Analysis and Demonstration of an IIP3 Improvement Technique for Low-Power RF Low-Noise Amplifiers

Chun-Hsiang Chang and Marvin Onabajo, *Senior Member, IEEE*

***Abstract*— This paper describes a linearization method to enhance the third-order distortion performance of a subthreshold common-source cascode low-noise amplifier (LNA) without extra power consumption by using passive components. An inductor between the gate of the cascode transistor and the power supply in combination with a digitally programmable capacitor between the gate and the drain of the cascode transistor enable to improve the third-order intermodulation intercept point (IIP3) of a subthreshold LNA. The theoretical mechanisms that underlie the linearity improvement are analyzed comprehensively under the consideration of the LNA's input stage, cascode stage, reverse isolation, and stability. A 1.8-GHz LNA was designed and fabricated using 0.11-μm CMOS technology to prove the concept. Measurement results reveal that the linearized low-power LNA has a 14.8-dB voltage gain, a 3.7-dB noise figure, and a −3.7-dBm IIP3 with a power consumption of 0.336 mW.**

***Index Terms*— Low-noise amplifier, subthreshold biasing, weak inversion, third-order intermodulation intercept point (IIP3) improvement, low-power radio frequency (RF) front-end design, digitally programmable tuning.**

I. Introduction

REQUIREMENTS for portable electronic devices with low-power radio frequency (RF) circuits are based on the needs to extend battery lifetimes [1], [2] or to operate with harvested energy [3], [4]. The low-noise amplifier (LNA) is a critical block in RF receiver front-ends because its specifications strongly impact the system-level performance of the complete receiver, including the overall noise and linearity. Transistors operated in the subthreshold (or weak inversion) region offer opportunities to minimize power consumption of CMOS RF front-end circuits. Over the past years, some of such LNAs and mixers were reported with very low power consumptions [5]–[14], which were made possible through high transconductance-to-drain current ratios (g_m/I_D) and low power supply voltages (V_{DD}). However, the prevalent design challenge associated with subthreshold RF front-end circuits has been linearity degradation. For example, in earlier published subthreshold LNAs and mixers [5]–[13], the third-order intermodulation intercept point (IIP3) is typically equal to or below −10 dBm.

The most prevalent IIP3 improvement methods for low-power LNAs in narrowband applications can be broadly divided into two categories. The first approach is to use an auxiliary transistor biased in the weak inversion region to cancel the third-order nonlinearity coefficient (g_3), but the main transistor has to be operated in strong inversion with higher linear transconductance ($g_1 = g_m$) than that in the auxiliary path [15], [16]. The second method is to operate the main transistor between the moderate inversion and subthreshold regions for finding the optimum bias zone [17]. However, linearization methods have not yet been reported with measurements of RF amplifiers using only transistors biased in the subthreshold region.

In this paper, measurement results are described together with further analysis for a subthreshold LNA linearization technique that only uses passive devices for the third-order nonlinearity coefficient reduction without additional power consumption. Furthermore, a digitally programmable IIP3 tuning topology is employed for application in RF front-end calibration methods [18], [19]. The measurement results were first presented in [20], and this extended article provides new insights from the detailed analysis of the LNA's input stage and cascode stage, including impacts of gate-to-bulk capacitances. Reverse isolation and stability considerations are also discussed based on newly derived analytical expressions and plots. Additional simulations reveal how the linearization method affects voltage gains associated with internal nodes of the LNA. The paper is organized as follows. Section II briefly summarizes key subthreshold RF circuit design aspects. The analyses of linearity, gain, noise, and input matching conditions for the linearized subthreshold LNA under investigation are presented in Section III. Chip measurement results are summarized in Section IV, and conclusions are made in Section V.

II. Subthreshold Design Considerations

The key distinguishing characteristics of subthreshold biasing compared to strong inversion biasing are stated below to summarize our prior simulation-based works [21]–[24].

Manuscript received May 24, 2017; revised October 25, 2017; accepted December 4, 2017. This work was supported by the National Science Foundation under Award 1451213. This paper was recommended by Associate Editor F. Lustenberger. *(Corresponding author: Marvin Onabajo.)*

C.-H. Chang was with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA. He is now with OmniVision Technologies, Santa Clara, CA 95054 USA (e-mail: chunhsiang.chang@ovt.com).

M. Onabajo is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: monabajo@ece.neu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2017.2781369

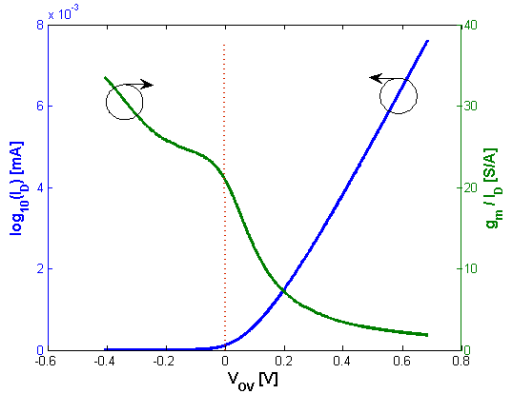

Fig. 1. Drain current (I_D) with logarithmic scale and current efficiency (g_m/I_D) vs. overdrive voltage ($V_{OV} = V_{GS} - V_{TH}$) of an NMOS transistor with 130nm channel length.

1) Higher power efficiency: transistors biased in subthreshold can provide a higher g_m/I_D ratio than when biased in strong inversion. As can be seen in Fig. 1, power-efficient subthreshold biasing involves the use of gate-to-source (V_{GS}) bias voltages below the threshold voltage (V_{TH}), where the overdrive voltage on the x-axis is $V_{OV} = V_{GS} - V_{TH}$. Hence, the drain-to-source voltage (V_{DS}) can be low with subthreshold biasing, which permits the use of reduced power supply voltages at the expense of slightly higher noise figure.
2) Increase of parasitic capacitances and relative changes of their values: In subthreshold mode of operation, the gate-to-source capacitance (C_{gs}) no longer dominates, implying that the gate-to-drain capacitance (C_{gd}) and the gate-to-bulk capacitance (C_{gb}) have to be taken into account for more sophisticated design. As visualized in Fig. 2(a), the simulated C_{gs}/C_{gg} ratio (where: $C_{gg} = C_{gs} + C_{gd} + C_{gb}$) decreases approximately 20% when g_m/I_D is swept such that the bias point changes from the strong inversion region to the subthreshold region. As a result, C_{gs} no longer dominates in the subthreshold region. Hence, the gate-drain capacitance C_{gd} and the gate-bulk capacitance C_{gb} should be taken into account for precise input impedance matching calculation and linearity estimation. Moreover, to achieve similar transconductance gains as in strong inversion it is required to increase the transistor widths, which results in higher parasitic capacitances and lower transition frequency (f_T). Fig. 2(b) shows that the transition frequency changes from 70 GHz to a few GHz when the transistor bias is varied from the strong inversion to the subthreshold region. In the past, transistors biased in subthreshold region were not seriously considered for RF circuit design because f_T was severely limited. However, newer CMOS process technologies have significantly improved f_T values, which has made it possible to design subthreshold circuits with operating frequencies up to several gigahertz using 130nm CMOS and smaller technology nodes.
3) Linearity degradation due to highly positive g_3/g_1 ratio, where $g_1 = g_m$ and g_3 is the third-order nonlinearity coefficient: For a weakly nonlinear transistor, the relationship of the small-signal gate-source voltage (v_{gs}) and the drain current (i_d) can be expressed by the first three power series terms [25], [26]:

