# Design of CMOS RF Low-Noise Amplifiers and Mixer for Wireless Applications

by

Lou Shuzuo

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in Department of Electronic and Computer Engineering

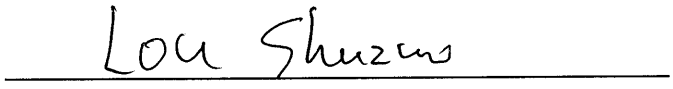
August 2007, Hong Kong

## Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.



Lou Shuzuo

Design of CMOS RF Low-Noise Amplifiers and Mixer for Wireless Applications

by
Lou Shuzuo

This is to certify that I have examined the above PhD thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.

**Prof. Howard C. Luong**, Thesis Supervisor

**Prof. Z. Lu**, Thesis Examination Committee Member (Chairman)

**Prof. K. S. Wong**, Thesis Examination Committee Member

**Prof. C. F. Chan**, Thesis Examination Committee Member

**Prof. Kevin J. Chen**, Thesis Examination Committee Member

**Prof. Ross D. Murch**, Thesis Examination Committee Member

**Prof. Khaled Ben Letaief**, Head of Department of Electronic and Computer Engineering

Department of Electronic and Computer Engineering

August 2007

# Acknowledgements

I would like to express my sincere gratitude to many people who have helped and supported me during last 5 years of my study in HKUST.

My supervisor, Prof. Howard Luong, introduced me into the exciting field of RF integrated circuit design, I am deeply indebted to him for his constant support and guidance during these years. When I encountered difficulties on the design and became discouraged, his suggestion and encouragement helped me overcome the obstacles in front of me. I am grateful for his great patience and encouragement.

I would like to thank Prof. Lu Zongli, Prof. Chan Cheong-fat, Prof. Wong Kam Sing, Prof. Kevin Chen Jing, and Prof. Ross D. Murch for being my thesis exam committee.

In addition, I wish to thank the laboratory technicians, Frederick Kwok, for his technical support on measurement equipments and PCB board making, Mr. Chan Kwok Wai for the measurement support, and Mr. S. F. Luk for his help in setting up the CAD tools. My colleagues, Dr. Lincoln Leung, Tay Zheng, Evelyn Wang, Alan Ng, Lu Dong Tian, Annby Rong, Shen Cheng, Chan Tat Fu, Camel Lok, and Kay Chui in the Analog Research Laboratory provided me a lot of useful discussion related to my work, and helped me kindly in my life as well. My former labmates, Dr. Vincent Cheung, Dr. Rachel Wang, Kwok Kachun, Patrick Wu, Gary Wong, Gerry Leung, Sophia Song, and Dennis Lau shared a lot of experience with me when I first came to the lab. I am grateful to them all.

Finally, I wish to thank my family for their support during my postgraduate study. I miss them very much.

# Table of Contents

Title Page

Authorization Page

Acknowledgements

Table of Contents

List of Figures

List of Tables

# List of Figures

# List of Tables

# Design of CMOS RF Low-Noise Amplifiers and Mixer for Wireless Applications

by Lou Shuzuo

Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology

## Abstract

The recent upsurge in the demand for low-power portable wireless communication products creates a lot of research opportunities. Popular applications include mobile phones, wireless local area networks (WLANs), wireless personal area networks (WPANs), radio frequency identification systems (RFIDs). On the other hand, complementary metal-oxide-semiconductor (CMOS) process has been advanced and become very attractive for radio-frequency (RF) circuits to be integrated with the digital integrated circuits, which can reduce the cost, size, and power dissipation for the wireless systems.

Receiver front-end (RFE), including low-noise amplifier (LNA) and mixer, plays an important role in wireless receivers. Not only does it operate at high frequencies, but it also needs to have low noise and good linearity while consuming low power. In this thesis, wideband and high linearity circuit techniques are proposed for CMOS RFE, including fully integrated LNAs for WLAN, cable TV, ultra-wideband (UWB), and RFID applications.

Inductive series peaking technique is proposed as wideband output loading to enable an UWB LNA to work from 3 to 8GHz. A $2^{nd}$-intermodulation (IM) injection technique is proposed to linearize the RFID LNA. A new mixer topology, with LO

switching pair at the bottom of the transconductance cells and with an inductor in series with the LO switching pair, is also proposed to achieve high performance in terms of frequency and linearity.

The LNAs and mixers proposed are all successfully integrated as part of the transceiver for WLAN, cable TV, UWB and RFID reader applications.

# Chapter 1 Introduction

## 1. 1. Background and Motivation

Decades of continuous development of wireless communication has brought about many different standards. Communication technology is moving toward a major milestone. The explosive growth of the wireless industry, global access to the internet, and the ever increasing demand for high speed data communication are spurring us toward rapid developments in communication technology. Wireless communication plays an essential role in the transformation to the next generation communication systems. Cellular phones, pagers, wireless local area networks (WLAN), global positioning system (GPS) handhelds, and short-range data communication devices employing Bluetooth and ultra wideband (UWB) technologies are all examples of portable wireless communication devices. Nowadays, driven by the big commercial demand for low-cost and low-power multi-standard portable devices, radio-frequency (RF) designers are urged to develop new methodologies that allow the design of such products.

