A Highly Linear Broadband CMOS LNA Employing Noise and Distortion Cancellation

Wei-Hung Chen, *Student Member, IEEE*, Gang Liu, *Member, IEEE*, Boos Zdravko, and Ali M. Niknejad, *Member, IEEE*

***Abstract*—A broadband inductorless low-noise amplifier (LNA) design that utilizes simultaneous noise and distortion cancellation is presented. Concurrent cancellation of the intrinsic third-order distortion from individual stages is exhibited with the common-gate and common-source cascade. The LNA IIP_3 is then limited by the second-order interaction between the common source and common gate stages, which is common in all cascade amplifiers. Further removal of this third-order distortion is achieved by incorporating a second-order-distortion-free circuit technique in the common gate stage. Implemented in 0.13 μm CMOS technology, this LNA achieved $+16$ dBm IIP_3 in both the 900 MHz and 2 GHz bands. Measurements demonstrate that the LNA has a minimum internal gain of 14.5 dB, noise figure of 2.6 dB from 800 MHz to 2.1 GHz while drawing 11.6 mA from 1.5 V supply voltage.**

***Index Terms*—Broadband LNA, distortion cancellation, noise cancellation, volterra series analysis.**

I. Introduction

THE emerging 4G telecom system envisages ubiquitous wireless connectivity that supports multiple radio standards across multiple frequency bands and features reconfigurability for agile service switching and adaptive power consumption in response to radio dynamics [1]. One key challenge in bringing out a multistandard multimode front-end resides in fulfilling the high linearity and low noise over a wide frequency range [2]. Even in a "digital" receiver application, employing discrete-time signal processing, a linear front-end amplifier is required to reject the noise and relax the performance of the subsequent samplers [3]–[5]. The straight solution for multistandard multimode front-ends employs parallel narrowband receiving paths at cost of die and board area, high pin count, and lack of reconfigurability. Much research activity has focused on the design of a tunable or broadband front-end amplifier. For instance [6] used multiband feedback in conjunction with a common-gate input stage to realize a tunable narrowband amplifier, so that the input pin can be shared. As opposed to the multiple narrowband approach, recent demonstration of ultra-wideband low-noise amplifiers (LNAs) ranging from hundreds of MHz up to 10 GHz suggests an alternative that uses single LNA for contiguous broadband signal receiving [7]–[10]. The inductorless resistive feedback LNA in [10] achieved below 3 dB noise figure (NF) and other performances comparable to its narrowband counterparts by exploiting high $f_T/f_{\max}$ transistors available from nanoscale CMOS. In [8] and [9] another broadband topology, i.e., the common-gate amplifier, was incorporated and achieved low noise figure with the aid of the noise cancellation scheme.

While the noise and bandwidth of nanoscale CMOS improve with scaling, unfortunately the linearity deteriorates with supply voltage and high-field mobility effects [12], [13]. Contrary to the diminishing device linearity, the multimode front-ends require high linearity to suppress the cross-modulation/inter-modulation due to the increased co-existence of adjacent blockers or transmitter on-chip leakage [14]. The required LNA out-of-band IIP_3 in current CDMA and WCDMA 3GPP systems is at least $+7.6$ [15] and $+0$ dBm [2], respectively and expected to grow in the multimode scenario. State-of-the-art broadband LNAs rely on non-baseline technology options, such as thick-oxide high V_t transistor [8] or elevated supply voltage [10], to deliver an appreciable IIP_3. A prior-art MOSFET linearization scheme manipulates the different polarity of the second-order g_m derivative from weak to strong inversion region and has achieved extraordinary IIP_3 in narrowband applications [16], [17]. Since the transconductance of the multi-gate transistor (MGTR) is inherently broadband, it is a good starting point in the pursuit for broadband linearization scheme.

Given the multiple tasks to be fulfilled in multimode LNAs, in this paper we demonstrate that a wideband LNA with high dynamic range can be realized if we incorporate the intrinsic device properties together with noise and distortion cancellation. An LNA topology suitable for performing both cancellations is presented. Design tradeoffs for optimum LNA noise cancellation are derived. The lowest achievable NF by noise cancellation is predicted by taking the LNA power consumption into consideration. We also show that for distortion cancellation, since a cascade amplifier is used in the noise cancellation technique to subtract the dominant amplifier noise, another source of third-order distortion originating from the second-order interaction dominates the residual distortion of the amplifier after the MGTR scheme is applied. By proper care in the transconductor design, we present an amplifier that highlights both noise and distortion cancellation. The target application for this design is for current multimode 3GPP systems, while the techniques are generally applicable to many wireless communication systems.

This paper is organized as follows. Section II begins with noise analysis of the proposed LNA. Section III continues

Manuscript received September 15, 2007; revised January 20, 2008. This work was supported by the National Science Foundation Infrastructure Grant 0403427, and project funding from the U.C. Micro program. Wafer fabrication was supported by Infineon Technologies.

W.-H. Chen and A. M. Niknejad are with the Berkeley Wireless Research Center, University of California, Berkeley, CA 94707 USA (e-mail: whchen@eecs.berkeley.edu).

G. Liu was with the Berkeley Wireless Research Center, University of California, Berkeley, CA 94707 USA, and is now with Marvell Semiconductor Inc., Santa Clara, CA 95054 USA.

B. Zdravko is with Infineon Technologies, D-81739 Munich, Germany.

Digital Object Identifier 10.1109/JSSC.2008.920335

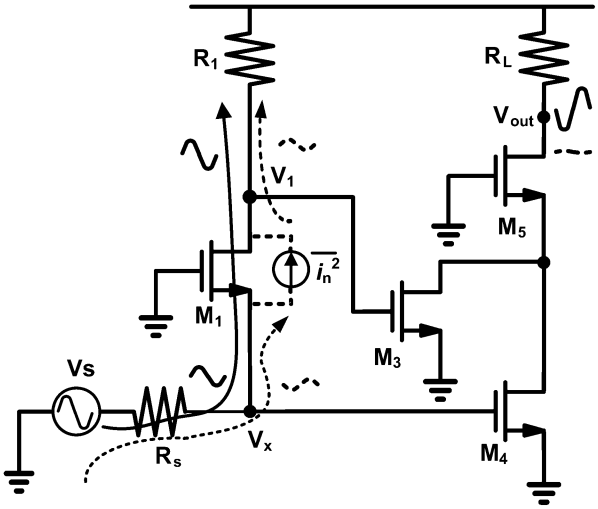

Fig. 1. Simplified small-signal schematic of the proposed LNA.

with Volterra series distortion analysis and explains multiple linearization techniques used for distortion cancellation. The implementation and measurement results are provided in Sections IV and V.

II. Broadband Noise Cancellation Design

Noise cancellation is known as an effective technique to improve the inferior NF in conventional broadband LNAs [18]. Fig. 1 shows the simplified small-signal schematic of the noise-canceling common-gate LNA. A similar implementation targeting UWB application was presented by [9] without focusing on the linearity, which is highlighted in this paper. Despite the design and publication of many noise-canceling LNAs, one unanswered question is the ultimate achievable LNA NF by this scheme. This is addressed in this section by extending the existing noise analysis techniques and taking applicable design practices into consideration [19]. It should be noted that in a narrowband application, the minimum achievable noise figure is simply the $\mathrm{NF}_{\min}$ of the transistor. But since the optimum source impedance for a noise match varies with frequency, it is difficult in practice to realize a broadband low noise amplifier with a good power match. Moreover, the possibility to also cancel the distortion of an amplifier is an exciting opportunity which motivated this research.

The principle of noise cancellation in Fig. 1 can be briefly explained as follows. The input signal, say, a current denoted by solid line, undergoes feed-forward voltage amplification by transistors M_3 and M_4 whereas the channel thermal noise of transistor M_1, the dotted signal, undergoes subtraction at output node due to two correlated but out-of-phase noise voltages at V_x and V_1. At frequencies well below f_T, the input resistance R_{in} and voltage gain A_v are given by

$$R_{\text{in}} = \frac{R_1 + r_{o1}}{1 + g_{m1} r_{o1}} \tag{1}$$

$$A_v = -\left(g_{m4} + g_{m3} \left(\frac{1 + g_{m1} r_{o1}}{1 + \frac{r_{o1}}{R_1}} \right) \right) \times R_L \tag{2}$$

In (2), the output conductance of transistor M_3 and M_4 is ignored, while it is important to consider the output conductance of M_1. Assuming only thermal noise caused by resistors and MOSFET channel current, the LNA noise factor F is derived and expressed in terms of A_v, R_{in} and device values

$$F = 1 + \frac{g_{m3}^2 R_1 + (\gamma/\alpha)\left(g_{m1} R_T^2 \left(g_{m3} \frac{R_1}{R_s} - g_{m4} \right)^2 + g_{m3} + g_{m4} \right) + \frac{1}{R_L}}{R_L^{-2} R_s^{-1} \cdot (R_s \,\|\, R_{\text{in}})^2 \cdot A_v^2} \tag{3}$$

where γ is the MOSFET noise parameter, $\alpha = g_m / g_{d0}$, and

$$R_T = R_s \| \frac{r_{o1}}{1 + R_1/R_s} \| \frac{1}{g_{m1}}$$

The terms in the numerator of (3) refer to the output noise due to R_1, M_1, M_3, M_4, and R_L, respectively. M_1's noise, conventionally as large as the one caused by the source resistance R_s in a non-cancellation case, now becomes a residual result of g_{m3} and g_{m4} subtraction. A residual factor

$$\delta = \frac{R_s}{R_1} \frac{g_{m4}}{g_{m3}} - 1 \tag{4}$$

defines the degree of cancellation of M_1 noise. $\delta = -1$ represents disabling M_4 and no cancellation while $\delta = 0$ represents full cancellation of M_1 noise. δ can be any value greater than -1. To capture the required input match, we introduce another error factor

$$\varepsilon_{rr} = \frac{R_{\text{in}}}{R_s} - 1 = \frac{2 S_{11}}{1 - S_{11}} \tag{5}$$