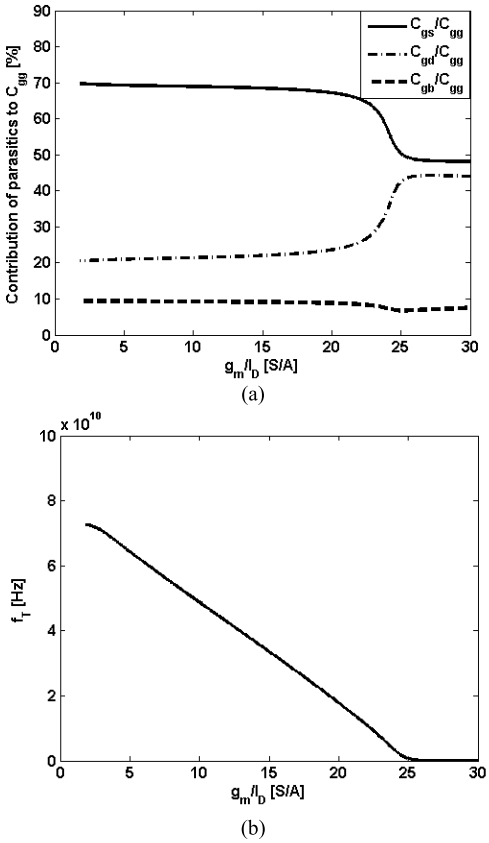

Fig. 2. (a) Contribution of parasitic capacitances to the total gate capacitance (C_{gg}) vs. g_m/I_D; (b) transition frequency (f_T) vs. g_m/I_D of an NMOS transistor with 130nm channel length.

$$i_d = g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3, \tag{1}$$

where g_1, g_2 and g_3 are the linear transconductance gain, second-order nonlinearity coefficient and third-order nonlinearity coefficient of the transistor, respectively. Notice that these parameters can be obtained by taking derivatives of the DC drain current (I_D) with respect to the DC gate-source voltage (V_{GS}) at the bias point:

$$g_1 = \frac{\partial I_D}{\partial V_{GS}}, \quad g_2 = \frac{1}{2!}\frac{\partial^2 I_D}{\partial V_{GS}^2}, \quad g_3 = \frac{1}{3!}\frac{\partial^3 I_D}{\partial V_{GS}^3}. \tag{2}$$

The sign of g_3 changes from negative to positive when the transistor biasing is changed from strong inversion to subthreshold [21]. In addition, the value of g_3/g_1 strongly depends on the g_m/I_D ratio when biasing transistors in the subthreshold region. This has created the need to develop adapted or new linearization methods compared to existing schemes for RF circuits with transistors biased in the strong inversion region.

III. Linearized Subthreshold LNA

Fig. 3 displays the schematic of the LNA under investigation, where inductor L_{g2} and digitally-programmable capacitor C_{gd2_ext} can improve the IIP3 in the presence of variations.

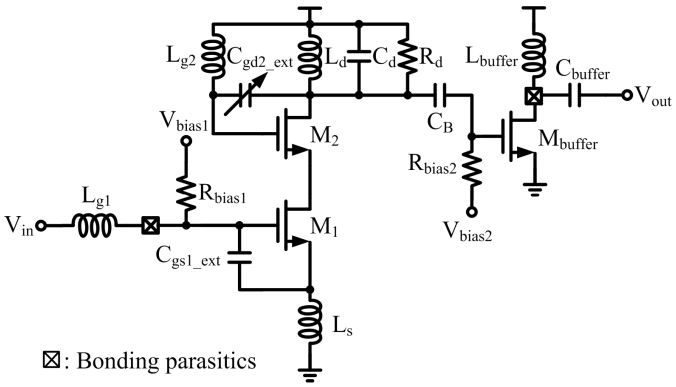

Fig. 3. Linearized subthreshold LNA, where the voltage at the source of M_2 is referred to as V_y.

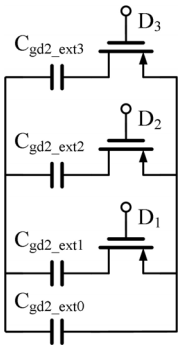

Fig. 4. Digitally-programmable capacitance (C_{gd2_ext}).

Inductor L_{g1}, inductor L_{buffer}, and capacitor C_{buffer} are off-chip components for impedance matching purposes. C_{gd2_ext} is implemented with a fixed metal-insulator-metal (MIM) capacitor (C_{gd2_ext0}) and a 3-bit digitally-programmable MIM capacitor (C_{gd2_ext1}, C_{gd2_ext2}, and C_{gd2_ext3}) as illustrated in Fig. 4. Preliminary simulations showed that MOS capacitors can also be employed to realize C_{gd2_ext}, but resulting in slightly increased LNA gain variation and less linearity improvement due to inherent nonlinearity and voltage-dependent variation of MOS capacitances compared to metal-insulator-metal capacitors. All passive devices (with frequency-dependent quality factor limitations) and active devices were simulated using foundry-supplied models. In this section, the bonding/package parasitics and buffer stage were neglected to simplify the small-signal analysis. It has been shown in [27] that an inductor between the gate of the cascode transistor and the power supply can improve stability of a common-source cascode LNA by creating a sharp notch in the transfer function of the reverse isolation (S12) around the operating frequency. In another related work [28], a fully differential common-source LNA topology with an inductor at the gate of the cascode transistor, combined with a cross-coupling capacitor between the gate of the cascode transistor and the source of the opposite cascode transistor, was introduced to decrease the noise figure, improve the linearity, and enhance the voltage gain. Nevertheless, this LNA was biased in the strong inversion region. The linearization method described in the next subsection was developed for subthreshold common-source cascode LNAs and does not require cross-coupling for nonlinearity cancellation.

