIP3 as a function of the LO frequency. Simulated IIP3 (line), modeled IIP3 that includes all conductance nonlinearity (squares), and modeled IIP3 that includes only the g_m nonlinearity (triangular) for $P_{\rm dc} = 3$ mW, $V_C = 0.3$ V, $V_{\rm LO} = 0.6$ V, and $\alpha = 0.1$.

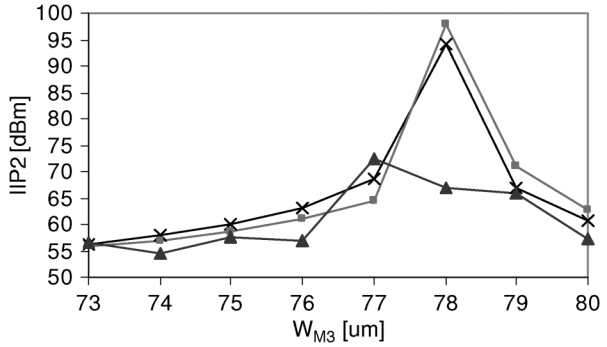

Fig. 22. Gilbert mixer's IIP2 as a function of the width of $\rm M_3$. Simulated IIP2 (line with cross), modeled IIP2 that includes all conductance nonlinearity and LO effect (line with squares), and modeled IIP2 that includes only the g_m nonlinearity and no LO slope effect (line with triangles) for $f_{\rm LO} = 1$ GHz, $V_C = 0.3$ V, $V_{\rm LO} = 0.6$ V, and $\alpha = 0.1$.

mV, and $V_{\rm off3}$ is 3 mV. Fig. 22 shows the simulated and calculated IIP2s as a function of the width of $\rm M_3$ for a constant $V_{\rm GT}$. For the LO signal, $f_{\rm LO} = 1$ GHz, $V_C = 0.3$ V, $V_{\rm LO} = 0.6$ V, and $\alpha = 0.1$. The IIP2 is estimated by 12 time-invariant nonlinearity calculations. The two-tone signals are at 1.01 and 1.014 GHz, and the IIP2 is extrapolated by sweeping the input power from -30 to -25 dBm. For an IIP2 lower than 65 dBm, the estimation error of our model is below 1 dB, while for an IIP2 higher than 70 dBm, the error is within 4 dB. For the model that only includes the g_m nonlinearity and without considering the LO slope effect, it does not predict the IIP2 peak. Fig. 23 shows a comparison of the simulated and calculated IIP2s as a function of the LO frequency from 1 to 5 GHz. For the LO signal, $V_C = 0.3$ V, $V_{\rm LO} = 0.6$ V, and $\alpha = 0.1$. The two-tone signals are at $f_{\rm LO} + 10$ MHz and $f_{\rm LO} + 14$ MHz, and the IIP2 is extrapolated by sweeping the input power from -30 to -25 dBm. The estimation error of our model increases with an increasing LO frequency, but it remains smaller than 4 dB, while the error is larger than 10 dB for the model that only includes the g_m nonlinearity and without considering the LO slope effect.

The estimation time of our model is compared with the simulation time using Spectre, shown in Table I. For noise and IIP3 estimation, our model speeds up the estimation time by a factor of about 40 since only one or two time-invariant circuit calculations are involved. Although IIP2 estimation takes 12 time-invariant nonlinearity calculations, the estimation time is still three times less in comparison with the circuit simulation.

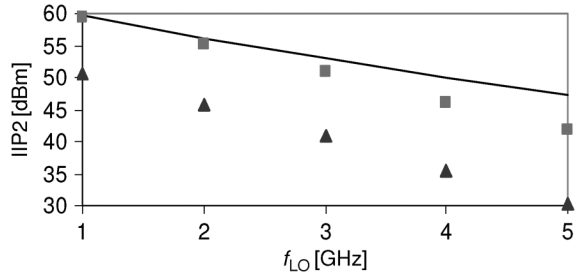

Fig. 23. Gilbert mixer's IIP2 as a function of the LO frequency. Simulated IIP2 (line), modeled IIP2 that includes all conductance nonlinearity and LO effect (line with squares), and modeled IIP2 that includes only the g_m nonlinearity and no LO slope effect (line with triangles) for $V_C = 0.3$ V, $V_{\rm LO} = 0.6$ V, and $\alpha = 0.1$.