For instance, $\varepsilon_{rr} = 0.22$ corresponds to $S_{11} = -20$ dB. Incorporating ε_{rr} and δ, we derive the relative noise factor of each device by dividing the individual output noise by that of the source resistance R_s

$$F_{R_1} = \frac{R_s}{R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{6}$$

$$F_{\mathrm{M}_1} = \frac{(\gamma/\alpha) g_{m1} R_T^2 \delta^2}{R_s} \times \Theta(\varepsilon_{rr}, \delta) \tag{7}$$

$$F_{\mathrm{M}_3} = \frac{(\gamma/\alpha)}{g_{m3} R_1} \times \frac{R_s}{R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{8}$$

$$F_{\mathrm{M}_4} = \frac{(\gamma/\alpha)(1+\delta)}{g_{m3} R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{9}$$

$$F_{R_L} = \frac{1}{g_{m3}^2 R_L R_1} \times \frac{R_s}{R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{10}$$

$$\Theta(\varepsilon_{rr}, \delta) = \frac{(2 + \varepsilon_{rr})^2}{((1 + \varepsilon_{rr})(1 + \delta) + 1)^2} \tag{11}$$

$$F = 1 + F_{R_1} + F_{\mathrm{M}_1} + F_{\mathrm{M}_3} + F_{\mathrm{M}_4} + F_{R_L} \tag{12}$$

The normalization factor $\Theta(\varepsilon_{rr}, \delta)$ is derived from the denominator of (3) and reflects the change of R_s output noise as the condition of input match and M_1 noise cancellation vary. Notice that $F_{\mathrm{M}_4}/F_{\mathrm{M}_3} = (1 + \delta)(R_1/Rs)$, and using (4) results in $F_{\mathrm{M}_4}/F_{\mathrm{M}_3} = g_{m4}/g_{m3}$. F_{R_L} is a fraction of F_{M_3} by an

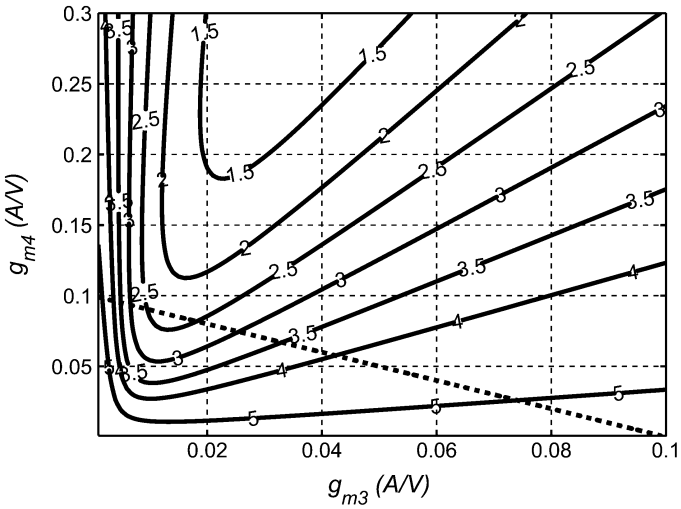

Fig. 2. NF contour plot with different g_{m3} and g_{m4} at $g_{m1} = 25$ mA/V, $R_1 = 464\ \Omega$, $g_m r_o = 22$, $\gamma/\alpha = 2.5$.

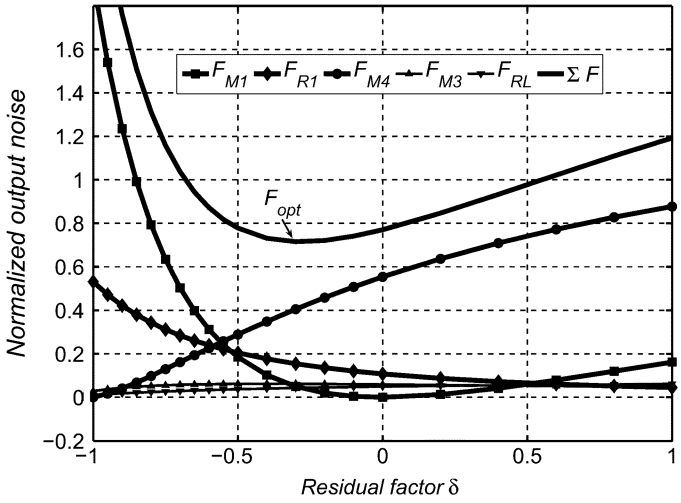

Fig. 3. Relative output noise with different δ and fixed LNA transconductor current efficiency $= 12.5$ mA/V.

amount of $(\gamma/\alpha)/g_{m3}R_L$, which is a consequence of the cascade effect of noise suppression. The dependency of F_{M_1} on δ^2 implies noise cancellation by proper design choice. Assuming $g_{m1}r_{o1} \gg 1$ and $r_{o1} \gg R_s$, R_T in (7) can be simplified as

$$R_T \approx \frac{R_s}{1 + (R_1/r_{o1}) + g_{m1}R_s} = \frac{R_s}{g_{m1} \times (R_{\mathrm{in}} + R_s)} \tag{13}$$

and F_{M_1} is rewritten as

$$F_{\mathrm{M}_1} = \frac{(\gamma/\alpha)\delta^2}{g_{m1}R_s} \times \frac{\Theta(\varepsilon_{rr}, \delta)}{(2 + \varepsilon_{rr})^2} \tag{14}$$

From (6) and (14), if both S_{11} and δ are held unchanged, the LNA NF improves due to reduced F_{R_1}, F_{M_1}, and F_{M_3} by simultaneously increasing g_{m1} and R_1[1]. Therefore, larger g_{m1} is always favored for lower NF in all cases of δ. The maximum g_{m1}, however, is bound by the available voltage headroom. For instance, $R_1 = 464\ \Omega$ is paired with $g_{m1} = 25$ mS given $\varepsilon_{rr} = 0.22$ and $g_{m1}r_{o1} = 22$. This results in R_1 voltage drop as large as 1.16 V assuming the transistor overdrive voltage is 0.2 V and $g_m/I_D = 2/V_{\mathrm{od}}$. For nanoscale CMOS technologies, this is more than the available headroom since the supply is around 1 V.

Fig. 2 plots the noise figure contour by varying g_{m3} and g_{m4} given that g_{m1} and R_1 are maximized. The dash line in Fig. 2 corresponds to a constant current consumption of 10 mA for M_3+M_4 if the same V_{od} assumption applies. The dash line intercepts the contour at different NF values as g_{m3}, or equivalently the residual cancellation ratio δ, is swept. An optimum ratio, δ_{opt}, occurs when the dash line intercepts the turn-around of the contour curves and the LNA NF reaches its local minimum for a given current consumption, defined as $\mathrm{NF}_{\mathrm{opt}}$. In other words, for a given NF, biasing M_3 and M_4 at the turn-around of the contour curve results in the lowest current consumption. By moving the dash line up and down and checking its contour interception, the $\mathrm{NF}_{\mathrm{opt}}$ at different levels of current consumption can also be determined. There exists an analytical solution for δ_{opt} by taking

$$\left(\frac{\partial F_{R_1}}{\partial \delta} + \frac{\partial F_{\mathrm{M}_1}}{\partial \delta} + \frac{\partial F_{\mathrm{M}_3}}{\partial \delta} + \frac{\partial F_{\mathrm{M}_4}}{\partial \delta} + \frac{\partial F_{R_L}}{\partial \delta}\right)\Bigg|_{\delta=\delta_{\mathrm{opt}}} = 0 \tag{15}$$

but the result does not present obvious circuit design insight. On the other hand, the implications of (6)–(10) are easier to perceive by referring to the graphic results of Fig. 3 as we traverse the dash line in Fig. 2. First, notice that F_{M_3} and F_{R_L} are small and show less variation as compared to F_{M_4} in most regions of interest because in practice $F_{M3}/F_{M4} = 1/(1+\delta)(R_s/R_1)$ and $F_{M3}/F_{RL} = 1/g_{m3}R_L$ are usually small. The determination of $\mathrm{NF}_{\mathrm{opt}}$ is simplified by neglecting F_{M_3} and F_{R_L} and focusing on F_{M_1}, F_{R_1}, and F_{M_4} alone. Starting with the non-cancellation case, i.e., M_4 off and $\delta = -1$, the LNA NF is dominated by M_1 and partly by R_1 due to the presumed practice that R_1/R_s is large.[2] As M_4 gradually turns on, M_4 amplifies the R_s noise but cancels M_1's, causing decrease in both F_{M_1} and F_{R_1}. Moreover, since only M_1 noise experiences cancellation, F_{M_1} decreases more rapidly than F_{R_1} with a ratio of

$$\frac{F_{\mathrm{M}_1}}{F_{R_1}} = \frac{(\gamma/\alpha)g_{m1}R_1R_T^2}{R_s^2}\delta^2 \tag{16}$$

until δ approaches 0 and F_{M1} becomes 0. Beyond $\delta = 0$, F_{R_1} maintains the same trend of decline but F_{M_1} rebounds from 0 because M_1 noise deviates from full cancellation in the opposite direction. Contrary to the parabolic F_{M_1} and decaying F_{R_1}, F_{M_4} shows a monotonic increase with δ. The reason for F_{M_4} to keep increasing is that, given the assumption of constant g_m in Fig. 3, g_{m3} decreases by the same amount equal to the increase of g_{m4} as δ varies. With equal change in g_{m4}, the loss of g_{m3}, however, is magnified by R_1/R_s when referring back to the LNA input. Therefore the total output noise by R_s decreases while at the same time the M_4 noise current increases. The upward F_{M_4} and downward F_{R_1} coupled with the noise-canceling

[1]In this case g_{m3} is made smaller for constant δ but the product of $g_{m3}R_1$ remains unchanged.

[2]Substituting $g_{m1} \approx 1/R_s$ and $\varepsilon_{rr} \approx 0$ in (6) and (7), F_{M_1} is greater than F_{R_1} if $R_1/R_s > 4/(\gamma/\alpha)$.

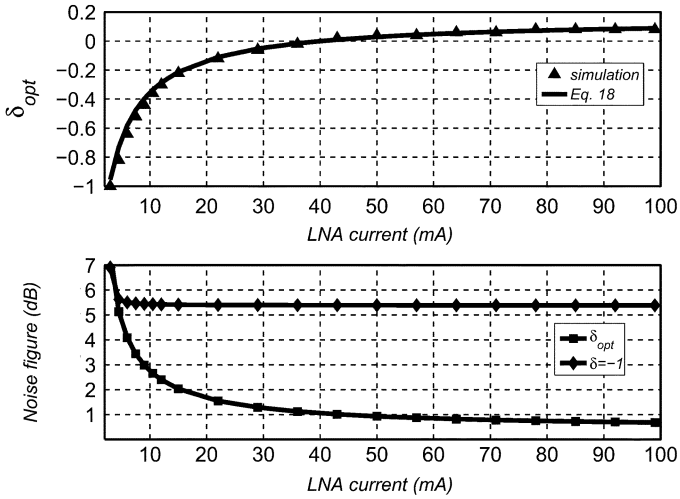

Fig. 4. NF and $\delta_{\rm opt}$ at different LNA current levels.