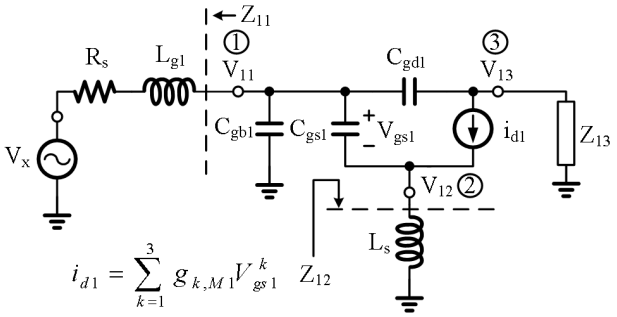

Fig. 5. Small-signal model of the LNA's input stage with M_1 under consideration of its nonlinear drain-to-source current.

A. Linearity Analysis of the Input Stage

The input stage (transistor M_1) and cascode stage (transistor M_2) of the LNA in Fig. 3 are split into two individual parts to simplify the linearity analysis. Fig. 5 shows the small-signal model of the input stage where the extra metal-insulator-metal capacitor ($C_{gs1_{ext}}$) is lumped into the parasitic capacitance C_{gs1}. The IIP3 of transistor M_1 can be derived after Volterra series analysis [15], [16], [28] as

$$IIP_{3,M1} = \frac{1}{6R_s \cdot |H_1(\omega)| \cdot |A_{11}(\omega)|^3 \cdot |\varepsilon_{M1}(\Delta\omega, 2\omega)|}, \quad (3)$$

where ω is the center frequency of the two intermodulation tones at ω_{RF1} and ω_{RF2}, $\Delta\omega$ is defined as $|\omega_{RF1} - \omega_{RF2}|$, and R_s is the antenna impedance of 50Ω. $H_1(\omega)$ is the third-order nonlinearity transfer function from V_{in} to the drain-source current (i_{d1}) of M_1, $A_{11}(\omega)$ is the linear transfer function from the input voltage (V_x) to the gate-to-source voltage (V_{gs1}), and $\varepsilon_{M1}(\Delta\omega, 2\omega)$ represents the nonlinear contribution from the second-order and third-order terms of transistor M_1. Minimization of the term $|\varepsilon_{M1}(\Delta\omega, 2\omega)|$ in (3) leads to improved IIP_3. For this reason, we will now focus on the analysis of the $\varepsilon(\Delta\omega, 2\omega)$ term for transistors M_1 and M_2. The $\varepsilon_{M1}(\Delta\omega, 2\omega)$ term of M_1 can be expressed as

$$\varepsilon_{M1}(\Delta\omega, 2\omega) = g_{3,M1} - g_{oB,M1}, \quad (4)$$

where (5)–(12) define the parameters, as shown at the bottom of the next page.

The variables $g_{1,M1}$, $g_{2,M1}$ and $g_{3,M1}$ are the linear transconductance gain, second-order nonlinearity coefficient and third-order nonlinearity coefficient of transistor M_1, respectively. The terms $g_{M1}(\Delta\omega)$ and $g_{M1}(2\omega)$ in (5) can be evaluated by substituting $\Delta\omega = |\omega_{RF1} - \omega_{RF2}|$ and $2\omega = (\omega_{RF1} + \omega_{RF2})$ for ω in (6) respectively. Note that contrary to [21], the parasitic capacitance C_{gb1} was included above to further improve the accuracy of the analysis and the insights gained from it, for which more detailed derivations are provided here. Conceptually, Fig. 6 visualizes that the mechanism of the partial third-order intermodulation cancellation in (4) entails changing the magnitude and phase of $g_{oB,M1}$ in (5) such that they are almost identical to those of $g_{3,M1}$, where $g_{oB2,M1}$ represents a better design point (with more cancellation of $g_{3,M1}$) than $g_{oB1,M1}$ as result of different parameters.

To continue the analysis, the following equations can be written after applying Kirchhoff's current law to nodes 1, 2,

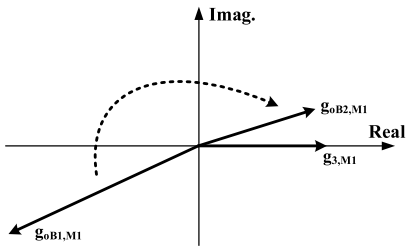

Fig. 6. Vector diagram of the third-order intermodulation cancellation, where $g_{oB1,M1}$ and $g_{oB2,M1}$ are $g_{oB,M1}$ realizations in (4) with different design parameters.

and 3 in Fig. 5:

$$\frac{V_x}{Z_{11}} - \left(j\omega C_{gs1} + j\omega C_{gd1} + j\omega C_{gb1} + \frac{1}{Z_{11}}\right) V_{11}, + j\omega C_{gs1} V_{12} + j\omega C_{gd1} V_{13} = 0 \quad (13)$$

$$j\omega C_{gs1} V_{11} - \left(j\omega C_{gs1} + \frac{1}{Z_{12}}\right) V_{12} + i_{d1} = 0, \quad (14)$$

$$j\omega C_{gd1} V_{11} - \left(j\omega C_{gd1} + \frac{1}{Z_{13}}\right) V_{13} - i_{d1} = 0. \quad (15)$$

Furthermore,

$$V_{gs1} = V_{11} - V_{12}. \quad (16)$$

Using (13) through (16) and the definitions of $g_{M1}(\omega)$ in (6), $Z(\omega)$ in (7), and $Z_{13}(\omega)$ in (10), $V_{gs1}(\omega)$ can be derived as the following function of V_x and i_{d1}:

$$V_{gs1}(\omega) = \frac{1}{g_{M1}(\omega)} \left[\frac{\left(1 + j\omega C_{gd1} Z_{13}(\omega)\right) V_x}{Z(\omega)} - i_{d1}\right]. \quad (17)$$

The relation between V_x and V_{13}, where V_x and V_{13} are the input and output voltages of the input stage with transistor M_1, can be expressed with a Volterra series as

$$V_{13}(\omega) = C_{11}(\omega) \circ V_x + C_{12}(\omega_1, \omega_2) \circ V_x + C_{13}(\omega_1, \omega_2, \omega_3) \circ V_x \ldots. \quad (18)$$