The front-end low-noise amplifier (LNA) and mixer are important components of any RF receiver. Fig. 1. 1 depicts the simplified structure of an RF receiver. The received signal is typically filtered, amplified by an LNA, and translated to the base-band by mixing with a local oscillator (LO). After being demodulated, the signal is applied to an analog-to-digital converter (ADC) which digitizes the analog signal. The digital signal is then processed in a digital signal processing unit (DSP). As the first active building block in the receiver front-end (RFE), the LNA should provide considerable gain while minimizing noise. The mixer needs to down-convert the signal with good noise performance and high linearity. As both LNA and mixer work

in high-frequency, they face the parasitic effect from active and passive devices. The parasitics can greatly affect sensitivity and noise parameters of the overall receiver.

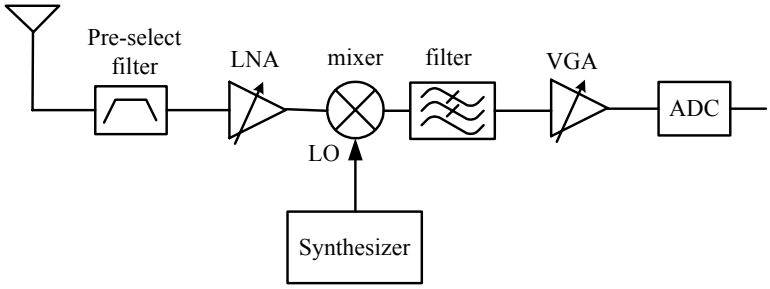



Fig. 1. 1 Block diagram of a generic RF receiver

**Choosing CMOS**

Currently, most commercial RF transceivers are implemented as multi-chip modules (MCMs) or system in packages (SiPs), using various technologies. Base-band and mixed-signal components (e.g., DAC, ADC, and DSP) are mainly implemented in complementary metal-oxide semiconductor (CMOS) technology, while RF and analog sections are typically implemented in silicon-germanium (SiGe) or gallium-arsenide (GaAs) technologies [1. 1]-[1. 3]. High quality passive filters are mostly realized as discrete components. MCM and SiP approaches suffer from many shortcomings, such as large size, high power consumption, and high integration cost. The aforementioned problems account for the global trend toward a single technology that can support a commercially viable single chip RF transceiver.

Historically, CMOS technology was not considered to a be good candidate for analog and RF applications. Relatively small transconductance, low drive capability, and poor quality of on-chip passive elements are among several limitations for this technology. However, the incredible growth of the digital industry due to the continuous scaling in CMOS technology has motivated designers to develop analog and RF CMOS circuits that can be integrated along with the digital circuitry. This has

led to the tremendous research and development in implementing single chip systems. Furthermore, the transit frequency ($f_t$) of MOS devices has increased due to the evolution of CMOS into deep-sub-micron technologies. This greatly improves the performance of integrated RF CMOS circuits.

Therefore, despite the inferior performance of RF CMOS circuits compared to their SiGe and GaAs counterparts, the dominance of CMOS in the digital world, combined with the feasibility of integrating digital/analog/RF circuits on a single chip and the potential cost and power advantage of this integration, provide reasonable motives to adopt CMOS over other technologies.

**Challenges**

Recently, there has been a tremendous effort to develop wireless devices that integrate multiple applications (phone, video-game console, navigator, digital camera, web browser, etc.) on a single chip. A variety of standards exist for each of these applications, such as global system for mobile communication (GSM) and universal mobile and telecommunication system (UMTS) for cellular telephony, IEEE 802.11a/b/g and HiperLNA2 for LAN access, Bluetooth for short-range communication, and GPS. The growing number of these wireless communication standards, promotes the need for a multi-standard transceiver. The RF front-end of such a receiver has to cover a wide range of different carrier frequencies. To achieve this goal, wideband performance of the receiver front-end is desired. A variety of architectures have been proposed to fulfill this requirement. One approach is to use a parallel combination of several tuned narrowband LNAs. This solution, although straightforward, is power hungery and area inefficient. Therefore, it is not particularly suitable for low-cost portable systems. Two other LNA architectures that can be used for multi-standard signal reception are concurrent LNA [1. 4] and tunable LNA [1. 5].

The former technique is applicable when the frequency bands of desired standards are well separated, while the latter design approach is complicated if a wideband tuning range is desired. An alternative solution is to design a single wideband LNA covering the entire band of interest, which is one subject of this research.

Linearity is another important parameter for LNA and mixer. It is a measure of the intermodulation (IM) distortion that LNA and mixer experience due to the presence of unwanted spurious tones in the vicinity of the desired frequency band. Despite considerable advances in technology, nonlinearity of devices continues to limit the performance of wireless systems. In the design of receivers, nonlinearity restricts the ability of a radio to receive weak signals in the presence of nearby stronger signals. Linearity is usually a tradeoff between gain, power consumption, and noise, therefore, low cost linearization technique is highly desired.

The objective of this thesis is to develop CMOS LNA and mixer design techniques for high gain, low noise, and high linearity performance with low power consumption in both narrowband and wideband applications, and integrate the front-end circuits with the analog part to realize the complete functionality of the RF receiver.

## 1. 2. Thesis Outline

The thesis is organized into 7 chapters. LNA and mixer for various applications are discussed, such as wireless local area network (WLAN), cable TV tuner, ultra wideband (UWB), and radio-frequency identification (RFID). As the ultimate goal for LNA design to cover all the commonly-used standards as a software-defined radio (SDR) LNA, since it covers all standards with different specifications and focus, all techniques for narrowband and wideband, low NF, high gain, high linearity are investigated. Each of the techniques was verified and demonstrated in single chip system before combining together for an SDR LNA.