$F_{\rm M_1}$ then give rise to an unique $\delta_{\rm opt}$ that produces the smallest $F_{\rm M_1}+F_{R_1}+F_{\rm M_4}$. If a lower $\rm NF_{opt}$ is to be pursued, the only degree of freedom in Fig. 3 is by reshaping $F_{\rm M_4}$ through allocating larger g_{m3} and g_{m4} since F_{R_1} and $F_{\rm M_1}$ are already fixed. For instance, by scaling up $\rm M_3$ and $\rm M_4$ current with the same ratio, the curve of $F_{\rm M_4}$ moves lower across the full range of δ. This shifts $\delta_{\rm opt}$ to the right and causes a lower $\rm NF_{opt}$.

The factor of LNA current consumption, or equivalently the total transconductance $\sum g_m = g_{m1}+g_{m3}+g_{m4}$, is taken into consideration by using (4) and replacing g_{m3} with $\sum g_m - g_{m1}/1+(1+\delta)(R_1/R_s)$. For instance, (9) can be rewritten as

$$F_{\rm M_4}=\frac{(\gamma/\alpha)\left((1+\delta)+(R_1/R_s)(1+\delta)^2\right)}{(\sum g_m-g_{m1})\times R_1}\times\Theta(\varepsilon_{rr},\delta) \tag{17}$$

Taking the zero sum of $\partial F_{R1}/\partial\delta$, $\partial F_{M1}/\partial\delta$, and $\partial F_{M4}/\partial\delta$, we can derive $\delta_{\rm opt}$ as a function of LNA current consumption

$$\delta_{\rm opt}=\frac{\frac{2R_s}{R_1}(1+\varepsilon_{rr})-\frac{(\gamma/\alpha)}{R_1(\sum g_m-g_{m1})}\left(\frac{2R_1}{R_s}-\varepsilon_{rr}\right)}{\frac{2(\gamma/\alpha)g_{m1}R_L^2}{R_s}(2+\varepsilon_{rr})+\frac{(\gamma/\alpha)}{R_1(\sum g_m-g_{m1})}\left(\frac{2R_1}{R_s}-\varepsilon_{rr}-1\right)} \tag{18}$$

Although (18) neglects the effect of F_{R_1} and $F_{\rm M_1}$, a comparison of (18) with the simulation result of (15) shows good agreement in Fig. 4. Furthermore, $\rm NF_{opt}$ as a function of total transconductance can be found by substituting $\delta_{\rm opt}$ in (12)

$$\begin{aligned}F_{\rm opt}=1+\Bigg(&\frac{R_s}{R_1}+\frac{(\gamma/\alpha)g_{m1}R_T^2\delta_{\rm opt}^2}{R_s}\\&+\frac{(\gamma/\alpha)\left(1+\delta_{\rm opt}+\frac{R_1}{R_s}(1+\delta_{\rm opt})^2\right)}{R_1(\Sigma g_m-g_{m1})}\\&\times\left(1+\frac{R_s}{R_1}\frac{1}{1+\delta}\left(1+\frac{1}{(\gamma/\alpha)g_{m3}R_L}\right)\right)\Bigg)\\&\times\Theta(\varepsilon_{rr},\delta_{\rm opt})\end{aligned} \tag{19}$$

A. Power-Efficient Noise Cancellation Design

In a mobile environment, it is desirable to minimize both LNA NF and its bias current at the same time. For conventional common-gate LNAs, e.g., $\rm NF_{\delta=-1}$ in Fig. 4, the noise figure barely changes with the overall LNA current because g_{m1} is constrained by input match and limits the NF[3]. With noise cancellation, this LNA NF is made smaller by reducing the value of $F_{\rm M_1}$ and F_{R_1} while maintaining the same input match and can be further reduced if more current consumption in the common-soucre stage is dissipated. The bias current–NF tradeoff of (19) is evaluated in Fig. 4. To the first order it shows an inverse relationship with current consumption, similar to the conventional $1/g_m$ dependency. Fig. 4 also indicates that designs at extreme δ values fail to fulfill low power and low noise simultaneously. In order to facilitate a power-efficient noise cancellation, recall in Fig. 3 that $F_{\rm M_1}$ starts out more prominent than F_{R_1} at $\delta=-1$ then gradually diminishes until $\delta=0$. The cross-over of $F_{\rm M_1}$ and F_{R_1} indicates that the otherwise dominant $\rm M_1$ noise is suppressed below the noise of R_1, manifesting the onset of appreciable $\rm M_1$ noise cancellation. δ_1 is found by equating $F_{\rm M_1}$ and F_{R_1}

$$\delta_1=-\frac{R_s}{R_T}\sqrt{\frac{1}{(\gamma/\alpha)g_{m1}R_1}}=-0.526 \tag{20}$$

The corresponding $\sum g_m$ is 67 mA/V by substituting δ_1 in (18). For designs with δ beyond δ_1, $F_{\rm M_1}$ and F_{R_1} continues to decrease with the increasing $\rm M_3$ and $\rm M_4$ bias current. This is also true for their first derivative. $\partial F_{R_1}/\partial\delta$ and $\partial F_{\rm M_1}/\partial\delta$, i.e., the NF reduction per delta or bias current change, are diminishing. In other words the power efficiency of the noise cancellation scheme deteriorates. The power efficiency becomes worst after δ exceeds 0. This happens because of the plus sign of $\partial F_{\rm M_1}/\partial\delta$. The $F_{\rm M_4}$ and F_{R_1} reduction per bias current increase is counteracted by the progressively increasing $F_{\rm M_1}$. Therefore, this region of operation should be avoided.

B. Low Noise Design Without Power Constraint

If the constraint on power efficiency is lifted, $\rm M_3$ and $\rm M_4$ current continues increasing and eventually reaches the point where $F_{\rm M_4}$, $F_{\rm M_3}$, and F_{R_L} become negligible. The ultimate LNA NF is then limited by R_1 and $\rm M_1$ exclusively. An estimate of δ for this lowest NF is found by neglecting $F_{\rm M_4}$, $F_{\rm M_3}$, and F_{R_L} in (15) and making $\partial F_{\rm M_1}/\partial\delta+\partial F_{R_1}/\partial\delta=0$

$$\delta_{\rm ultimate}=\frac{R_s}{(\gamma/\alpha)g_{m1}R_1R_T^2}\times\frac{1+\varepsilon_{rr}}{2+\varepsilon_{rr}}=+0.15 \tag{21}$$

Although in Fig. 4 the corresponding NF at $\delta_{\rm ultimate}$ appears to be less than 1 dB, in practice it is unfeasible because the required voltage headroom is too high. Let us consider a practical bias current, say, 20 mA, and sets the upper limit in Fig. 4. With this modification, the lowest NF with simultaneous broadband input match achieves sub 2 dB and is comparable to its narrowband counterpart of common-source stage with source inductor degeneration.

[3]From (14), $F_{\rm M_1}\approx(\gamma/\alpha)$.

III. Wideband Distortion Cancellation Design

In the circuit of Fig. 1, the distortion of the LNA output voltage is caused by the nonlinear transistor drain current since the on-chip passive resistors R_1 and R_L are ideally linear. The nonlinearity of MOSFET drain current comes from the nonlinear transconductance g_m as well as the nonlinear drain conductance g_{ds}. By designing small resistance value in shunt with the nonlinear transistor output resistance, the distortion by nonlinear g_{ds} is less noticeable and can be neglected in order to simplify the LNA design. For small signal operation, the nonlinear transconductance of NMOS is represented by a power series

$$i_{ds} = g_m \times v_{gs} + \frac{g'_m}{2!} \times v_{gs}^2 + \frac{g''_m}{3!} \times v_{gs}^3 + \cdots \tag{22}$$

The PMOS has the same g_m characteristics but the subscript notation in (22) should be interchanged to reflect its opposite carrier type. The third-order nonlinear coefficient, g''_m, features a well-known linearity sweet spot where the g''_m value crosses zero as the transistor changes bias from weak to moderate inversion. Exploiting g''_m sign inversion around the sweet spot, zero g''_m can also be realized by biasing a transistor pair, one of which at weak inversion with positive g''_m and the other at moderate inversion with negative g''_m. Although no such sweet spot exists for g'_m, the second-order nonlinear current is usually neutralized by employing the differential pair configuration such that the g'_m-induced current of each transistor appears common-mode and is rejected by taking the output differentially. Similar action is taken in the transistor pair composed of one NMOS and one PMOS transistor, both common-source configured and sharing the drain output, to achieve very good second-order linearity [20]. Consider the well-known inverter-type amplifier with single-ended input v_{in} applied to the gate of both PMOS and NMOS. As (23) shows, the single-ended output current i_{out} is free from second-order distortion if g'_m of PMOS and NMOS is well matched

$$\begin{aligned} i_{\text{out}} &= i_{ds,n} + i_{ds,p} \\ &= \left(g_{m,N} \times v_{\text{in}} + \frac{g'_{m,N}}{2} \times v_{\text{in}}^2 + \frac{g''_{m,N}}{6} \times v_{\text{in}}^3 \right) \\ &\quad - \left(g_{m,P} \times (-v_{\text{in}}) + \frac{g'_{m,P}}{2} \right. \\ &\quad \left. \times (-v_{\text{in}})^2 + \frac{g''_{m,P}}{6} \times (-v_{\text{in}})^3 \right) \\ &= (g_{m,N} + g_{m,P}) \times v_{\text{in}} + \frac{(g'_{m,N} - g'_{m,P})}{2} \\ &\quad \times v_{\text{in}}^2 + \frac{(g''_{m,N} + g''_{m,P})}{6} \times v_{\text{in}}^3 \end{aligned} \tag{23}$$

In narrowband applications the second-order nonlinearity is usually not of primary concern because it results in distortion that is outside the frequency of interest. However, due to the cascade/feedback configuration in noise-canceling LNAs, the second-order nonlinearity also contributes third-order distortion and that overpowers the residual distortion after the linearisation techniques on intrinsic third-order distortion are applied. In order to achieve very low third-order distortion, it is then equally important to linearize both second- and third-order nonlinearities.