On the other hand, the relationship between the drain current (i_{d1}) and the gate-to-source voltage (V_{gs1}) of transistor M_1 can be written in terms of its linear transconductance ($g_{1,M1}$) and its nonlinear transconductance components ($g_{2,M1}$, $g_{3,M1}$, ...):

$$i_{d1}\left(V_{gs1}\right) = g_{1,M1} V_{gs1} + g_{2,M1} V_{gs1}^2 + g_{3,M1} V_{gs1}^3 + \ldots. \quad (19)$$

Furthermore, the relation between V_x and V_{gs1} in Fig. 5 can also be expressed with a Volterra series as

$$V_{gs1}(\omega) = A_{11}(\omega) \circ V_x + A_{12}(\omega_1, \omega_2) \circ V_x + A_{13}(\omega_1, \omega_2, \omega_3) \circ V_x \ldots. \quad (20)$$

From [15], [16], and [28] , the linear transfer functions $A_{11}(\omega)$ and $C_{11}(\omega)$ above can be determined by applying a single tone [$V_x(\omega) = e^{j\omega t}$] in the analysis, which results in:

$$A_{11}(\omega) = \frac{1}{g_{1,M1} + g_{M1}(\omega)} \left[\frac{1 + j\omega C_{gd1} Z_{13}(\omega)}{Z(\omega)}\right], \quad (21)$$

$$C_{11}(\omega) = Z_{13}(\omega) \frac{j\omega C_{gd1} d(\omega) - [d(\omega) + e(\omega)]\, g_{1,M1} A_{11}(\omega)}{b(\omega) + c(\omega) + j\omega\, C_{gs1} Z(\omega)}; \quad (22)$$

where:

$$b(\omega) = 1 + j\omega C_{gd1} Z_{13}(\omega), \quad (23)$$

$$c(\omega) = [j\omega C_{gs1} + j\omega C_{gd1} + j\omega C_{gb1} - \omega^2 C_{gd1} C_{gb1} Z_{13}(\omega)] \cdot Z_{11}(\omega), \quad (24)$$

$$d(\omega) = 1 + j\omega C_{gs1} Z_{12}(\omega), \quad (25)$$

$$e(\omega) = [j\omega C_{gs1} + j\omega C_{gd1} + j\omega C_{gb1} - \omega^2 C_{gs1} C_{gb1} Z_{12}(\omega)] \cdot Z_{11}(\omega). \quad (26)$$

B. Linearity Analysis of the Cascode Stage

Fig. 7 depicts the small-signal model of the cascode stage, where the extra MIM capacitor C_{gd2_ext} is merged with the parasitic capacitance C_{gd2}. A cascode device whose gate is connected to an AC ground (i.e., V_{DD}) typically only has a small impact on the overall linearity of a cascode

$$g_{oB,M1} = \frac{2}{3} g_{2,M1}^2 \left[\frac{2}{g_{1,M1} + g_{M1}(\Delta\omega)} + \frac{1}{g_{1,M1} + g_{M1}(2\omega)}\right], \quad (5)$$

$$g_{M1}(\omega) = \frac{1 + j\omega C_{gd1} \cdot [Z_{11}(\omega) + Z_{13}(\omega)] + j\omega C_{gs1} \cdot [Z_{11}(\omega) + Z_{12}(\omega)] + j\omega C_{gb1} \cdot [1 + j\omega C_{gd1} Z_{13}(\omega)] \cdot Z_{11}(\omega)}{Z(\omega)}, \quad (6)$$

$$Z(\omega) = Z_{12}(\omega) + j\omega C_{gb1}[1 + j\omega C_{gd1} Z_{13}(\omega)] Z_{11}(\omega)\, Z_{12}(\omega) + j\omega C_{gd1} [Z_{11}(\omega)\, Z_{12}(\omega) + Z_{11}(\omega)\, Z_{13}(\omega) + Z_{12}(\omega)\, Z_{13}(\omega)], \quad (7)$$

$$Z_{11}(\omega) = R_s + j\omega L_{g1}, \quad (8)$$

$$Z_{12}(\omega) = j\omega L_s, \quad (9)$$

$$Z_{13}(\omega) = \frac{1 + j\omega C_{gd2} Z_{23}(\omega) + [j\omega C_{gs2} + j\omega C_{gd2} - \omega^2 C_{gs2} C_{gd2} Z_{23}(\omega)] \cdot Z_{22}(\omega)}{g_{1,M2} + j\omega C_{gs2} + [j\omega C_{gd2} g_{1,M2} - \omega^2 C_{gd2} C_{gs2}] \cdot [Z_{22}(\omega) + Z_{23}(\omega)]}, \quad (10)$$

$$Z_{22}(\omega) = \left(j\omega L_{g2}\right) // \left(j\omega C_{gb2}\right)^{-1}, \quad (11)$$

$$Z_{23}(\omega) = R_d // (j\omega L_d) // (j\omega C_d)^{-1}. \quad (12)$$

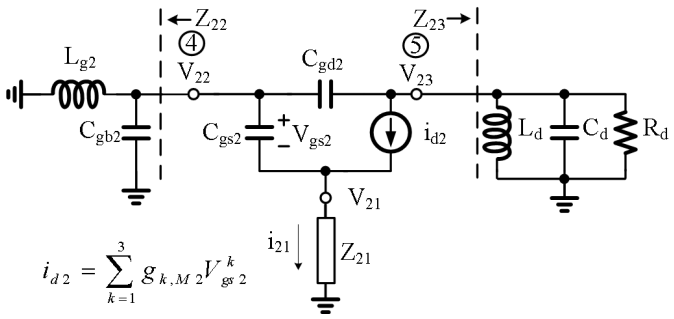

Fig. 7. Small-signal model of the LNA's cascode stage with M_2 under consideration of its nonlinear drain-to-source current.

common-source LNA. On the other hand, the cascode stage with additional components at the gate of M_2 in Fig. 3 has a significant impact on the overall linearity performance. Subthreshold RF designs typically employ wide transistors to achieve sufficiently high transconductances. Hence, increasing the width/length ratio of M_2 is not a feasible option to reduce its impact on linearity because the adverse effects of the parasitic capacitances on gain and reverse isolation would become worse. However, nonlinearity cancellation in the cascode stage is realized with the proposed design technique to improve third-order linearity. The linearity effect of the cascode device was only briefly analyzed in [20], and the following analysis in this section provides further insights into internal transfer functions that affect the LNA's voltage gain (Section III-C). By applying Volterra series [28] analysis, the upper A_{IIP3} limit (i.e., max. IIP3 point in volts) due to transistor M_2 can be derived as

$$A_{\mathrm{IIP3,M2}}^{2} = \frac{4}{3} \cdot \frac{1}{|H_2(\omega)| \cdot |A_{21}(\omega)|^3 \, |\varepsilon_{M2}(\Delta\omega, 2\omega)|}. \quad (27)$$