Chapter 2 reviews the fundamentals of LNA and mixer design, such as noise figure and input matching.

Chapter 3 discusses narrowband LNA design. Topology consideration and noise optimization with on-chip inductors are then presented. Linearization technique is also proposed including analysis using 2$^{nd}$ IM injection.

Chapter 4 presents wideband LNA design. To enable large gain-bandwidth product (GBW), wideband loading networks are used and discussed, including T-coil and inductive peaking. Wideband input matching topologies are also reviewed and compared.

Mixer design is discussed in Chapter 5. New topology is proposed with LO switches at the bottom of the transconductors, so that the mixer is immune to parasitic capacitor and suitable for high-frequency and wideband operation.

Receiver front-end design for software-defined radio (SDR) is discussed in Chapter 6. Here, we will build the SDR front-end to cover many standards below 10 GHz. Finally, the conclusion and the future works are presented in Chapter 7.

## REFERENCES


[1. 1] Y. Mimino, M. Hirata, K. Nakamura, K. Sakamoto, Y. Aoki, and S. Kuroda, "High gain-desity K-band P-HEMT LNA MMIC for LMDS and satellite communication," in *IEEE Radio Frequency Integrated Circuits Symp. Dig. Papers*, pp. 209-212, 2000.

[1. 2] Y. Yun, M. Nishijima, M. Katsumo, H. Ishida, K. Minagawa, T. Nobusada, and T. Tanaka, "A fully integrated broad-band amplifier MMIC employing a novel chip-size package," *IEEE Trans. Microwave Theory Tech.*, Vol. 50, pp. 2930-2937, Dec. 2002.

[1. 3] P. Marsh, S. Chu, S. Lardizabal, R. Leoni III, S. Kang, R. Wohlert, A. Bowlby, W. Hoke, R. McTaggart, C. Whelan, P. Lemonias, P. McIntosh, and T. Kazior, "Low noise metamorphic HEMT devices and amplifiers on GaAs substrates," in *IEEE Microwave Theory and Techniques Symp. Dig. Papers*, pp. 105-108, 1999.

[1. 4] H. Hashemi, and A. Hajimiri, "Concurrent multiband low-noise amplifiers – theory, design, and application," *IEEE Trans. Microwave Theory and Techniques*, Vol. 50, No. 1, pp. 288-301, Jan. 2002.

[1. 5] S. Andersson, P. Caputa, and C. Svensson, "A tuned, inductorless, recursive filter LNA in CMOS," in *Proc. European Solid-State Circuits Conference*, pp. 351-354, Sep. 2002.

# Chapter 2 RF Receiver Fundamentals

This chapter provides some general background to facilitate the discussion of the LNA and mixer. Important parameters for evaluating the performance of the RF front-end will be described.

## 2. 1. Noise Figure (NF)

The signals received at the antenna most likely are very weak and have to be amplified in order to driver the mixer so as not to deteriorate the signal-to-noise ratio (SNR) of the received signal. Noise figure is a measure of the amount of the noise added after the signal goes through a circuit, and it is defined as the ratio of the available output noise power to available output noise due to the source (a 50-Ω resistor), i.e.

$$F = \frac{S_i / N_i}{S_o / N_o} = \frac{SNR_{in}}{SNR_{out}} = \frac{S_i}{N_i}\frac{N_i G + N_{c,o}}{S_i G} = 1 + \frac{N_{c,o}}{N_i G} = 1 + \frac{N_{c,i}}{N_i} \tag{2. 1}$$

where $N_{c,i}$ and $N_{c,o}$ are circuit's input-referred and output-referred noise, respectively. $G$ is the gain of the circuit and $N_i$ is the source noise power. NF is usually expressed in dB and noise factor $F$ is the corresponding value in linear scale, i.e.

$$NF(in\ dB) = 10 \cdot \log F \tag{2. 2}$$

There exists optimum source admittance for small NF. A general two-port noise model is shown in Fig. 2. 1. A noisy two-port is driven by a source that has an admittance $Y_s$ and an equivalent shunt noise current $i_s$. The noises from internal noise sources are all referred to the circuit input as $e_n$ and $i_n$ because we are concerned only with overall input-output behavior. Therefore, the expression for noise figure becomes

$$F = \frac{\overline{i_s^2} + \overline{\left|i_n + Y_s e_n\right|^2}}{\overline{i_s^2}} \tag{2. 3}$$

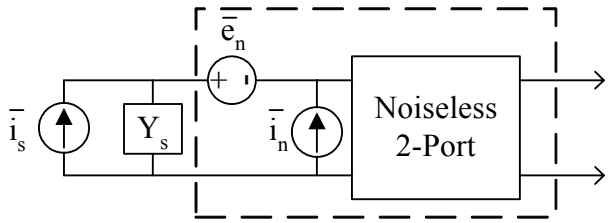



Fig. 2. 1 Noisy two-port driven by noisy source

In order to accommodate the possibility of correlations between $e_n$ and $i_n$, express $i_n$ as the sum of two components. One, $i_c$, is correlated with $e_n$, and the other, $i_u$, isn't:

$$i_n = i_c + i_u \tag{2. 4}$$

Since $i_c$ is correlated with $e_n$, it can be written as:

$$i_c = Y_c e_n \tag{2. 5}$$

where the constant $Y_c$ is the correlation admittance.