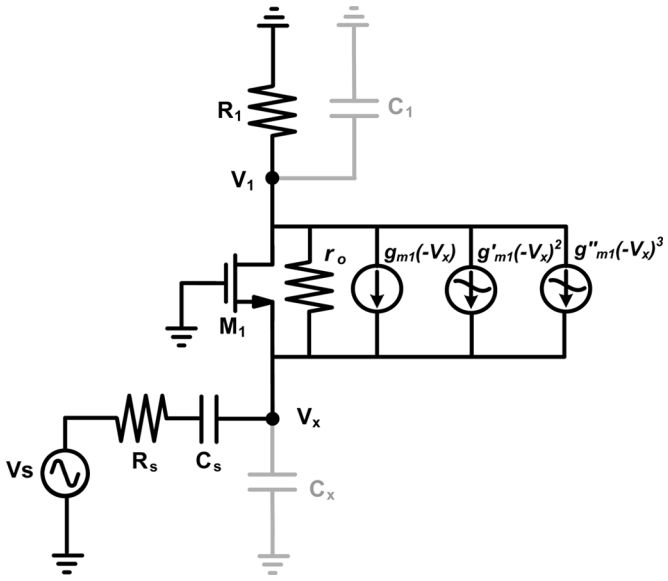

Fig. 5. Common-gate schematic for distortion analysis.

A. Simplified Distortion Analysis

The schematic of the common-gate amplifier for distortion analysis is shown in Fig. 5. C_1 and C_x are the parasitic capacitance associated at the drain and source of transistor M_1. C_s models the input coupling capacitor and can be generalized to the series network between R_s and V_x. We focus on the output current of the nonlinear transconductance and limit the analysis up to the third order. The nonlinearity of the output conductance is ignored assuming that R_1 and R_L are relatively small. To capture memory effects at high frequency, we absorb the bandwidth limiting capacitance into C_1 and C_x and apply the Volterra series around the common-gate stage. For the common-source stage with cascode transistor, the memoryless Taylor series is applied to reduce the analysis complexity, an assumption that will be checked later. By denoting

$$\begin{aligned} V_x &= A_1(s) \circ V_s + A_2(s_1, s_2) \circ V_s^2 \\ &\quad + A_3(s_1, s_2, s_3) \circ V_s^3 \end{aligned} \tag{24}$$

$$\begin{aligned} V_1 &= B_1(s) \circ V_s + B_2(s_1, s_2) \circ V_s^2 \\ &\quad + B_3(s_1, s_2, s_3) \circ V_s^3 \end{aligned} \tag{25}$$

and solving for the KCL equations, the first, second, and third-order Volterra series kernels are derived as shown in (26)–(32) on the next page, where $Z_s(s)$=R_s + $(1/sC_s)$, $Z_x(s) = (1/sC_x)$ and $Z_1(s) = R_1 \| (1/sC_1)$. $\overline{A_1(s_1)A_2(s_2, s_3)}$ represents the second-order interaction operator. Details of the derivation of (26)–(32) are presented in the Appendix. V_1 and V_x are then amplified at the common-source stage by

$$\begin{aligned} V_{\text{out}} &= \left(\left(g_{m3} \times V_1 + \frac{g'_{m3}}{2!} \times V_1^2 + \frac{g''_{m3}}{3!} \times V_1^3 \right) \right. \\ &\quad \left. + \left(g_{m4} \times V_x + \frac{g'_{m4}}{2!} \times V_x^2 + \frac{g''_{m4}}{3!} \times V_x^3 \right) \right) \\ &\quad \times Z_L(s) \end{aligned} \tag{33}$$

$Z_L(s)$ is the impedance at the LNA output. The fundamental and third-order V_{out} expressions are given by

$$\begin{aligned} &V_{\text{out,fund}} \\ &\quad = ((A_1(s) \circ V_s) \times g_{m4} + (B_1(s) \circ V_s) \times g_{m3}) \\ &\qquad \times Z_L(s) \end{aligned} \tag{34}$$

$$\begin{aligned} &V_{\text{out,3rd}} \\ &\quad = \Big(\Big(\big(A_3(s_1,s_2,s_3) \circ V_s^3\big) \times g_{m4} \\ &\qquad + \big(B_3(s_1,s_2,s_3) \circ V_s^3\big) \times g_{m3}\big) \\ &\qquad + \left((A_1(s) \circ V_s)^3 \times \frac{g''_{m4}}{6}\right. \\ &\qquad \left. + (B_1(s) \circ V_s)^3 \times \frac{g''_{m3}}{6}\right) \\ &\qquad + \left(\left(\overline{A_1(s_1)A_2(s_2,s_3)} \circ V_s^3\right) \times g'_{m4}\right. \\ &\qquad \left. + \left(\overline{B_1(s_1)B_2(s_2,s_3)} \circ V_s^3\right) \times g'_{m3}\right)\Big) \times Z_L(s) \end{aligned} \tag{35}$$

In order to deconvolve (35), first consider low frequencies, the ratio of $A_1(s)/B_1(s)$, $A_2(s_1,s_2)/B_2(s_1,s_2)$ and $A_3(s_1,s_2,s_3)/B_3(s_1,s_2,s_3)$ reduces to R_{in}/R_1, $-R_s/R_1$ and $-R_s/R_1$, respectively. Each of these terms contributes to the third-order V_{out} distortion. The first term in (35) is due to M_1's third-order distortion and cancels out in the same vein as the mechanism that cancels M_1's channel noise. The second term is due to M_3 and M_4's third-order distortion and is canceled by properly biasing M_3 and M_4 with different g''_m polarity. The criteria of these two cancellation can be formulated as

$$\frac{g_{m4}}{g_{m3}} = \frac{R_1}{R_s}, \quad \text{and} \quad \frac{g''_{m4}}{g''_{m3}} = \frac{R_1}{R_{\text{in}}} = \frac{R_1}{R_s} \times \frac{1}{1+\epsilon_{rr}} \tag{36}$$

By properly choosing M_3 and M_4's bias voltage and transistor size, the criteria of (36) can be met. Up until this point, the optimum size and bias voltage of M_3 and M_4 are already set to allow the aforementioned noise and distortion cancellation to take effect.

The third expression in (35) plays the role of second-order interaction which mixes the first- and second-order voltages at V_x and V_1 by g'_{m3} and g'_{m4}. Despite the reverse polarity between $\overline{A_1(s_1)A_2(s_2,s_3)}$ and $\overline{B_1(s_1)B_2(s_2,s_3)}$, and the same polarity of g'_{m3} and g'_{m4}, the residual value of the third term in (35) can still be substantial because g'_{m3} and g'_{m4} are fixed once M_3 and M_4 are designed to satisfy the criteria of (36). To achieve complete cancellation of (35), given the first two cancellations need to remain valid, NMOS M_1 can be replaced with a PMOS/NMOS pair such that $A_2(s_1,s_2)$ and $B_2(s_1,s_2)$ becomes null by a zeroed equivalent g'_{m1}.

The values of g_m and its derivatives for different transistors are extracted from Spectre simulation using the foundry's BSIM4 model. A calculation on IIP_3 versus second-order interaction is performed by substituting the g_m information into (35) and varying the value of g'_m of the common-gate transistor. The original single NMOS common-gate transistor in the noise cancellation design has g'_{m1} of 42 mA/V^2. Adding the PMOS together with the NMOS transistor allows the equivalent g'_m to be altered to reshape the LNA IIP_3 drastically in Fig. 6. Comparing (4) and (36), the (g_{m4}/g_{m3}) requirement in noise cancellation is different from that in distortion cancellation by $1+\delta$. Because a nonzero δ is chosen, the peak IIP_3 in Fig. 6 occurs off the center.

As far as the frequency dependency of (30)–(32) is concerned, to the first degree, $(Z_{\text{in}}(s)/Z_1(s))$ and $(Z_1(s)/Z_s(s)\,||\,Z_x(s))$, the ratio for transistors M_3 and M_4 in distortion cancellation, is insensitive to frequency change, which enlarges the bandwidth of the distortion cancellations scheme.

B. Complete Distortion Analysis

The LNA full schematic is shown in Fig. 7 which incorporates PMOS M_2 with NMOS M_1 in the common-gate stage to realize the cancellation of second-order distortion. The DC bias

$$A_1(s) = \frac{Z_1(s) + r_{o1}}{H(s)} \tag{26}$$

$$A_2(s_1,s_2) = \frac{\frac{1}{2}g'_{m1}r_{o1}Z_s(s_1+s_2)A_1(s_1)A_1(s_2)}{H(s_1+s_2)} \tag{27}$$

$$A_3(s_1,s_2,s_3) = \frac{-Z_s(s1+s2+s3)r_{o1}\left(-g'_{m1}\overline{A_1(s_1)A_2(s_2,s_3)} + \frac{1}{6}g''_{m1}A_1(s_1)A_1(s_2)A_1(s_3)\right)}{H(s_1+s_2+s_3)} \tag{28}$$

$$H(s) = Z_s(s)(1+g_{m1}r_{o1}) + (Z_1(s)+r_{o1})\left(1+\frac{Z_s(s)}{Z_x(s)}\right) \tag{29}$$

and

$$B_1(s) = \frac{Z_1(s)\times(1+g_{m1}r_{o1})}{Z_1(s)+r_{o1}}A_1(s) \tag{30}$$

$$B_2(s_1,s_2) = \frac{-Z_1(s_1+s_2)}{Z_x(s_1+s_2)\,||\,Z_s(s_1+s_2)}A_2(s_1,s_2) \tag{31}$$

$$B_3(s_1,s_2,s_3) = \frac{-Z_1(s_1+s_2+s_3)}{Z_x(s_1+s_2+s_3)\,||\,Z_s(s_1+s_2+s_3)}A_3(s_1,s_2,s_3) \tag{32}$$

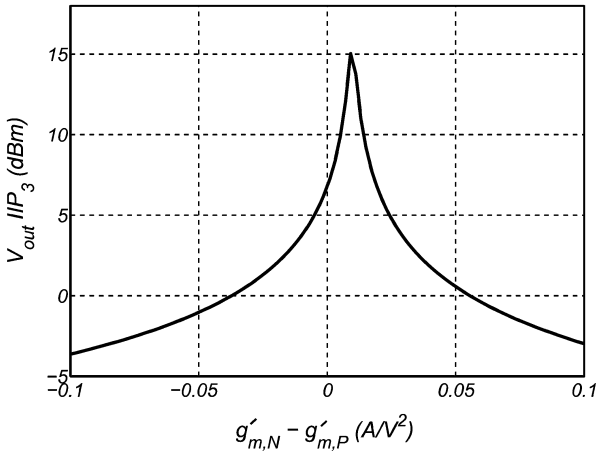

Fig. 6. $V_{\rm out}$IIP$_3$ versus g'_m of the common-gate transistors.