The definition of $\varepsilon_{M2}(\Delta\omega, 2\omega)$ is the same as in (4) and can be rewritten as

$$\varepsilon_{M2}(\Delta\omega, 2\omega) = g_{3,M2} - g_{oB,M2}, \quad (28)$$

where:

$$g_{oB,M2} = \frac{2}{3} g_{2,M2}^{2} \left[\frac{2}{g_{1,M2} + g_{M2}(\Delta\omega)} + \frac{1}{g_{1,M2} + g_{M2}(2\omega)} \right], \quad (29)$$

$$g_{M2}(\omega) = \frac{\begin{array}{l} 1 + j\omega C_{gd2} Z_{23}(\omega) \\ + \left(j\omega C_{gs2} + j\omega C_{gd2}\right) \cdot \left[1 + j\omega C_{gd2} Z_{23}(\omega)\right] \\ \cdot Z_{22}(\omega) \\ + \omega^2 C_{gs2} C_{gd2} Z_{22}(\omega) \, Z_{23}(\omega) \end{array}}{j\omega C_{gd2} Z_{22}(\omega) \, Z_{23}(\omega)}. \quad (30)$$

Parameters $g_{1,M2}$, $g_{2,M2}$ and $g_{3,M2}$ are the linear gain, second-order nonlinearity coefficient and third-order nonlinearity coefficient of transistor M_2. The terms $g_{M2}(\Delta\omega)$ and $g_{M2}(2\omega)$ in (29) can be evaluated by substituting $\Delta\omega = |\omega_{RF1} - \omega_{RF2}|$ and $2\omega = (\omega_{RF1} + \omega_{RF2})$ for ω in (30) respectively. The linear transfer function $A_{21}(\omega)$ in equation (27) is derived next. By applying Kirchhoff's current law to nodes 4 and 5 in Fig. 7, the following equations can be obtained:

$$j\omega C_{gs2}(V_{22} - V_{21}) + \frac{V_{22}}{Z_{22}} + j\omega C_{gd2}(V_{22} - V_{23}) = 0, \quad (31)$$

$$j\omega C_{gd2}(V_{23} - V_{22}) + \frac{V_{23}}{Z_{23}} + i_{d2} = 0. \quad (32)$$

It can be noted that

$$V_{gs2} = V_{22} - V_{21}. \quad (33)$$

From (31) through (33) and the previous $g_{M2}(\omega)$, $Z_{22}(\omega)$ and $Z_{23}(\omega)$ definitions, $V_{gs2}(\omega)$ can be found in terms of V_{21} and i_{d2} as follows:

$$V_{gs2}(\omega) = \frac{1}{g_{M2}(\omega)} \left[\frac{f(\omega) V_{21}}{j\omega C_{gd2} Z_{22}(\omega) \, Z_{23}(\omega)} - i_{d2} \right], \quad (34)$$

where:

$$f(\omega) = \left(1 + j\omega C_{gd2} Z_{23}(\omega)\right) \cdot \left[1 + j\omega C_{gd2} Z_{22}(\omega)\right] + \omega^2 C_{gd12} Z_{22}(\omega) \, Z_{23}(\omega). \quad (35)$$

The relationship between V_{21} and V_{23} in Fig. 7, where V_{21} and V_{23} are the input and output voltages of transistor M_2, can be written with Volterra series:

$$V_{23}(\omega) = C_{21}(\omega) \circ V_{21} + C_{22}(\omega_1, \omega_2) \circ V_{21} + C_{23}(\omega_1, \omega_2, \omega_3) \circ V_{21} \ldots. \quad (36)$$

The relation between the drain current (i_{d2}) and the gate-to-source voltage (V_{gs2}) of transistor M_2 is

$$i_{d2}\left(V_{gs2}\right) = g_{1,M2} V_{gs2} + g_{2,M2} V_{gs2}^{2} + g_{3,M2} V_{gs2}^{3} + \ldots. \quad (37)$$

Furthermore, the relationship between V_{21} and V_{gs2} can be expressed by applying Volterra series as

$$V_{gs2}(\omega) = A_{21}(\omega) \circ V_x + A_{22}(\omega_1, \omega_2) \circ V_x + A_{23}(\omega_1, \omega_2, \omega_3) \circ V_x \ldots. \quad (38)$$

Correspondingly, the linear transfer functions $A_{21}(\omega)$ and $C_{21}(\omega)$ can be determined through single-tone analysis [using $V_{21}(\omega) = e^{j\omega t}$], which are:

$$A_{21}(\omega) = \frac{1}{g_{1,M2} + g_{M2}(\omega)} \left[\frac{f(\omega)}{j\omega C_{gd2} Z_{22}(\omega) \, Z_{23}(\omega)} \right], \quad (39)$$

$$C_{21}(\omega) = Z_{23}(\omega) \left[\begin{array}{l} g_{1,M2} A_{21}(\omega) \, \dfrac{1 + \left(j\omega C_{gs2} + j\omega C_{gd2}\right)}{f(\omega)} \\ - \dfrac{\omega^2 C_{gs2} C_{gd2} Z_{22}(\omega)}{f(\omega)} \end{array} \right]. \quad (40)$$

Fig. 8 visualizes the numerical calculations of $|\varepsilon_{M2}(\Delta\omega, 2\omega)|$ in equation (28) versus L_{g2} for three values of C_{gd2_ext} based on the above equations. In this design example, an L_{g2} value around 3.5 nH leads to optimum IIP3. In addition to the cancellations associated with equations (28) and (4) for the cascode stage and input stage respectively, the effectiveness of the linearization is impacted by higher-order nonlinearities and interactions between the stages. Depending on a particular LNA's design parameters, its center frequency, and the process technology; either transistor M_1 or M_2 can be the linearity bottleneck with subthreshold biasing.