Combining (2. 3)~(2. 5), the noise figure becomes

$$F = \frac{\overline{i_s^2} + \overline{\left|i_u + (Y_c + Y_s)e_n\right|^2}}{\overline{i_s^2}} = 1 + \frac{\overline{i_u^2} + \left|Y_c + Y_s\right|^2 \overline{e_n^2}}{\overline{i_s^2}} \tag{2. 6}$$

If the independent noise sources are treated as thermal noise produced by an equivalent resistance or conductance:

$$R_n \equiv \frac{\overline{e_n^2}}{4kT\Delta f}, \quad G_u \equiv \frac{\overline{i_u^2}}{4kT\Delta f}, \quad G_s \equiv \frac{\overline{i_s^2}}{4kT\Delta f} \tag{2. 7}$$

(2. 6) becomes:

$$F = 1 + \frac{G_u + \left|Y_c + Y_s\right|^2 R_n}{G_s} = 1 + \frac{G_u + \left[(G_c + G_s)^2 + (B_c + B_s)^2\right]R_n}{G_s} \tag{2. 8}$$

where $Y_c$ and $Y_s$ are decomposed into a sum of a conductance $G$ and a susceptance $B$. Therefore, from (2. 8), the optimum source admittance $Y_s$ can be written as:

$$\begin{aligned} B_s &= -B_c = B_{opt} \\ G_s &= \sqrt{G_u / R_n + G_c^2} = G_{opt} \end{aligned} \tag{2. 9}$$

The source admittance for minimizing NF is usually not the same as that for maximizing power transfer, degradation in power transfer needs to be accepted when noise performance is to be optimized.

In the receiver path, as the signal propagates from the antenna to digital back-ends, different blocks may introduce noise to the signal. The overall NF of the receiver depends on the NF of each block as well as the gain of preceding stages. Intuitively, larger signals are less susceptible to noise, and this is why the larger gain of one stage makes the noise of the following stage less important. Friis shows that the overall NF of a cascaded system (such as the one shown in Fig. 2. 2) is given by [2. 1]:

$$NF_{tot} = NF_1 + \frac{NF_2 - 1}{A_1} + \frac{NF_3 - 1}{A_1 A_2} + \cdots \tag{2. 10}$$

where $NF_i$ and $A_i$ are the NF and available power gain of each stage in linear scale, respectively. It accounts for the fact that the low noise of the front-end stages, i.e., low $NF_1$ and $NF_2$ are of great importance in the receiver design.

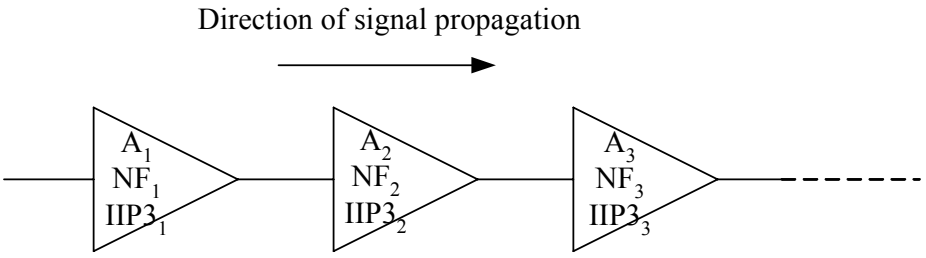



Fig. 2. 2 Cascade of several blocks

For a passive reciprocal network, the NF is equal to its loss. If we calculate the overall NF of a lossy filter followed by an LNA, the overall NF $NF_{tot} = L \cdot NF_{LNA}$, where $L$ is the loss of the filter. This relationship is useful when we do NF de-embedding from the measurement result.

Sensitivity is defined as the minimum signal level that a circuit or system can detect with certain SNR. $NF_{tot}$ determines the sensitivity of the overall receiver. This relation is analytically given by

$$Sensitivity\ (dBm) = -174dBm/Hz + NF_{tot} + 10\log B + 10\log(SNR_{out}) \qquad (2.\ 11)$$

where -174dBm/Hz is the available noise power from the antenna (the noise floor), $B$ is the bandwidth of the desired signal, and the last term is the minimum acceptable SNR at the receiver output, which is a function of minimum required bit-error-rate (BER) at the output of the demodulator. As can be seen from (2. 10) and (2. 11), low NF of the LNA greatly improves the sensitivity of the overall receiver.

## 2. 2. Linearity

While many analog and RF circuits can be approximated with a linear model to obtain their response to small signals, nonlinearities often lead to important phenomena. If we assume the circuits we analysis have the transfer function

$$y(t) \approx a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) \qquad (2.\ 12)$$

The coefficients $a_2$ and $a_3$ provide information on the nonlinearity. When a sinusoidal signal $A$*cosωt is applied to the system in (2. 12), the output y(t) would be

$$y(t) = \frac{a_2 A^2}{2} + \left( a_1 + \frac{3a_3 A^2}{4} \right) A\cos\omega t + \frac{a_2 A^2}{2}\cos 2\omega t + \frac{a_3 A^3}{4}\cos 3\omega t \qquad (2.\ 13)$$

From (2. 13), the output contains not only the fundamental frequency term, but also many higher order harmonics caused by $x^2(t)$ and $x^3(t)$.