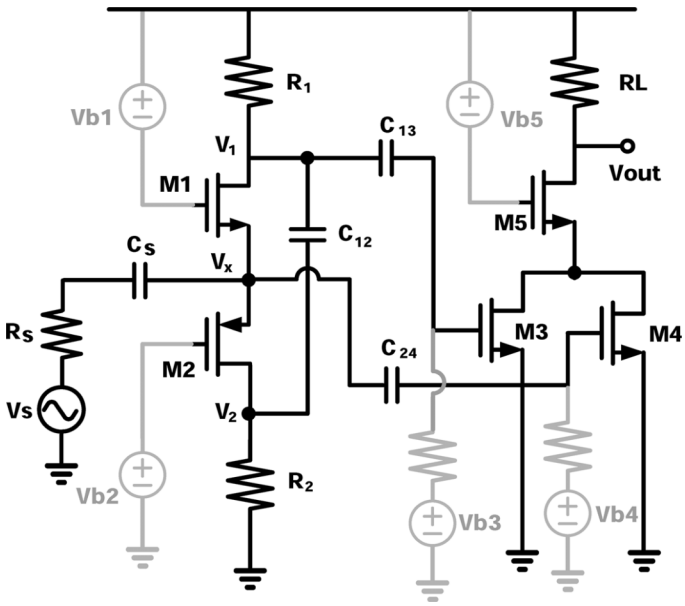

Fig. 7. Complete LNA schematic.

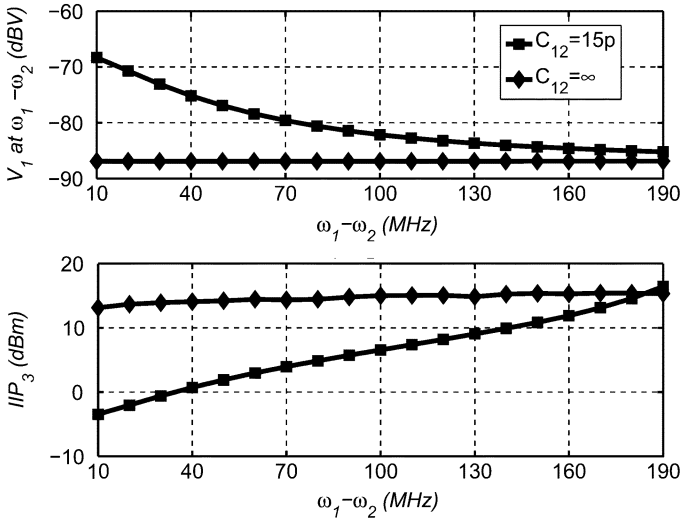

Fig. 8. Effect of C_{12} and frequency spacing on LNA IIP$_3$.

current of the PMOS-NMOS cascode is reused and the bias voltages are resolved by placing the NMOS transistor on the top of the PMOS transistor. AC equivalence of the drain node of $\rm M_1$ and $\rm M_2$ at signal frequency is achieved by the coupling capacitor C_{12}. Transistors $\rm M_3$ and $\rm M_4$ are sized for the noise and intrinsic third-order distortion cancellations. Referring the previous distortion analysis to the new schematic, at RF signal frequency and beyond, C_{12} shorts out two drain nodes, making the equivalent circuit no different from the one shown in Fig. 5. On the other hand, for frequencies much lower than RF, C_{12} presents a high impedance path between two drain nodes and breaks the LNA into two standing-alone common-gate amplifiers. This change of scenario leads the low frequency portion of the g'_m-induced current to deviate apart from the designed path in the previous simplified distortion analysis. Assuming two tones at ω_1 and ω_2 in an IIP$_3$ test, at higher $IM2$ frequency $\omega_1+\omega_2$, $\rm M_2$ distortion current goes across C_{12} and loops back through $\rm M_2-C_{12}-M_1-M_2$. Since $\rm M_2$ g'_m current is absorbed into $\rm M_1$, there will be no second-order distortion current flowing across R_1 and R_L if both transistors' g'_m is matched. Contrarily, at lower IM$_2$ frequency $\omega_1-\omega_2$, the impedance of C_{12} is large enough so the current through C_{12} is hindered. Instead, the $\rm M_1$ and $\rm M_2$ distortion current flows through R_1 and R_L and causes distortion voltage of V_x and V_1. A full Volterra series analysis taking C_{12} into account is conducted and also presented in the Appendix. It verifies a non-canceled second-order distortion in $B_2(s_1,s_2)$ and $C_2(s_1,s_2)$ if the impedance of C_{12} is not negligible. To reduce the IIP$_3$ impairment due to insufficient g'_m cancellation, the frequency difference $\omega_1-\omega_2$ or the capacitance of C_{12} needs to be large. For instance, 15 pF was picked for C_{12} in this implementation. The drawback of using large capacitor is the bandwidth reduction due to the parasitics associated with the capacitor. Fortunately the most stringent LNA IIP$_3$ is usually governed by the out-of-band blockers, which usually resides away from the desired receiving band by several tens to hundreds of MHz. This wide frequency spacing, instead of KHz-apart in-band blockers, substantially reduces the required capacitance value for C_{12}.

The above 2-tone property is checked in the simulation by replacing all coupling capacitors with ideal ones of very large capacitance. Fig. 8 shows a flat IIP$_3$ with infinite C_{12} whereas the IIP$_3$ with finite C_{12} losses as much as 15 dB and the $IM2$ voltage of V_1 increases by 18 dB as the tone spacing shrinks.

There is yet another second-order interaction that is not accounted for in the previous simplified analysis. Instead of the unilateral assumption in (33), the $\rm M_3$ and $\rm M_4$ second-order nonlinear current may flow back to the gate through overlap capacitor C_{gd} and mix with the linear voltage of V_1 and V_x to produce third-order distortion at $V_{\rm out}$. Accounting for this second-order interaction, the IIP$_3$ of the common-source stage alone is modified by an offset from g''_m,

$$\text{IIP}_3 \propto \frac{1}{\varepsilon((s_1-s_2),(s_1+s_2))} \tag{37}$$

$$\varepsilon((s_1-s_2),(s_1+s_2)) = g''_m - \frac{2g'^2_m}{3}(2K(s_1-s_2)+K(s_1+s_2)) \tag{38}$$

where $K(\omega)$ is the network function depending on the source and drain impedances [21]. From Fig. 8 this effect is much less

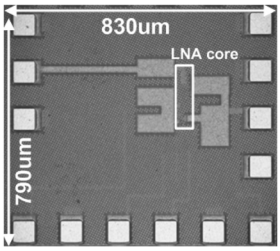

Fig. 9. Chip microphotograph.

TABLE I
DEVICE DIMENSION

M_1	$(6.52\mu\text{m}/0.12\mu\text{m})\times 6$	R_1	450Ω
M_2	$(6.52\mu\text{m}/0.12\mu\text{m})\times 17$	R_2	250Ω
M_3	$(6.52\mu\text{m}/0.12\mu\text{m})\times 30$	R_L	55Ω
M_4	$(6.52\mu\text{m}/0.12\mu\text{m})\times 50$	$C_{12,13,24}$	15 pF
M_5	$(6.52\mu\text{m}/0.12\mu\text{m})\times 30$	C_s	25 pF

noticeable as only little variation of IIP_3 is observed after C_{12} is replaced with ideal capacitor.

IV. IMPLEMENTATION

A 0.8–2.1 GHz LNA for multimode multiband 3GPP system based on the above noise and distortion cancellations was designed and fabricated in 0.13 μm CMOS technology. It was built on standard-V_t transistors and metal–insulator–metal (MIM) capacitors. Device sizes are summarized in Table I. For better device matching, transistors M_1–M_5 were duplicated from the unit transistor of identical dimension and layout. The capacitance of the MIM capacitors was chosen large enough to minimize the second-order distortion. Due to the intrinsic high linearity of this LNA, adding an on-chip output buffer for measurement would degrade the linearity performance to be observed. Therefore resistor R_L was sized close to 50 Ω so that direct measurement by placing the probe tip on the output pad is feasible. Because of the low value of R_L, simulation shows negligible IIP_3 difference when the output is checked as it is without 50 Ω probe resistance. The LNA gain measured with 50 Ω probe resistance is 6 dB smaller than otherwise available on-chip. The chip microphotograph, shown in Fig. 9, occupied an area of 830×790 μm^2 while the circuit core was only 320×310 μm^2 including the MIM capacitors.

V. MEASUREMENT RESULTS

The chip was measured by on-wafer probing. The two-port S-parameters shown in Fig. 10 are in good agreement with the simulation. The gain roll-off is attributed to the large parasitics, roughly 500 fF, associated with the MIM capacitors. The gain roll-off can be mitigated by introducing inductive compensation as demonstrated in [9]. The S_{11} at 2.1 GHz is -8.5 dB and can be improved with the addition of a 1.5–2.5 nH inductor at the LNA input, which can be absorbed into the package/board bondwires. In simulation adding this bond-wire has only a minor impact on NF and IIP_3.

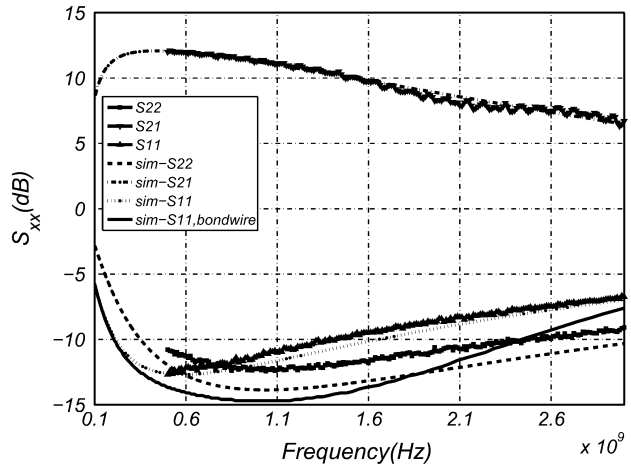

Fig. 10. Measured S parameters.

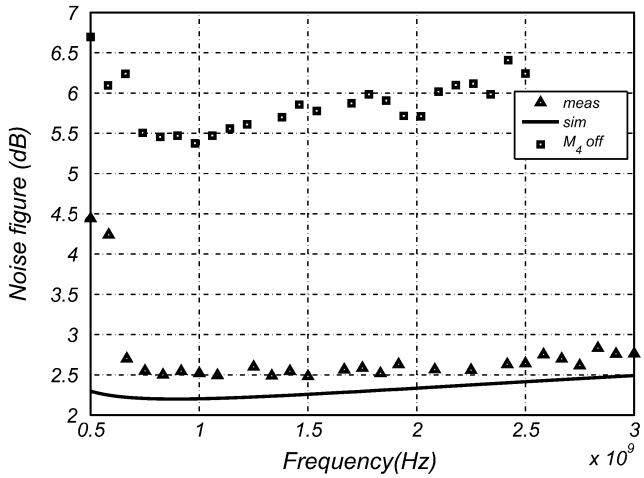

Fig. 11. Measured noise figure.