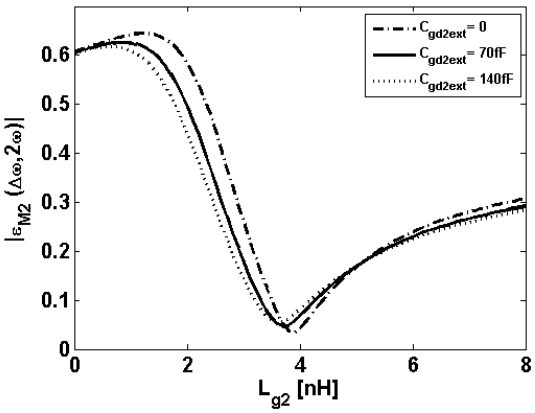

Fig. 8. Calculation results of $|\varepsilon(\Delta\omega, 2\omega)|$ in equation (28) for L_{g2} with three C_{gd2_ext} combinations in the cascode stage (with M_2).

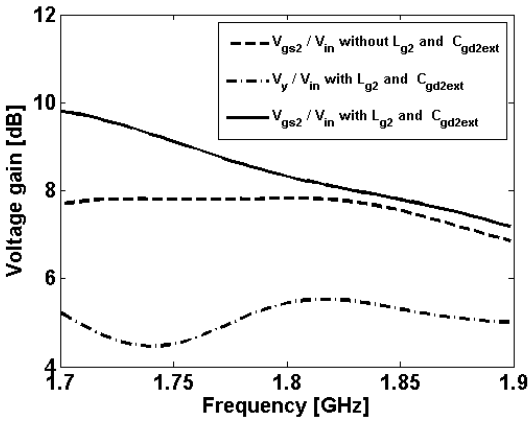

Fig. 9. Simulated voltage gains from V_{in} to V_y (at the source of M_2) and V_{gs2} of the LNA with and without L_{g2} and C_{gd2_ext} (ideal components).

In this design, M_2 imposes more severe linearity degradation. In general, numerical evaluations of equations (4) and (28) for a few C_{gd2_ext} values can serve as a helpful first design step, as it allows the selection of the L_{g2} range for an initial design by observing the inductor values for optimum M_1 and M_2 linearity. While the above equations provide a theoretical foundation for the proposed linearization technique to identify tradeoffs based on key parameters, in practice a designer can select a reasonable C_{gd2_ext} value and sweep L_{g2} in post-layout circuit simulations with accurate device models and extracted parasitics. A standard IIP3 metric can be monitored during the simulations in lieu of the $|\varepsilon_{M2}(\Delta\omega, 2\omega)|$ term. In our previous research [21], two LNA designs (with and without this linearized technique) were compared based on simulations. The LNAs were designed with the same power consumption, and have the same voltage gain in the typical corner case. The comparison revealed that the conventional cascode LNA has a slightly better noise performance (0.4 dB) than the linearized LNA. However, the linearized LNA has a significantly higher IIP3 (e.g., 11.2 dB better in [21]) with negligible difference in the other performance parameters. The related reverse isolation (S12) and stability aspects for the selection of L_{g2} and C_{gd2_ext} values are discussed in Section III-F.

C. Voltage Gain

The voltage gain of the linearized LNA can be separated to identify the contributions associated with transistors M_1 and M_2. In sections III-A and III-B, the linear transfer functions from V_x to V_{13} (in Fig. 5) and from V_{21} to V_{23} (in Fig. 7) were derived, which represent the frequency-dependent voltage gains $C_{11}(\omega)$ and $C_{21}(\omega)$ of the two stages. From equations (22) and (40), these voltage gains can be combined to determine the overall LNA gain:

$$Av\left(\omega\right) = \left|C_{11}\left(\omega\right)\right| \times \left|C_{21}\left(\omega\right)\right|. \tag{41}$$

In addition to the nonlinearity cancellation analyzed above, a secondary mechanism leads to linearity enhancement thanks to the extra components at the gate of M_2. Fig. 9 displays the simulated voltage gains from V_{in} to V_y (at the source of M_2) and V_{gs2} (in Fig. 3). For the LNA without L_{g2}, the voltage gain from V_{in} to V_y is same as that from V_{in} to V_{gs2} but with opposite phase. The LNA with $L_{g2} = 3.5$ nH and $C_{gd2\_ext} = 150$ fF has a lower voltage gain V_y/V_{in} than the conventional cascode common-source LNA, but both LNAs have a similar V_{gs2}/V_{in} gain for frequencies around and above the 1.8 GHz operating frequency. Hence, the attenuation at V_y and reduced signal swing at this node (due to L_{g2} and C_{gd2_ext}) contributes to the linearity improvement.

D. Input Matching Network

The input matching of a subthreshold common-source LNA was analyzed in [23] without inductor L_{g2}. For the modified LNA (with linearization) presented in this paper, it can be shown that the input impedance under consideration of the extra components can be estimated as

$$Z_{in}\left(\omega\right) = j\omega L_{g1} + Z_{in}^{*}\left(\omega\right) // \frac{1}{j\omega C_{MF}}; \tag{42}$$

where

$$Z_{in}^{*}\left(\omega\right) = \frac{1}{j\omega C_{gs1}} + j\omega L_s + \frac{g_{1,M1} L_s}{C_{gs1}}, \tag{43}$$

$$C_{MF} = \left(1 + A\left(\omega\right)\right) \cdot C_{gd1} + C_{gb1}, \tag{44}$$

$$A\left(\omega\right) = G_{1,M1-eff}\left(\omega\right) \cdot Z_{13}\left(\omega\right), \tag{45}$$

$$G_{1,M1-eff}\left(\omega\right) = \frac{g_{1,M1}}{1 + j\omega L_s (g_{1,M1} + j\omega C_{gs1})} - j\omega C_{gd1}, \tag{46}$$

and $Z_{13}(\omega)$ is defined in equation (10).

E. Noise

The noise factor analysis for the subthreshold common-source LNA with inductive source degeneration has been reported in [29] with the following result:

$$F = 1 + C_t^2 \times \frac{\omega_o^2 R_s \gamma n^2 V_T}{I_D} \left[\frac{\delta\alpha^2}{5\gamma} \left(1 + Q_{in}^2\right) \frac{C_{gs1}^2}{C_t^2} + 1 - 2\left|c\right| \frac{C_{gs1}}{C_t} \sqrt{\frac{\delta\alpha^2}{5\gamma}} \right], \tag{47}$$

where $C_t = C_{gs1} + C_{gs1\_ext}$, ω_0 is the operating frequency, γ and δ are the channel and gate noise coefficients, $\alpha = g_{1,M1}/g_{d0,M1}$, $g_{d0,M1}$ is the channel conductance with zero drain-source voltage, V_T is the thermal voltage, Q_{in} is the quality factor of the input matching network, and c is the correlation parameter between the gate and channel noise currents.