Due to nonlinearities, harmonics, gain compression, desensitization, blocking, cross modulation and intermodulation (IM) arise.

### 2.2.1 1-dB Compression Point

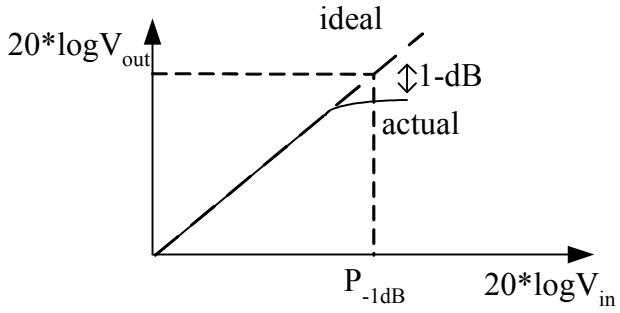



Fig. 2. 3 Plot of 1-dB compression point

1-dB compression point is defined as the input signal amplitude at which the output signal deviates from the ideal response by 1-dB, as depicted in Fig. 2. 3. In (2. 13), the first-order coefficient includes two terms, the desired gain $a_1$ and the undesired term $3a_3A^3/4$. For small input amplitude *A*, the first term dominates and the output is linearly dependent on the input. However, due to nonlinearity, when the input signal amplitude is large, the gain of the fundamental frequency begins to diminish because $a_1$ and $a_3$ are usually of opposite sign, thus, the term $a_1$ and $3a_3A^3/4$ cancel each other. This is called gain compression. If $a_1$ and $a_3$ are of the same sign, gain expansion will happen.

1-dB compression point of the front-end needs to be designed larger than the maximum input signal so as not to cause large distortion in the received signal.

### 2.2.2 Third-order Intermodulation Distortion

The corruption of signals due to third-order IM (IM3) of two nearby interferers is common and critical, as shown in Fig. 2. 4, a weak signal accompanied by two strong interferers experiences third-order nonlinearity, then one of the IM product falls in the band of interest, corrupting the desired component while the modulation is operating on the amplitude of the signal. This effect degrades the performance even if the modulation is on the phase (because zero-crossing points are still affected).

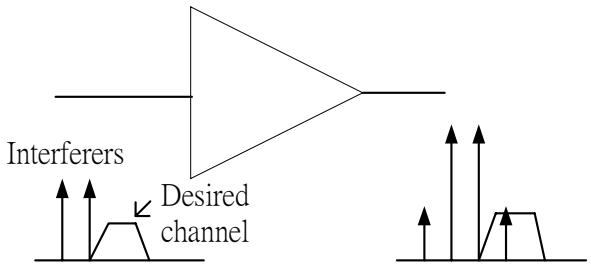



Fig. 2. 4 Illustration of third-order intermodulation products

Third-order intercept point (IP3) is defined to characterize third-order IM behavior. IP3 is measured by a two-tone test in which input amplitude *A* is chosen to be sufficiently small so that higher-order nonlinear terms are negligible and the gain is relatively constant and equal to $a_1$. With *A* increasing, plotting fundamental and IM3 outputs vs. input signal on a logarithmic scale, we see that the magnitude of the IM products grows at 3 times the rate as compared to the fundamental outputs. The IP3 is defined to be at the intersection of the two lines, as shown in Fig. 2. 5. The horizontal coordinate of this point is called the input IP3 (IIP3), and the vertical coordinate is called the output IP3 (OIP3).

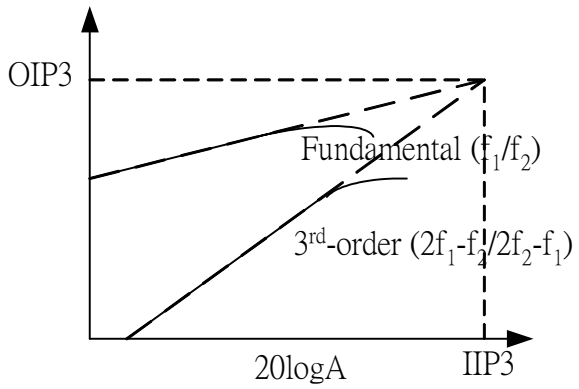



Fig. 2. 5 The plot of IIP3

The overall linearity of a receiver consisting of a cascade of several blocks depends on the gain as well as the linearity of each stage. This can analytically be shown by considering the chain of Fig. 2. 2, and expressing the input-output relation of the

different stages in the chain. However, finding a closed-form expression for the linearity of overall system is rather difficult. If two stages in cascade are considered, the overall IIP3 can be written as [2. 1]:

$$\frac{1}{A_{IIP3,tot}^2} = \frac{1}{A_{IIP3,1}^2} + \frac{3a_2 b_2}{2b_1} + \frac{a_1^2}{A_{IIP3,2}^2} \tag{2. 14}$$

where $A_{IIP3,i}$ is the IIP3 of the $i$-th stage ($i$ = 1, 2) in linear scale. $a_1$ and $a_2$ are the 1st and 2nd order coefficients of 1st stage's transfer function, $b_1$ and $b_2$ are the 1st and 2nd order coefficients of the 2nd stage's transfer function.