The measured noise figure of the LNA was about 2.6 dB for frequencies below 2.1 GHz. Noise cancellation was verified in Fig. 11 by turning off one of the common source transistors. Noise figure reduction as large as 3 dB was observed in the measurement. It should be noted that turning off one common source stage also changed the LNA gain. The IIP_3 of the LNA was measured with a WCDMA compliant blocker test. Two sinusoidal tones located at 1.76 GHz and 1.95 GHz modeled a strong AM blocker and transmitter on-chip leakage. They resulted in an IM_3 product at 2.14 GHz, which fell in the receiver band. Although there is no corresponding blocker test at 900 MHz band, the same two-tone spacing was kept for IIP_3 test at 900 MHz band in order to examine the IM_3 cancellation over a wide frequency range. The measured result in Fig. 12 shows very close IM_3 behavior in both bands which confirms the cancellation scheme is effective over wide frequency range. Both tests achieved a high IIP_3 of $+16$ dBm. The IM_3 cancellation scheme holds effective for blocker power level as large as -20 dBm. This relaxes the stringent isolation requirement imposed on the duplexer ahead of the LNA. The measured LNA $P_{1\text{ dB}}$ was -12 dBm, as higher order terms, e.g., IM_5, IM_7, also contributed to $P_{1\text{ dB}}$ and were not exploited in this scheme.

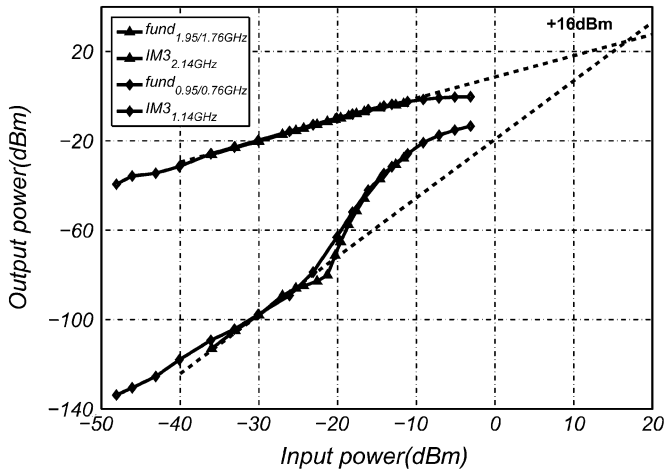

Fig. 12. Measured IIP_3.

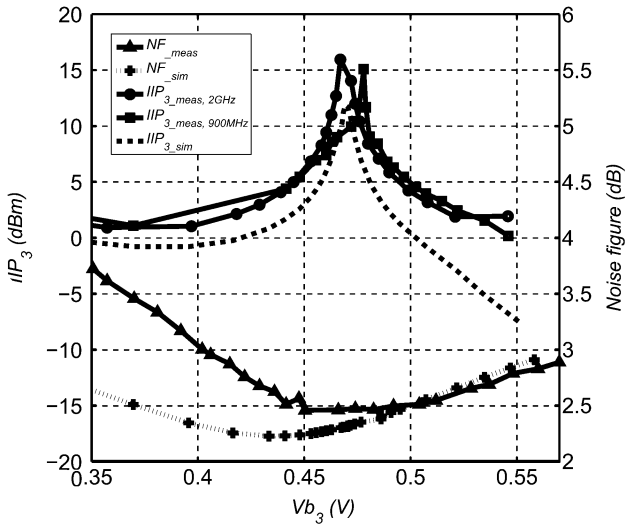

Fig. 13. Measured and simulated IIP_3 and NF sensitivity of Vb_3.

It is important to examine the sensitivity of the distortion and noise cancellation schemes. One way to emulate the effects of matching and threshold variation is to vary the bias point Vb_3. Fig. 13 compares the measured IIP_3 sensitivity versus the post-layout simulation. Discrepancy of NF in the lower Vb_3 region is noticed. A lower Vb_3 biases transistor M_3 into deep subthreshold region and drives the NF contour to the left border in Fig. 2 where it becomes more susceptible to parameter variations. Taking process and bias variations into consideration, an IIP_3 greater than $+5$ dBm and NF below 2.6 dB was achieved within a Vb_3 bias window of 50 mV. Furthermore, the variations in the common-gate stage are expected to cause less performance change since mainly the third cancellation criteria in (35) is affected. This was examined in Fig. 14 by varying Vb_1 and Vb_2 separately.

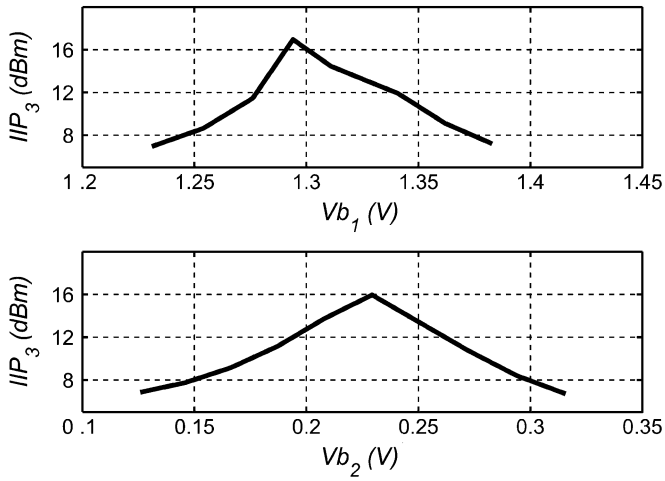

Fig. 14. Measured IIP_3 sensitivity of Vb_1 and Vb_2.

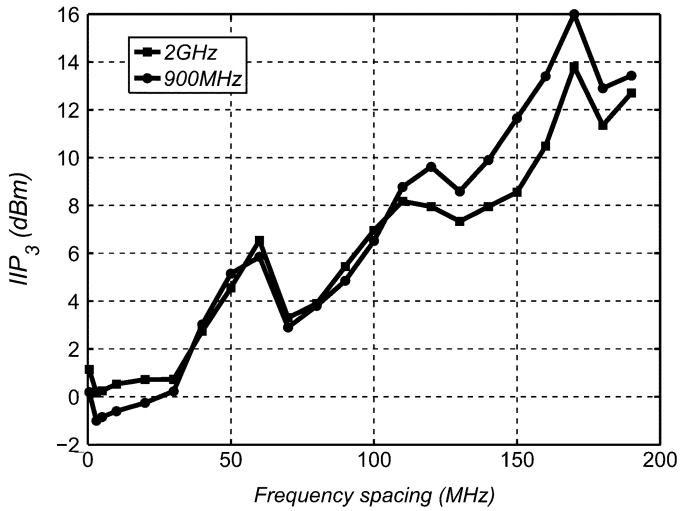

Fig. 15. Measured IIP_3 dependency on the two-tone spacing.

The IIP_3 dependency on two-tone frequencies is demonstrated in Fig. 15[4]. As previously mentioned, the cancellation of the second-order distortion degenerates as the frequency spacing between two tones narrows. For small tone spacing, only the cancellation of the transistors' intrinsic third-order distortion is effective. In that scenario the measured LNA IIP_3 is around 0 dBm and suffices most in-band IIP_3 specifications. The LNA behaves much more linearly for out-of-band blocker testing where the more stringent linearity is usually required. The LNA draws 11.6 mA from 1.5 V supply in high linearity and low noise mode. In the end, Table II summarizes the measured performances and compares this design to recently published wideband LNAs.

VI. Conclusion

A broadband very low IM_3 distortion LNA exploiting simultaneous noise and distortion cancellations was presented. Design tradeoffs in a noise-canceling common-gate

[4]The IIP_3 was calculated based on $\mathrm{IIP}_3 = \mathrm{input\ signal\ power} + 0.5 \times \mathrm{IM}_3$ at one input level and showed slight difference from the graphical extraction result in Fig. 12.

TABLE II
PERFORMANCE COMPARISON OF BROADBAND LNA

	Technology	Bandwidth	Gain	NF	IIP_3	S_{11}	Supply voltage	Power
[8]	0.13μm CMOS	$0.4-5$ GHz	19 dB	3 dB	$+1$ dBm	-10 dB	1.8 V	11.7 mW
[9]	0.18μm CMOS	$1.2-11.9$ GHz	9.7 dB	4.5 dB	-6.2 dBm	-11 dB	1.2 V	20 mW
[10]	0.09μm CMOS	$.5-8.2$ GHz	25 dB	$1.9-2.6$ dB	-4 dBm	-10 dB	2.7 V	41.85 mW
[11]	0.13μm CMOS	$3.1-10.6$ GHz	15.1 dB	2.5 dB	-5.1 dBm	-9.9 dB	1.2 V	9 mW
[18]	0.25μm CMOS	$2-1600$ MHz	13.7 dB	2.5 dB	0 dBm	no	2.5 V	35 mW
[22]	0.13μm CMOS	$2-5.2$ GHz	16 dB	$4.7-5.7$ dB	-6 dBm	-9 dB	2 V	38mW
[23]	0.065μm CMOS	$0.2-5.2$ GHz	15.6 dB	<3.5 dB	>0 dBm	<-10 dB	1.2 V	21 mW
This work	0.13μm CMOS	$0.8-2.1$ GHz	14.5 dB	2.6 dB	$+16$ dBm	-8.5 dB	1.5 V	17.4 mW

common-source cascade was investigated with and without a power constraint. A minimum noise figure below 2 dB for broadband common-gate LNA was predicted by maximally utilizing the available voltage headroom in a given technology. Full Volterra series analysis of the proposed LNA was conducted. It indicated that after cancellation of the intrinsic third-order distortions, second-order distortion dominates the residual IM_3 through second-order interaction. An inverter-like PMOS-NMOS pair was adopted to mitigate this effect. This LNA achieved a peak IIP_3 of +16 dBm over a wide bandwidth from 800–2100 MHz, while maintaining a noise figure below 2.6 dB. A prototype fabricated in a 0.13 μm CMOS technology without utilizing high-V_{dd} or thick-oxide transistors confirmed the amplifier linearity and noise cancellations.