F. Reverse Isolation and Stability

Compared to conventional common-source cascode LNAs, the described linearization method requires an inductor at the gate of the cascode transistor. As shown in [27], the reverse isolation of such an LNA can be improved in the desired frequency band with proper sizing of the inductor at the gate of the cascode transistor. To analytically estimate the impact on reverse isolation, the transfer function from V_{out_LNA} to V_y (at the source of M_2) in Fig. 3 can be derived from the small-signal circuit in Fig. 10(a):

$$H(s) = \frac{V_y}{V_{out\_LNA}} = \frac{s^3 + b_2 s^2 + b_0}{a_3 s^3 + a_2 s^2 + a_1 s + a_0}, \qquad (48)$$

where:

$$\begin{aligned}
a_3 &= 1 + C_{gb2}/C_{gd2},\\
a_2 &= (r_{o2} + Z_{M1} + g_{m2} r_{o2} Z_{M1})/(C_{gs2} r_{o2} Z_{M1})\\
&\quad + (r_{o2} + Z_{M1})/(C_{gd2} r_{o2} Z_{M1})\\
&\quad + C_{gb2}(1 + g_{m2} r_{o2} + r_{o2}/Z_{M1})/(C_{gs2} C_{gd2} r_{o2}),\\
a_1 &= 1/(C_{gd2} L_{g2}),\\
a_0 &= (1 + g_{m2} r_{o2} + r_{o2}/Z_{M1})/(C_{gs2} C_{gd2} r_{o2} L_{g2}),\\
b_2 &= 1/(C_{gs2} r_{o2}) + 1/(C_{gd2} r_{o2}) + g_{m2}/C_{gs2}\\
&\quad + C_{gb2}/(C_{gs2} C_{gd2} r_{o2}),\\
b_0 &= 1/(C_{gs2} C_{gd2} r_{o2} L_{g2}),
\end{aligned}$$

r_{o2} is the drain-source resistor of transistor M_2, and Z_{M1} is the equivalent impedance looking into the drain of transistor M_1. Note that L_{g2} and C_{gd2_ext} have to be chosen properly for enhanced reverse isolation in the desired frequency band. To simplify the assessment of L_{g2}'s effect on the reverse isolation from V_{out_LNA} to V_y, Z_{M1} can be replaced by the drain-source resistance (r_{o1}) of M_1. However, L_{g2} was modeled as a non-ideal inductor based on the information in [30] and [31]. The macromodel simulation results in Fig. 10(b) show that the reverse isolation (V_y/V_{out_LNA}) has a notch and a peak as predicted by equation (48). Note that the depth of the notch and magnitude of the peak are overestimated with the simulations of the macromodel with an inductor having a high quality factor of 12.5. With a reduced L_{g2} quality factor of 7.1, the peak in this example reduces by 5.3 dB. In practice, the peak and notch depth are further reduced by parasitic capacitances from routing and devices that are not included in the model, which further improves stability by lowering the quality factor of the network at the gate of transistor M_2. In addition, the peak in particular is further reduced by transconductance degradation at high frequencies.

The stability factor of the LNA is defined in [28] and [29]:

$$K = \frac{1 + |\Delta|^2 - |S11|^2 - |S22|^2}{2 \cdot |S21| \cdot |S12|}, \qquad (49)$$

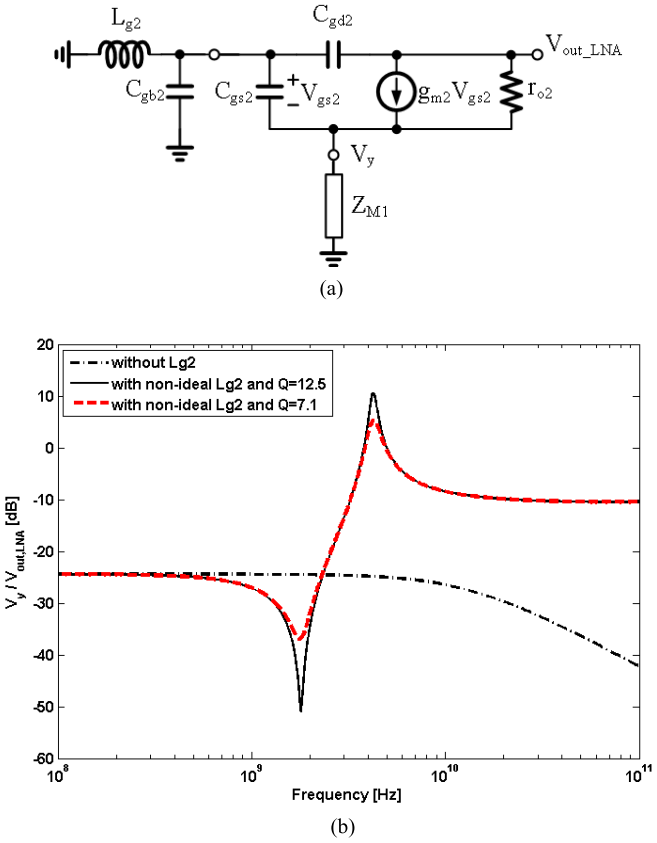

Fig. 10. (a) Simplified small-signal model of the cascode stage for reverse isolation analysis, and (b) simulated reverse isolation from V_{out_LNA} to V_y with and without L_{g2}.

where $\Delta = S11 \cdot S22 - S12 \cdot S21$. The unconditional stability requirement is $K > 1$ and $|\Delta| < 1$. Note that S11 and S22 are close to zero when the input and output of the LNA are matched to the source and load impedances. Based on the measured S-parameters in Section IV, the value of $|\Delta|$ is less than 1 and the value of K is more than 1 in the frequency range from 0.1 GHz to 8.5 GHz. The $|\Delta|$ and K values are 0.05 and 17.67 at 1.8GHz, respectively.

From LNA simulations without buffer, the reverse isolation at 1.8 GHz with $L_{g2} = 3.5$ nH and $C_{gd2_{ext}} = 150$ fF is slightly better (-29.7 dB) than without L_{g2} and C_{gd2_ext} (-27.4 dB). As S12 decreases, the value of K increases and $|\Delta|$ decreases, resulting in better stability. However, it is important to consider that the values of L_{g2} and C_{gd2_ext} can degrade reverse isolation and stability if they are not carefully selected. If L_{g2} and C_{gd2_ext} become too large, then the peak of the transfer function in equation (48) moves from higher to lower frequency, which can cause a stability problem.