TABLE I
MODEL ESTIMATION TIME

	Estimation time (second)	Simulation time (second)
Noise	0.06	2.2
IIP3	0.07	3.7
IIP2	1.68	7.4

VI. CONCLUSION

A simple closed-form model for the fast and accurate estimation of the noise, IIP3, and IIP2 of the active mixer has been presented. The mixer noise can be estimated by two ac noise calculations, with an error that is smaller than 2 dB, while the calculation time is about 40 times shorter than when using a commercial simulator. The model has shown that the decreasing transistor output resistance in deep-submicrometer technologies, rather than the output capacitance, is a dominant reason for the flicker-noise leakage. Any flicker-noise cancellation technique should include the effect of output resistance. By properly increasing W and L of $\rm M_3$ and by using longer switch pair transistors $(\rm M_1/M_2)$, the noise performance can be improved, while no degradation on gain is introduced. The mixer IIP3 can be estimated by one time-invariant nonlinearity calculation, with an error that is smaller than 1 dBm, while the calculation time is reduced with a factor of 50. The slope of the LO has a little effect on the IIP3. However, the LO slope, together with the cross-modulation nonlinearity and output conductance nonlinearity in the triode region, significantly contribute to the single-ended IM2 of the mixer. Therefore, the accuracy of the IIP2 estimation is highly dependent on the good capture of the LO waveform. Neglecting the LO slope or only considering the transconductance nonlinearity will overestimate the single-side IM2 and will significantly underestimate the differential IIP2. Other than introducing mismatches to the mixer, tuning the LO slope can be a new approach in IIP2 calibration.

APPENDIX

A general distortion model for amplifiers that is presented in [29] is utilized to derive the IM3 and IM2 of the cascode amplifier shown in Fig. 24, where $V_{\rm IN}(\cos\omega_1 t + \cos\omega_2 t)$ is the two-tone input signal.

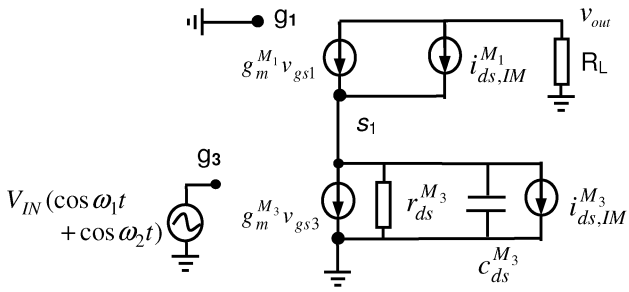

Fig. 24. Model for the IM2/IM3 calculation of the cascode amplifier.

For an IM3 to the first order, we include all third-order resistive nonlinearities between the drain–source terminals, and the output IM3 is given by

$$\begin{aligned} v_{\mathrm{IM_3}} &= H^{M_3}(2\omega_1-\omega_2)\cdot i_{\mathrm{ds,IM3}}^{M_3} + H^{M_1}(2\omega_1-\omega_2)\cdot i_{\mathrm{ds,IM3}}^{M_1} \\ &\approx \frac{-3\times V_{\mathrm{IN}}^3 R_L}{4\times\left(1+g_m^{M_1} Z_{\mathrm{ds}}^{M_3}\right)} \\ &\quad\times\left[Z_{\mathrm{ds}}^{M_3} g_m^{M_1}\cdot\left(g_{m3}^{M_3}-g_{x21}^{M_3}\right)+g_{m3}^{M_1}-g_{\mathrm{ds3}}^{M_1}\left(g_{m1}^{M_3}R_L\right)^3\right. \\ &\quad\left.+g_{x12}^{M_1}\cdot\left(g_{m1}^{M_3}R_L\right)^2-g_{x21}^{M_1}\cdot g_{m1}^{M_3}R_L\right] \end{aligned} \tag{33}$$

where $i_{\mathrm{ds,IM3}}^{M_1}$ and $i_{\mathrm{ds,IM3}}^{M_3}$ are the IM3 current components of transistor $\mathrm{M_1}$ and $\mathrm{M_3}$, respectively, and $H^{M_1}(2\omega_1-\omega_2)$ and $H^{M_3}(2\omega_1-\omega_2)$ are the gains from IM3 current components to the voltage output.