A careful examination of (2. 14) reveals that, the 1st term represents the IM3 that is generated at the 1st stage output and amplified by the 2nd stage linearly; the 3rd term represents the IM3 that is generated by the 2nd stage; the middle term represents the IM3 that comes from IM2 at the 1st stage output mixing with the fundamental in the 2nd stage through its 2nd order nonlinearity. In many RF systems, each stage in a cascade has a narrow frequency band. Thus, IM2 falls out of the band and is heavily attenuated, the middle term described in (2. 14) becomes negligible (can be attributed to small $a_2$), giving

$$\frac{1}{A_{IIP3,tot}^2} \approx \frac{1}{A_{IIP3,1}^2} + \frac{a_1^2}{A_{IIP3,2}^2} \tag{2. 15}$$

This equation readily gives a general expression for three or more stages:

$$\frac{1}{A_{IIP3,tot}^2} \approx \frac{1}{A_{IIP3,1}^2} + \frac{a_1^2}{A_{IIP3,2}^2} + \frac{a_1^2 b_1^2}{A_{IIP3,3}^2} + \cdots \tag{2. 16}$$

If each stage in a cascade has a gain greater than unity, then the nonlinearity of the following stage becomes more critical. This expression also states that the high gain of LNA degrades the overall linearity of the system. This is in contrast with the NF scenario in which the high gain of the LNA improves the overall NF.

## 2. 3. Input Matching

To deliver the maximum power from the antenna to the LNA, matching to the impedance of antenna, e.g., 50-Ω, is required at the input port of the LNA. For wideband applications, this impedance matching should be obtained over a wide frequency range and is usually a major challenge considering the noise and power consumption requirement.

To quantify the degree of the impedance matching, it is customary to introduce the concept of voltage standing wave ratio (VSWR) [2. 2]:

$$VSWR = \frac{1+|\Gamma|}{1-|\Gamma|} \tag{2. 17}$$

where Γ is the reflection coefficient and is defined as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} \tag{2. 18}$$

In this expression, $Z$ is the actual input impedance and $Z_0$ is the characteristic impedance of the source, which is usually equal to 50-Ω. Perfect matching ($Z = Z_0$) results in $\Gamma = 0$ (-∞ dB) and equivalently VSWR = 1. However, for practical purposes $|\Gamma| < -10$ dB is usually sufficient to meet the matching requirement.

As LNA and mixer can be viewed as two-port networks with input and output ports, scattering parameters (S parameters) are more frequently used in characterizing the performance, which defines the four variables as the incident/reflected input/output voltage wave. The definition of S parameters exploits the fact that a transmission line terminated in its characteristic impedance does not reflect any power at its termination [2. 2]. Consider the block diagram of a two-port network shown in Fig. 2. 6, where $Z_0$ is the impedance of the source and the load terminations, $E_{ii}$ and $E_{ri}$ are the magnitude of incident and reflected voltage waves at $i$-th port ($i$ = 1, 2). The S parameter coefficients are expressed as

$$b_1 = s_{11}a_1 + s_{12}a_2$$
$$b_2 = s_{21}a_1 + s_{22}a_2 \qquad (2.\ 19)$$

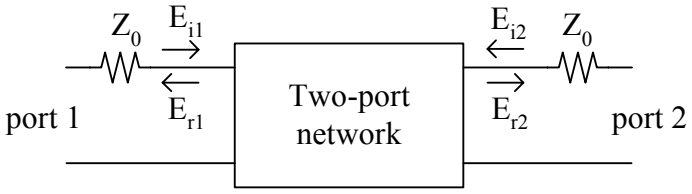



Fig. 2. 6 S parameters definition of two-port networks

where
$$a_1 = E_{i1}/\sqrt{Z_0} \qquad a_2 = E_{i2}/\sqrt{Z_0}$$
$$b_1 = E_{r1}/\sqrt{Z_0} \qquad b_2 = E_{r2}/\sqrt{Z_0} \qquad (2.\ 20)$$

The normalization to the square root of $Z_0$ makes the square of magnitude of $a_i$ and $b_i$ equal to the incident and reflected power at both ports.

Assume port 1 and port 2 are the input and output ports, respectively. It can be shown [2. 2] that $s_{11} = \Gamma$ and $s_{21}$ represents the forward gain of the network. We can predict that a good amplifier should possess a large $s_{21}$ to achieve high gain, small $s_{11}$ and $s_{22}$ to possess good input and output matching, and very small $s_{12}$ to ensure stability and reverse isolation. S parameters are convenient and useful in characterizing two-port networks at high frequency, we usually measure S parameters using network analyzer for LNA and mixer testing.

## 2. 4. Image Rejection

Image signal is a problem related to frequency conversion. A mixer is usually used in a receiver to down-convert the signal from RF frequency to IF frequency, as shown in Fig. 2. 7. For example, if the RF signal is $A_{RF}*\cos\omega_{RF}t$ and the LO signal is $A_{LO}*\cos\omega_{LO}t$, then the output of the mixer is proportional to

$$A_{RF}\cos\omega_{RF}t * A_{LO}\cos\omega_{LO}t = 1/2 * A_{RF}A_{LO}[\cos(\omega_{LO} - \omega_{RF})t + \cos(\omega_{LO} + \omega_{RF})t] \quad (2.\ 21)$$