APPENDIX A
EXPANDED VOLTERRA SERIES ANALYSIS

Derivation of the Volterra series kernels for the circuit in Fig. 7 is presented below. The Volterra series kernels for the simplified schematic in Fig. 1 is deduced from the derived results by assuming $Z_{12}(s)=0$. Referring to Fig. 7 and repeating (24)–(25) for V_x, V_1 and V_2,

$$V_x = A_1(s_1)\circ V_s + A_2(s_1,s_2)\circ V_s^2 + A_3(s_1,s_2,s_3)\circ V_s^3 \tag{A.1}$$

$$V_1 = B_1(s_1)\circ V_s + B_2(s_1,s_2)\circ V_s^2 + B_3(s_1,s_2,s_3)\circ V_s^3 \tag{A.2}$$

$$V_2 = C_1(s_1)\circ V_s + C_2(s_1,s_2)\circ V_s^2 + C_3(s_1,s_2,s_3)\circ V_s^3 \tag{A.3}$$

the following KCL equations are established:

$$i_{m1}+\frac{V_x-V_1}{r_{o1}}+i_{m2}+\frac{V_x-V_2}{r_{o2}}+\frac{V_x}{Z_x(s)} = \frac{V_s-V_x}{Z_s(s)} \tag{A.4}$$

$$i_{m1}+\frac{V_x-V_1}{r_{o1}} = \frac{V_1}{Z_1(s)}+\frac{V_1-V_2}{Z_{12}(s)} \tag{A.5}$$

$$i_{m2}+\frac{V_x-V_2}{r_{o2}} = \frac{V_2}{Z_2(s)}+\frac{V_2-V_1}{Z_{12}(s)} \tag{A.6}$$

where i_{m1} and i_{m2} are the small signal current flowing into the source of the transistors.

$$i_{m1} = -\left(g_{m1}(-V_x)+\frac{g'_{m1}}{2}(-V_x)^2+\frac{g''_{m1}}{6}(-V_x)^3\right) \tag{A.7}$$

$$= g_{m1}V_x-\frac{g'_{m1}}{2}V_x^2+\frac{g''_{m1}}{6}V_x^3$$

$$i_{m2} = g_{m2}V_x+\frac{g'_{m2}}{2}V_x^2+\frac{g''_{m2}}{6}V_x^3 \tag{A.8}$$

Substituting i_{m1} and i_{m2} with their g_m polynomial, we set up the following for the first-order Volterra kernel.

$$g_{m1}A_1(s)+\frac{A_1(s)-B_1(s)}{r_{o1}}+g_{m2}A_1(s)+\frac{A_1(s)-C_1(s)}{r_{o2}}+\frac{A_1(s)}{Z_x(s)} = \frac{1-A_1(s)}{Z_s(s)} \tag{A.9}$$

$$g_{m1}A_1(s)+\frac{A_1(s)-B_1(s)}{r_{o1}} = \frac{B_1(s)}{Z_1(s)}+\frac{B_1(s)-C_1(s)}{Z_{12}(s)} \tag{A.10}$$

$$g_{m2}A_1(s)+\frac{A_1(s)-C_1(s)}{r_{o2}} = \frac{C_1(s)}{Z_2(s)}+\frac{C_1(s)-B_1(s)}{Z_{12}(s)} \tag{A.11}$$

At the frequency of interest, $Z_{12}(s)$ is negligible such that $B_1(s)=C_1(s)$. The last two KCL equations can be combined and result in a concise answer for $A_1(s)$ and $B_1(s)$.

$$A_1(s) = \frac{(Z_1(s)\,||\,Z_2(s))+(r_{o1}\,||\,r_{o2})}{H(s)} \tag{A.12}$$

$$B_1(s) = \frac{Z_1(s)\,||\,Z_2(s)}{\left(\frac{Z_1(s)\,||\,Z_2(s)+(r_{o1}\,||\,r_{o2})}{1+(g_{m1}+g_{m2})(r_{o1}\,||\,r_{o2})}\right)}A_1(s) \tag{A.13}$$

$$H(s) = Z_s(s)(1+(g_{m1}+g_{m2})(r_{o1}\,||\,r_{o2})) + ((Z_1(s)\,||\,Z_2(s))+(r_{o1}\,||\,r_{o2})) \times\left(1+\frac{Z_s(s)}{Z_x(s)}\right) \tag{A.14}$$

Next, consider the following equations for the second-order Volterra series kernel.

$$g_{m1}A_2(s_1,s_2)-\frac{g'_{m1}}{2}A_1(s_1)A_1(s_2)$$

$$+\frac{A_2(s_1,s_2)-B_2(s_1,s_2)}{r_{o1}}$$
$$+g_{m2}A_2(s_1,s_2)+\frac{g'_{m2}}{2}A_1(s_1)A_1(s_2)$$
$$+\frac{A_2(s_1,s_2)-C_2(s_1,s_2)}{r_{o2}}+\frac{A_2(s_1,s_2)}{Z_x(s_1+s_2)}$$
$$=\frac{-A_2(s_1,s_2)}{Z_s(s_1+s_2)} \tag{A.15}$$

$$g_{m1}A_2(s_1,s_2)-\frac{g'_{m1}}{2}A_1(s_1)A_1(s_2)$$
$$+\frac{A_2(s_1,s_2)-B_2(s_1,s_2)}{r_{o1}}=\frac{B_2(s_1,s_2)}{Z_1(s_1+s_2)}$$
$$+\frac{B_2(s_1,s_2)-C_2(s_1,s_2)}{Z_{12}(s_1+s_2)} \tag{A.16}$$

$$g_{m2}A_2(s_1,s_2)+\frac{g'_{m2}}{2}A_1(s_1)A_1(s_2)$$
$$+\frac{A_2(s_1,s_2)-C_2(s_1,s_2)}{r_{o2}}=\frac{C_2(s_1,s_2)}{Z_2(s_1+s_2)}$$
$$+\frac{C_2(s_1,s_2)-B_2(s_1,s_2)}{Z_{12}(s_1+s_2)} \tag{A.17}$$

$A_2(s_1,s_2)$ and $B_2(s_1,s_2)$ are found as shown in (A.18)-(A.19) at the bottom of the page, where

$$\triangle A_2(s_1,s_2)$$
$$=\frac{1}{2}Z_{12}(s_1+s_2)A_1(s_1)A_1(s_2)$$
$$\times\frac{Z_s(s_1+s_2)}{Z_1(s_1+s_2)+Z_2(s_1+s_2)}$$
$$\times\left((g'_{m1}-g'_{m2})(r_{o1}\,\|\,r_{o2})\right.$$
$$\left.+\frac{g'_{m1}r_{o1}Z_2(s_1+s_2)-g'_{m2}r_{o2}Z_1(s_1+s_2)}{r_{o1}+r_{o2}}\right) \tag{A.20}$$

$$\triangle B_2(s_1,s_2)$$
$$=-\frac{1}{2}Z_{12}(s_1+s_2)A_1(s_1)A_1(s_2)$$
$$\times\frac{Z_1(s_1+s_2)}{Z_1(s_1+s_2)+Z_2(s_1+s_2)}\frac{1}{r_{o1}+r_{o2}}$$
$$\times\left(g'_{m1}r_{o1}(Z_2(s_1+s_2)+r_{o2})\left(1+\frac{Z_s(s_1+s_2)}{Z_x(s_1+s_2)}\right)\right.$$
$$+Z_s(s_1+s_2)(g'_{m2}r_{o2}(1+g_{m1}r_{o1})$$
$$\left.+g'_{m1}r_{o1}(1+g_{m2}r_{o2}))\right) \tag{A.21}$$

$$\triangle H(s_1,s_2)$$
$$=Z_{12}(s_1+s_2)\frac{Z_s(s_1,s_2)}{Z_1(s_1,s_2)+Z_2(s_1,s_2)}\frac{1}{r_{o1}+r_{o2}}$$
$$\times\left(\frac{(r_{o1}+Z_1(s_1+s_2))(r_{o2}+Z_2(s_1+s_2))}{Z_x(s_1+s_2)\,\|\,Z_s(s_1+s_2)}\right.$$
$$+((1+g_{m1}r_{o1})(r_{o2}+Z_2(s_1+s_2))$$
$$\left.+(1+g_{m2}r_{o2})(r_{o1}+Z_1(s_1+s_2)))\right) \tag{A.22}$$

$C_2(s_1,s_2)$ is found by interchanging the element notation in $B_2(s_1,s_2)$ because of the circuit duality. Notice that $\triangle A_2(s_1,s_2)$, $\triangle B_2(s_1,s_2)$ and $\triangle H(s_1,s_2)$ drop out if $Z_{12}(s_1+s_2)$ is zero. The second-order distortion then is canceled if g'_{m1} is matched to g'_{m2}.

Continue with the third-order Volterra analysis

$$g_{m1}A_3(s_1,s_2,s_3)+\frac{g''_{m1}}{6}A_1(s_1)A_1(s_2)A_1(s_3)$$
$$-g'_{m1}\overline{A_1(s_1)A_2(s_2,s_3)}$$
$$+\frac{A_3(s_1,s_2,s_3)-B_3(s_1,s_2,s_3)}{r_{o1}}$$
$$+g_{m2}A_3(s_1,s_2,s_3)+\frac{g''_{m2}}{6}A_1(s_1)A_1(s_2)A_1(s_3)$$
$$+g'_{m2}\overline{A_1(s_1)A_2(s_2,s_3)}$$
$$+\frac{A_3(s_1,s_2,s_3)-C_3(s_1,s_2,s_3)}{r_{o2}}$$
$$=-\frac{A_3(s_1,s_2,s_3)}{Z_s(s_1+s_2+s_3)} \tag{A.23}$$

$$g_{m1}A_3(s_1,s_2,s_3)+\frac{g''_{m1}}{6}A_1(s_1)A_1(s_2)A_1(s_3)$$
$$-g'_{m1}\overline{A_1(s_1)A_2(s_2,s_3)}$$
$$+\frac{A_3(s_1,s_2,s_3)-B_3(s_1,s_2,s_3)}{r_{o1}}$$
$$=\frac{B_3(s_1,s_2,s_3)}{Z_1(s_1+s_2+s_3)}$$
$$+\frac{B_3(s_1,s_2,s_3)-C_3(s_1,s_2,s_3)}{Z_{12}(s_1+s_2+s_3)} \tag{A.24}$$