For the IM2, we include all the nonlinearities between the drain–source terminals, and the output IM2 is given by

$$\begin{aligned} v_{\mathrm{IM_2}} &= H^{M_3}(\omega_1-\omega_2)\cdot i_{\mathrm{ds,IM2}}^{M_3} + H^{M_1}(\omega_1-\omega_2)\cdot i_{\mathrm{ds,IM2}}^{M_1} \\ &\approx V_{\mathrm{IN}}^2\frac{-r_{\mathrm{ds}}^{M_3} g_m^{M_1} R_L}{1+g_m^{M_1} r_{\mathrm{ds}}^{M_3}}\cdot i_{\mathrm{ds,IM2}}^{M_3} + \frac{-R_L}{1+g_m^{M_1} r_{\mathrm{ds}}^{M_3}}\cdot i_{\mathrm{ds,IM2}}^{M_1} \\ &\approx \frac{R_L\cdot V_{\mathrm{IN}}^2}{1+g_m^{M_1} r_{\mathrm{ds}}^{M_3}} \\ &\quad\cdot\left\{r_{\mathrm{ds}}^{M_3} g_m^{M_1}\cdot\left[-g_{m2}^{M_3}-g_{\mathrm{ds2}}^{M_3}\cdot\left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)^2\right.\right. \\ &\quad\left.+g_{x11}^{M_3}\cdot\left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)\right] \\ &\quad+\left[-g_{m2}^{M_1}-g_{\mathrm{ds2}}^{M_1}\cdot\left(g_{m1}^{M_3}R_L\right)^2\right. \\ &\quad\left.\left.+g_{x11}^{M_1}\cdot g_{m1}^{M_3}R_L\right]\right\} \end{aligned} \tag{34}$$

where $i_{\mathrm{ds,IM2}}^{M_1}$ and $i_{\mathrm{ds,IM2}}^{M_3}$ are the IM2 current components of transistor $\mathrm{M_1}$ and $\mathrm{M_3}$, respectively, and $H^{M_1}(\omega_1-\omega_2)$ and $H^{M_3}(\omega_1-\omega_2)$ are the gains from the IM2 current component to the voltage output. Assuming that the second-order nonlinearity of $\mathrm{M_1}$ and $\mathrm{M_3}$ are on the same order of magnitude and $(H^{M_3}(\omega_1-\omega_2)/H^{M_1}(\omega_1-\omega_2)) \approx r_{\mathrm{ds}}^{M_3} g_m^{M_1} \gg 1$, then the IM2 contribution of $\mathrm{M_3}$ is dominant. Equation (34) can be simplified to

$$\begin{aligned} v_{\mathrm{IM2,+}} &\approx \frac{r_{\mathrm{ds}}^{M_3} g_m^{M_1} R_L}{1+r_{\mathrm{ds}}^{M_3} g_m^{M_1}}\cdot V_{\mathrm{IN}}^2 \\ &\quad\cdot\left[-g_{m2}^{M_3}-g_{\mathrm{ds2}}^{M_3}\cdot\left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)^2+g_{x11}^{M_3}\cdot\left(\frac{g_{m1}^{M_3}}{g_{m1}^{M_1}}\right)\right]. \end{aligned} \tag{35}$$

REFERENCES

[1] S. Chehrazi, A. Mirzaei, and A. A. Abidi, "Noise in current-commutating passive FET mixers," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 2, pp. 332–344, Feb. 2010.

[2] S. Chehrazi, A. Mirzaei, and A. A. Abidi, "Second-order intermodulation in current-commutating passive FET mixers," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 12, pp. 2256–2568, Dec. 2009.

[3] H. Khatri, P. S. Gudem, and L. E. Larson, "Distortion in current commutating passive CMOS downconversion mixers," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 11, pp. 2671–2681, Nov. 2009.

[4] A. V. Do, C. C. Boon, M. A. Do, K. S. Yeo, and A. Cabuk, "An energy-aware CMOS receiver front end for low-power 2.4-GHz applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 10, pp. 2675–2684, Oct. 2010.

[5] N. Kim, V. Aparin, and L. E. Larson, "A resistively degenerated wideband passive mixer with low noise figure and high $\mathrm{IIP_2}$," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 4, pp. 820–829, Apr. 2010.