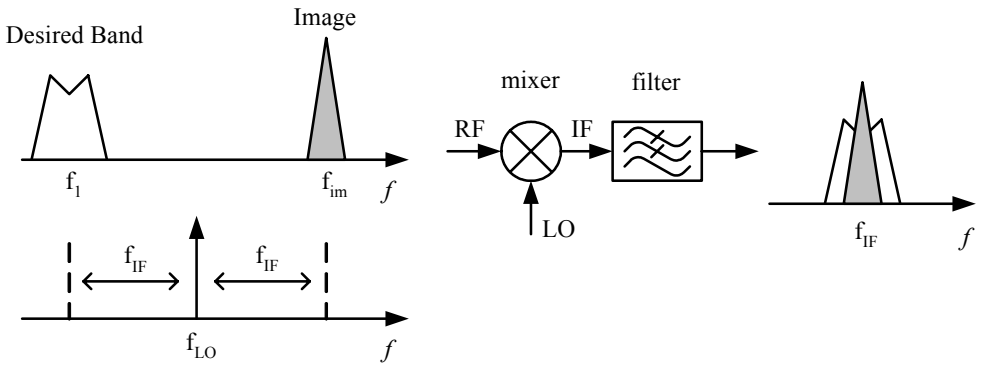



Fig. 2. 7 Problem of image

The component at frequency ($\omega_{LO}+\omega_{RF}$) in (2. 21) is filtered out by the filter. The component at frequency ($\omega_{LO}-\omega_{RF}$) is the desired signal, and IF frequency is $\omega_{IF}$ = ($\omega_{LO}-\omega_{RF}$). If there is a signal at frequency $\omega_{IM}$ = ($\omega_{LO}+\omega_{IF}$), it will be converted to IF frequency as well and can't be distinguished with the desired signal. This signal is called the image signal.

Image signal can be rejected by the image-reject (IR) filter before mixing to alleviate the image mixing problem, as shown in Fig. 2. 8.

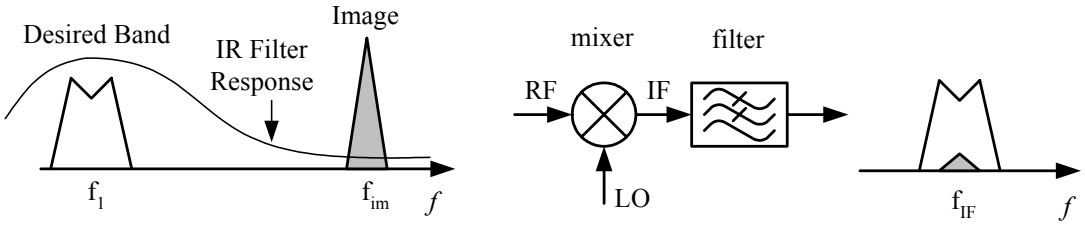



Fig. 2. 8 Image rejection filter to remove image

## 2. 5. High-Frequency Figures of Merit for Active Device

In the specific case of high-frequency performance, two figures of merit are particularly popular. These are $\omega_T$ and $\omega_{max}$, which are the frequencies at which the current and power gain of the active device are extrapolated to fall to unity. If $i_d$ and $i_g$

represent the output drain current and input gate current, respectively, the following equation results:

$$\left|\frac{i_d}{i_g}\right| \approx \left|\frac{v_{in} g_m}{v_{in} s(C_{gs} + C_{gd})}\right| = \frac{g_m}{\omega(C_{gs} + C_{gd})} \tag{2. 22}$$

where $g_m$ is the transconductance of the active device, $C_{gs}$ and $C_{gd}$ are the gate-source and gate-drain parasitic capacitors. (2. 22) has a value of unity at a frequency

$$\omega_T = \frac{g_m}{C_{gs} + C_{gd}} \tag{2. 23}$$

Computing $\omega_{max}$ is quite difficult, it is shown in [2. 3] that

$$\omega_{\max} \approx \frac{1}{2}\sqrt{\frac{\omega_T}{r_g C_{gd}}} \tag{2. 24}$$

where $r_g$ is the gate resistance.

$\omega_T$ and $\omega_{max}$ show important characteristic of the active devices. If $\omega_T$ and $\omega_{max}$ are close to the operating frequency $\omega$, active devices can not provide enough gain, gain and noise of circuits will be significantly degraded. It is always a tradeoff between good circuit performance and high cost of advanced devices with higher $\omega_T$ and $\omega_{max}$.

$g_m$, $C_{gs}$ and $C_{gd}$ are all bias dependent parameters, so are $\omega_T$ and $\omega_{max}$. Fig. 2. 9 from [2. 4] shows this dependence. Thus, bias point of the active devices needs to be carefully chosen to optimize both the circuit performance and power consumption.