IV. Experimental Results

A 1.8 GHz linearized subthreshold LNA has been designed and fabricated in Dongbu 0.11μm CMOS technology. Fig. 11 displays the chip micrograph of the LNA with an area of 810 μm $\times$ 770 μm. Table I lists the key design parameters of the LNA. It consumes a 480 μA current (with exclusion of the buffer) from a 0.7 V power supply instead of the nominal 1.2 V supply voltage for this technology. In order to limit the linearity degradation due to the output buffer that was designed

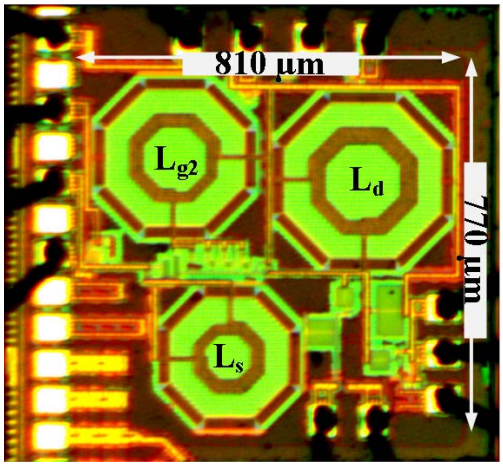

Fig. 11. Chip micrograph of fabricated linearized subthreshold LNA.

TABLE I
LNA DESIGN PARAMETERS

Component	Value
V_{DD}	0.7 V
I_D	480 µA
$g_{m,M1}/I_D$	22 S/A
L_{g1}	7.5 nH
L_{g2}	3.5 nH
L_s	2.4 nH
C_{gs1_ext}	130 fF
C_{gd2_ext0}	70 fF
C_{gd2_ext1}	20 fF
C_{gd2_ext2}	40 fF
C_{gd2_ext3}	80 fF
L_d	6.4 nH
C_d	88 fF
R_d	720 Ω
W/L per finger ($M_{1,2}$)	6µm / 0.13µm
Number of fingers ($M_{1,2}$)	64

to test the LNA, a 1.2 V supply is used for the buffer. The L_{g2} value of this design was selected to be 3.5 nH (with a quality factor of 6.5 at 1.8 GHz), and final post-layout simulations were performed with foundry-supplied device models for all on-chip components. The prototype chip was bonded to a conventional QFN16 package that was assembled on a printed circuit board for measurements.

A. Performance

The control switch settings for C_{gd2_ext} of $D_3D_2D_1 = 100$ resulted in the best linearity after fabrication process variations. Fig. 12 shows the measured scattering parameters of the linearized subthreshold LNA. S11 and S22 are both below −10 dB at 1.8 GHz. The measured voltage gain of 9.5 dB at 1.8 GHz is the combination of the LNA and buffer. Also note that S12 is under −40 dB around the frequency of interest. Fig. 13 displays the plot of the measured noise figure (NF) that is 6.3 dB at 1.8 GHz with the buffer. The voltage gain and noise figure of the LNA are 14.8 dB and 3.7 dB after de-embedding the effects of the buffer stage loss and SMA cables. As part of the de-embedding process, the simulated gain and noise figure of the LNA and buffer combination were compared to the measurement results. The overall noise figure difference between simulations and measurements was

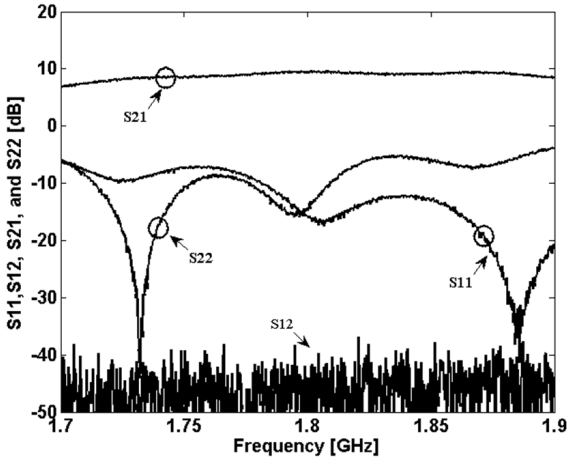

Fig. 12. Measured scattering parameters of the LNA with buffer stage (−5.3 dB gain).

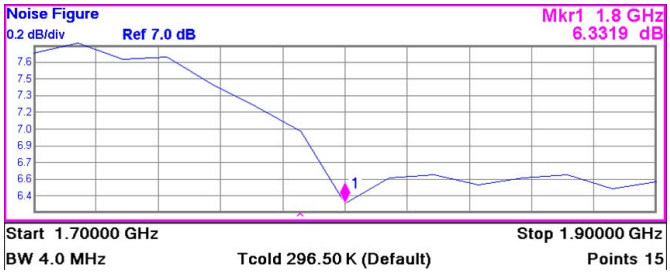

Fig. 13. Measured noise figure of the LNA with buffer stage (−5.3 dB gain).

within 0.8 dB, which supports that the measured LNA gain is at least 14.8 dB because any significant reduction of the LNA gain would significantly degrade the overall noise figure of the combined LNA and buffer stages. Notice that the −5.3 dB gain of the buffer in Fig. 3 is in the presence of parasitics due to the bonding, the integrated circuit package, and the PCB at the interface to the measurement equipment with 50Ω termination. Fig. 14 shows the measured input IIP3 of the LNA, and the output spectrum from a test with a two-tone input signal (1.8 GHz and 1.7995 GHz) and an input power of −35 dBm. Fig. 15 contains the plot of output power measurements from a power level sweep of a single 1.8 GHz tone to determine the 1-dB compression point (P_{1dB}) of the LNA. The IIP3 and P_{1dB} of the linearized LNA are −3.7 dBm and −12.6 dBm, respectively. Fig. 16 confirms that $K > 1$ and $|\Delta| < 1$ within the 8.5 GHz range of the network analyzer used for the measurements.

Table II summarizes the performance of narrowband low-power RF LNAs with operating frequencies ranging from 1 to 3 GHz in comparison to the presented design. Most of the reported low-power LNA measurement results reveal that the IIP3 is constricted to below −10 dBm except in this work as well as in [17] and [32]. Compared to this work with packaged chip measurements at 1.8 GHz on a printed circuit board, the 1 GHz LNA in [32] was measured on a probe station. Furthermore, the design in [32] contains a single transistor and three 11 nH inductors. Even though the cascode stage of a subthreshold LNA has negative impact on linearity as discussed in Section III, we opted to include the cascode transistor to maintain adequate reverse isolation (S12) in this work with linearization enhancement. On a different