[6] W. Cheng, A. J. Annema, and B. Nauta, "A multi-step P-cell for LNA design automation," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2008, pp. 2550–2553.

[7] H. Darabi and A. A. Abidi, "Noise in RF-CMOS mixers: A simple physical model," *IEEE J. Solid-State Circuits*, vol. 35, no. 1, pp. 15–25, Jan. 2000.

[8] M. T. Terrovitis and R. G. Meyer, "Noise in current-commutating CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 34, no. 6, pp. 772–783, Jun. 1999.

[9] T. Melly, A.-S. Porret, C. C. Enz, and E. A. Vittoz, "An analysis of flicker noise rejection in low-power and low-voltage CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 36, no. 1, pp. 102–109, Jan. 2001.

[10] J. Lerdworatawee and W. Namgoong, "Generalized linear periodic time-varying analysis for noise reduction in an active mixer," *IEEE J. Solid-State Circuits*, vol. 42, no. 6, pp. 15–25, Jun. 2007.

[11] M. T. Terrovitis and R. G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 35, no. 10, pp. 1461–1473, Jun. 2000.

[12] D. Manstretta, M. Brandolini, and F. Svelto, "Second-order intermodulation mechanisms in CMOS downconverters," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 394–406, Mar. 2003.

[13] G. Theodoratos, A. Vasilopoulos, G. Vitzilaios, and Y. Papananos, "Calculating distortion in active CMOS mixers using Volterra series," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 2249–2252.

[14] P. Dobrovolny, G. Vandersteen, P. Wambacq, and S. Donnay, "Analysis and white-box modeling of weakly nonlinear time-varying circuits," in *Proc. Des., Autom. Test Eur. Conf. Exhib.*, 2003, pp. 624–629.

[15] J. Yoon, H. Kim, C. Park, J. Yang, H. Song, S. Lee, and B. Kim, "A new RF CMOS Gilbert mixer with improved noise figure and linearity," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 3, pp. 626–631, Mar. 2008.

[16] J. Park, C.-H. Lee, B. Kim, and J. Laskar, "Design and analysis of low flicker-noise CMOS mixers for direct-conversion receivers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4372–4380, Dec. 2006.

[17] C. D. Hull and R. Meyer, "A systematic approach to the analysis of noise in mixers," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 40, no. 12, pp. 909–919, Dec. 1993.

[18] M. T. Terrovitis, K. S. Kundert, and R. G. Meyer, "Cyclostationary noise in radio-frequency communication systems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 11, pp. 1666–1671, Jan. 2002.

[19] H. Darabi and J. Chiu, "A noise cancellation technique in active RF-CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2628–2632, Dec. 2005.

[20] B. Razavi, "Design considerations for future RF circuits," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 741–744.

[21] A. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[22] M. Brandolini, P. Rossi, D. Sanzogni, and F. Svelto, "A +78 dBm IIP2 CMOS direct downconversion mixer for fully integrated UMTS receivers," *IEEE J. Solid-State Circuits*, vol. 41, no. 3, pp. 552–559, Mar. 2006.
[23] K. Kivekas, A. Parssinen, J. Ryynanen, J. Jussila, and K. Halonen, "Calibration techniques of active BiCMOS mixers," *IEEE J. Solid-State Circuits*, vol. 37, no. 6, pp. 766–769, Jun. 2002.
[24] M. W. Hwang, G. H. Cho, S. Y. Yoo, J. C. Lee, S. M. Ock, S. K. Min, S. H. Beck, K. Lim, S. Han, and J. Lee, "A high IIP2 direct-conversion receiver using even-harmonic reduction technique for cellular CDMA/PCS/GPS applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 9, pp. 2934–2943, Oct. 2008.
[25] K. Dufrene, Z. Boos, and R. Weigel, "Digital adaptive IIP2 calibration scheme for CMOS downconversion mixers," *IEEE J. Solid-State Circuits*, vol. 43, no. 11, pp. 2434–2445, Nov. 2008.
[26] W. Kim, S. G. Yang, Y. K. Moon, J. Yu, H. Shin, W. Choo, and B. H. Park, "IP2 calibrator using common mode feedback circuitry," in *Proc. IEEE ESSCIRC*, 2005, pp. 231–234.
[27] K. Dufrene and R. Weigel, "A novel IP2 calibration method for low-voltage downconversion mixers," in *Proc. IEEE RFIC Symp.*, 2006, pp. 289–292.
[28] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*. Norwell, MA: Kluwer, 1998.
[29] W. Cheng, A. J. Annema, J. A. Croon, D. B. M. Klaasen, and B. Nauta, "A general weak nonlinearity model for LNAs," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2008, pp. 221–224.
[30] [Online]. Available: http://www.nxp.com/models/mos_models/psp/