$$g_{m2}A_3(s_1,s_2,s_3)+\frac{g''_{m2}}{6}A_1(s_1)A_1(s_2)A_1(s_3)$$
$$+g'_{m2}\overline{A_1(s_1)A_2(s_2,s_3)}$$
$$+\frac{A_3(s_1,s_2,s_3)-C_3(s_1,s_2,s_3)}{r_{o1}}$$
$$=\frac{C_3(s_1,s_2,s_3)}{Z_2(s_1+s_2+s_3)}$$
$$+\frac{C_3(s_1,s_2,s_3)-B_3(s_1,s_2,s_3)}{Z_{12}(s_1+s_2+s_3)} \tag{A.25}$$

$$A_2(s_1,s_2)=\frac{\frac{1}{2}(g'_{m1}-g'_{m2})(r_{o1}\,\|\,r_{o1})Z_s(s_1+s_2)A_1(s_1)A_1(s_2)+\triangle A_2(s_1,s_2)}{H(s_1+s_2)+\triangle H(s_1,s_2)} \tag{A.18}$$

$$B_2(s_1,s_2)=\frac{-\frac{Z_1(s_1+s_2)\,\|\,Z_2(s_1+s_2)}{Z_x(s_1+s_2)\,\|\,Z_s(s_1+s_2)}\left(\frac{1}{2}(g'_{m1}-g'_{m2})(r_{o1}\,\|\,r_{o1})Z_s(s_1+s_2)A_1(s_1)A_1(s_2)\right)+\triangle B_2(s_1,s_2)}{H(s_1+s_2)+\triangle H(s_1,s_2)} \tag{A.19}$$

$$A_3(s_1,s_2,s_3)=\frac{-Z_s(r_{o1}\,\|\,r_{o2})(-(g'_{m1}+g'_{m2})\overline{A_1(s_1)A_2(s_2,s_3)}+\frac{1}{6}(g''_{m1}+g''_{m2})A_1(s_1)A_1(s_2)A_1(s_3))}{H(s_1+s_2+s_3)}$$

$$B_3(s_1,s_2,s_3)=\frac{-Z_1(s_1+s_2+s_3)}{Z_x(s_1+s_2+s_3)\,\|\,Z_s(s_1+s_2+s_3)}A_3(s_1,s_2,s_3)$$

Assume again that $Z_{12}(s_1+s_2+s_3)$ is negligible then we arrive at the equation shown at the top of the page.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center, in particular Prof. E. Alon for helpful discussions in the noise analysis.

REFERENCES

[1] T. B. Zahariadis, "Migration toward 4G wireless communications," *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 6–7, Jun. 2004.

[2] M. Brandolini, P. Rossi, D. Manstretta, and F. Svelto, "Toward multi-standard mobile terminals - fully integrated receivers requirements and architectures," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 3, pt. 2, pp. 1026–1038, Mar. 2005.

[3] R. B. Staszewski, K. Muhammad, D. Leipold, C.-M. Hung, Y.-C. Ho, J. L. Wallberg, C. Fernando, K. Maggio, R. Staszewski, T. Jung, J. Koh, S. John, I. Y. Deng, V. Sarda, O. Moreira-Tamayo, V. Mayega, R. Katz, O. Friedman, O. E. Eliezer, E. de-Obaldia, and P. T. Balsara, "All-digital TX frequency synthesizer and discrete-time receiver for Bluetooth radio in 130-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2278–2291, Dec. 2004.

[4] K. Muhammad, Y. C. Ho, T. Mayhugh, C.-M. Hung, T. Jung, I. Elahi, C. Lin, I. Deng, C. Fernando, J. Wallberg, S. Vemulapalli, S. Larson, T. Murphy, D. Leipold, P. Cruise, J. Jaehnig, M.-C. Lee, R. B. Staszewski, R. Staszewski, and K. Maggio, "A discrete time quad-band GSM/GPRS receiver in a 90 nm digital CMOS process," in *Proc. IEEE Custom Integrated Circuits Conf.*, San Jose, CA, Sep. 2005, pp. 809–812.

[5] T. Chalvatzis, E. Gagnon, M. Repeta, and S. P. Voinigescu, "A low-noise 40-GS/s continuous-time bandpass delta sigma ADC centered at 2 GHz for direct sampling receivers," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, pp. 1065–1075, May 2007.

[6] A. Liscidini, M. Brandolini, D. Sanzogni, and R. Castello, "A 0.13 μm CMOS front-end for DCS1800/UMTS/802.11b-g with multi-band positive feedback low noise amplifier," in *2005 Symp. VLSI Circuits Dig.*, Kyoto, Japan, Jun. 2005, pp. 406–409.

[7] A. Bevilacqua and A. M. Niknejad, "An ultrawideband CMOS low-noise amplifier for 3.1–10.6-GHz wireless receivers," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2259–2268, Dec. 2004.

[8] S. Chehrazi, A. Mirzaei, R. Bagheri, and A. A. Abidi, "A 6.5 GHz wide-band CMOS low noise amplifier for multi-band use," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2005, pp. 801–804.

[9] C.-F. Liao and S. I. Liu, "A broadband noise-canceling MOS LNA for 3.1–10.6-GHz UWB receiver," *IEEE J. Solid-State Circuits*, vol. 42, no. 2, pp. 329–339, Feb. 2007.

[10] J.-H. C. Zhan and S. S. Taylor, "A 5 GHz resistive-feedback CMOS LNA for low-cost multi-standard applications," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2006, pp. 721–722.

[11] M. T. Reiha and J. R. Long, "A 1.2 V reactive-feedback 3.1–10.6 GHz low-noise amplifier in 0.13 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, pp. 1023–1033, May 2007.

[12] P. H. Woerlee, M. J. Knitel, R. van Langevelde, D. B. M. Klaassen, L. F. Tiemeijer, and A. J. Scholten, "RF-CMOS performance trends," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, Aug. 2001.

[13] K. Lee, I. Nam, I. Kwon, J. Gil, K. Han, S. Park, and B.-I. Seo, "The impact of semiconductor technology scaling on CMOS RF and digital circuits for wireless application," *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1415–1422, Jul. 2005.

[14] A. Pärssinen, "System design for multi-standard radios," in *IEEE ISSCC Girafe Forum*, Feb. 2006.

[15] W. Y. Ali-Ahmad, "RF system issues related to CDMA receiver specifications," *RF Design*, pp. 22–32, Sep. 1999.

[16] V. Aparin and L. E. Larson, "Modified derivative superposition method for linearizing FET low noise amplifiers," in *2004 IEEE Radio Frequency Integrated Circuits (RFIC) Systems Dig. Papers*, Fort Worth, TX, Jun. 2004, pp. 105–108.

[17] T. W. Kim, B. Kim, and K. Lee, "Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 223–229, Jan. 2004.

[18] F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Noise canceling in wideband CMOS LNAs," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 406–407.

[19] F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, *Wideband Low Noise Amplifiers Exploting Thermal Noise Cancellation*. Amsterdam, The Netherlands: Springer, 2005.

[20] I. Nam, B. Kim, and K. Lee, "CMOS RF amplifier and mixer circuits utilizing complementary characteristics of parallel combined NMOS and PMOS devices," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 5, pp. 1662–1671, May 2005.

[21] V. Aparin and L. E. Larson, "Linearization of monolithic LNAs using low-frequency low-impedance input termination," in *Proc. 29th Eur. Solid-State Circuits Conf. (ESSCIRC 2004)*, Estoril, Portugal, Sep. 2003, pp. 137–140.

[22] R. Gharpurey, "A broadband low-noise front-end amplifier for ultra wideband in 0.13-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1983–1986, Sep. 2005.

[23] S. C. Blaakmeer, E. A. M. Klumperink, B. Nauta, and D. M. W. Leenaerts, "An inductorless wideband balun-LNA in 65nm CMOS with balanced output," in *Proc. 33rd Eur. Solid-State Circuits Conf. (ESSCIRC 2007)*, Munich, Germany, Sep. 2007, pp. 364–367.

Wei-Hung Chen received the B.S. and M.S. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1998 and 2000, respectively. Since 2003, he has been working toward the Ph.D. degree in electrical engineering and computer science at the University of California, Berkeley.

His research interest is in analog/RF circuits design for communications. During the summers of 2005 and 2006, he interned in the wireless group at Maxim Integrated Products, Sunnyvale, CA, where he worked on WiMAX transceiver chip designs.

Mr. Chen received the Outstanding Student Design Award from Analog Devices Inc. and the IEEE MTT-S Graduate Student Fellowship, both in 2007.

Gang Liu (M'07) received the B.E. degree from Tsinghua University, China, in 1998, and the Ph.D. degree from the University of California at Berkeley in 2006.

He is now with Marvell Semiconductor Inc., where he is involved with wireless system development. His main research interests include analog/digital circuits, and integrated circuits for systems.

Boos Zdravko received the B.S.E.E. degree from the University of Zagreb, Croatia, and the M.S.E.E. degree from the Eindhoven International Institute, Eindhoven, The Netherlands, in 1982 and 1991, respectively.

From 1983 to 1989, he was with the Institute for Electronics in Zagreb where he was involved in the UHF frequency synthesisers development. From 1991 to 1998, he was with Philips Consumer Electronics in Eindhoven where he developed DAB front-ends for DAB 452 and DAB 752 receivers. In 1989, he joined Siemens Semiconductors/Infineon Technologies where he was developing RF concepts for the advanced multimode transceivers. He is currently a Senior Principal Wireless Communications RF Engineer. His technical interests are in the digital transmitter and receiver concepts in deep-submicron CMOS technology. He holds several EU and U.S. patents.

Ali M. Niknejad (M'00) was born in Tehran, Iran. He received the B.S.E.E. degree from the University of California, Los Angeles, in 1994, and the Master's and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1997 and 2000.

During his graduate studies, he authored *ASITIC*, a CAD tools that aids in the simulation and design of passive circuit elements such as inductors into silicon integrated circuits. *ASITIC* is actively used by industry and academic research and development centers. After graduation from Berkeley, he worked in industry focusing on the design and research of analog RF integrated circuits and devices for wireless communication applications. He is currently an Associate Professor in the EECS Department at UC Berkeley and co-director of the Berkeley Wireless Research Center and the BSIM Research Group. His research interests lie within the area of wireless and broadband communications, including the implementation of integrated communication systems in silicon using CMOS, SiGe, and BiCMOS processes. Focus areas of his research include analog and RF circuits, device physics and modeling, and numerical techniques in electromagnetics, with an emphasis on the analysis and modeling of active and passive devices at microwave frequencies for IC applications.

Prof. Niknejad has served as an associate editor of the IEEE JSSC and is now serving on the TPC for CICC and ISSCC. He was co-recipient of the Outstanding Technology Directions Paper at ISSCC 2004 for co-developing a modeling approach for devices up to 65 GHz.