Wei Cheng was born in Wuhan, China, in 1980. He received the B.E. degree in measuring technology and instrument from Xiamen University, Xiamen, China, in 2002 and the M.Sc. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 2006, where he has been working toward the Ph.D. degree on the subject of RF building block modeling in the IC-Design Group since 2006.

From 2002 to 2003, he was an academic employee with the Department of Mechanical and Electrical Engineering, Xiamen University.

Anne Johan Annema (M'00) received the M.Sc. degree in electrical engineering and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1990 and 1994, respectively.

In 1995, he joined Philips Research, Eindhoven, The Netherlands, where he worked on a number of physics-electronics-related projects ranging from low-power low-voltage circuits, fundamental limits on analog circuits in conjunction with process technologies, and high voltage in baseline CMOS to feasibility research of future CMOS processes for analog circuits. Since 2000, he has been an Associate Professor with the IC-Design Group, Department of Electrical Engineering, University of Twente, with main focus on physics, analog- and mixed-signal electronics, and deep-submicrometer technologies and their joint feasibility aspects for accurate RF applications. He is also a Part-time Consultant in the industry, and in 2001, he cofounded ChipDesignWorks. He holds ten patents in circuit design.

Dr. Annema was the recipient of the 2003 and 2006 Opel (educational award).

Jeroen A. Croon was born in Delft, The Netherlands, in 1975. He received the M.Sc. degree in applied physics from the Delft University of Technology, Delft, in 1998 and the Ph.D. degree in electrical engineering from the Catholic University of Leuven, Leuven, Belgium, in 2004 for his study on the matching properties of deep-submicrometer MOSFETs.

After obtaining his M.S. degree, he started working at IMEC, Belgium. After obtaining his Ph.D. degree, he worked with IMEC on the characterization of MOSFETs fabricated on germanium substrates and on the study of variability in advanced CMOS processes. Since 2005, he has been with the Compact Modeling Group, NXP Semiconductors Research, Eindhoven, The Netherlands, working on the modeling of RF circuit blocks.

Bram Nauta (M'91–SM'03–F'07) was born in Hengelo, The Netherlands, in 1964. He received the M.Sc. degree (*cum laude*) in electrical engineering and the Ph.D. degree on the subject of analog CMOS filters for very high frequencies from the University of Twente, Enschede, The Netherlands, in 1987 and 1991, respectively.

In 1991, he joined the Mixed-Signal Circuits and Systems Department, Philips Research, Eindhoven, The Netherlands, where he worked on high-speed AD converters and analog key modules. In 1998, he returned to the University of Twente as a Full Professor, heading the IC-Design Group, which is part of the CTIT Research Institute. Moreover, he is also a Part-time Consultant in the industry, and in 2001, he cofounded ChipDesignWorks. His Ph.D. dissertation was published as a book entitled as *Analog CMOS Filters for Very High Frequencies* (Springer, 1993), and he received the "Shell Study Tour Award" for his Ph.D. work. His current research interest is high-speed analog CMOS circuits.

Dr. Nauta is a member of the technical program committees of the International Solid State Circuits Conference (ISSCC), where he is the European Chair, the European Solid State Circuit Conference, and the Symposium on VLSI circuits. He was the corecipient of the ISSCC 2002 and 2009 "Van Vessem Outstanding Paper Award.' He served as a distinguished lecturer of the IEEE and an elected member of the IEEE-SSCS AdCom. From 1997 to 1999, he served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: Analog and Digital Signal Processing. After this, he served as a Guest Editor and an Associate Editor (2001–2006) of the IEEE JOURNAL OF SOLID-STATE CIRCUITS, where he has been the Editor-in-Chief since 2007.