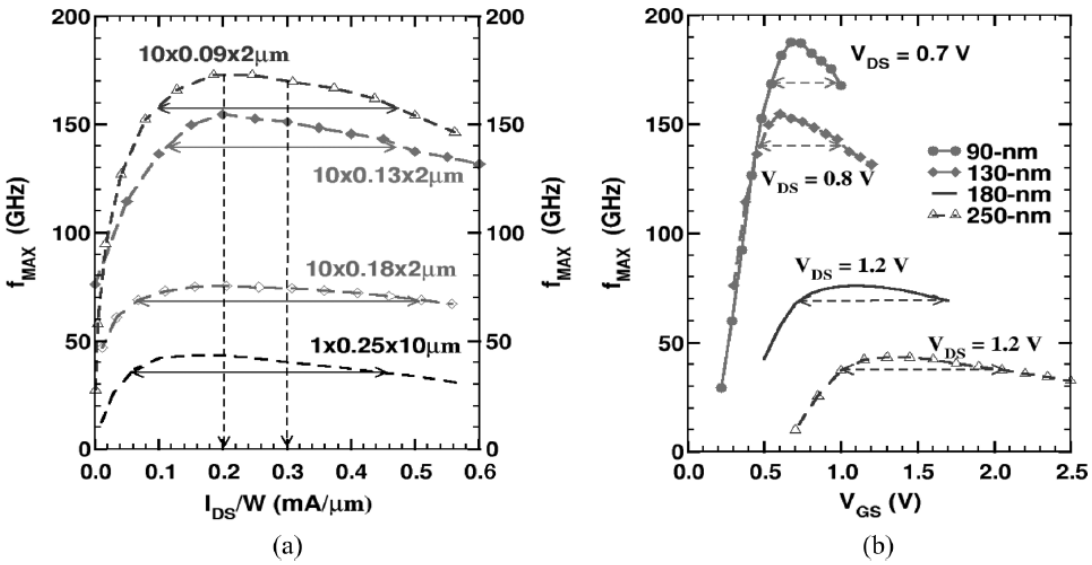



Fig. 2. 9 (a) Measured $f_{max}$ as a function of drain current per unit gate width for nMOSFETs across different technology nodes. (b) Measured $f_{max}$ versus $V_{GS}$ across different technology nodes [2. 4]

## 2. 6. Difference between Narrow and Wideband Circuits Design

In telecommunications, bandwidth that is low, relative to the transmission medium, is called narrowband. Taking IEEE 802.11a WLAN [3. 1] as a narrowband example, it occupies frequency bands 5.15-5.25 GHz, 5.25-5.35 GHz, and 5.725-5.825 GHz, the carrier frequency (~5 GHz) is 50 times that of the frequency bandwidth.

Wideband radio technologies also attract lots of attention, ultra-wideband (UWB) is one among them, it is used to refer to any radio technology having bandwidth larger than 500 MHz or 25% of the center frequency, according to Federal Communication Commission (FCC).

Due to the difference in the targeted bandwidth, different design considerations result for narrowband and wideband LNA design.

In narrowband design, parasitic capacitors are usually tolerable due to the application of LC tuned networks, whereas in the wideband case, parasitic capacitors greatly reduce the achievable bandwidth, and LC resonating tank usually cannot be

applied owing to its narrowband nature. Wideband circuits have the limitation of gain-bandwidth product (GBW), while in narrowband circuits, gain can be made large with small power consumption, because a parallel tuned LC tank can provide large load impedance at operating frequency.

Careful design needs to be conducted to optimize narrowband LNA at operating frequency, for gain, NF, impedance matching, and so on. Inductor/capacitor components tuning may be required in the tuned networks in the presence of process variation. Similarly, careful design is needed for wideband LNA to maximize the GBW, maintain good NF and impedance matching over the entire bandwidth.

## REFERENCES


[2. 1] B. Razavi, *RF Microelectronics*, Prentice Hall, 1998.

[2. 2] D. M. Pozar, *Microwave Engineering*, 2nd ed., Wiley, 1998.

[2. 3] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[2. 4] T. Yao, M. Q. Gordon, K. K. W. Tang, K. H. K. Yau, M. Yang, P. Cshvan, and S. P. Voinigescu, "Algorithmic design of CMOS LNAs and PAs for 60-GHz radio," *IEEE J. Solid-State Circuits*, Vol. 42, No. 5, pp. 1044-1057, May 2007.

# Chapter 3 Narrowband LNA Design

## 3. 1. Introduction

In a receiver chain, LNA is usually the first active signal-processing block after the antenna. The amplitude of the received signal at the input of the LNA may vary from nV (less than -130 dBm for GPS signals) to tens of mV (large interferers accompanying the signal). The LNA should be capable of amplifying all these signals without causing any significant distortion. Furthermore, LNA noise characteristic dominates the overall noise performance of the receiver. This requires that very little noise from the LNA is introduced to the entire receiver. Another major requirement of the LNA is to provide a large gain to suppress the noise of subsequent blocks.

LNAs are usually preceded by passive filters for out-of-band rejections and channel selection. The transfer function of such filters is usually a function of their termination impedance. This imposes the requirement of certain input impedance, such as 50 Ω, on the LNA. Input impedance matching consideration is particularly important because it usually determines the LNA topology.

In this chapter, the commonly used LNA topologies for narrowband application are reviewed and compared. Noise optimization and linearity improvement technique will be addressed.

The noise optimization procedure for the inductive source degeneration LNA is presented, taking into account lossy on-chip inductor's noise contribution. The linearity improvement technique is proposed and analyzed using $2^{nd}$ order intermodulation (IM2) injection, which is applied in a narrowband LNA for radio-frequency identification (RFID) reader working at 900 MHz.

## 3. 2. Passive Components

Passive components used in the RF circuits, including on-chip inductors, T-coils, and switched-capacitor arrays will be discussed in this section. Design consideration and optimization will be presented.

### 3.2.1 On-Chip Inductor

The loss of the on-chip inductors is usually caused not only by the resistance loss of the metal layer but also by the loss in the silicon substrate. Modern CMOS processes usually consists of a heavily doped epitaxy layer which is highly conductive, the eddy current induced by the magnetic field of the inductor onto the substrate directly increases the loss of the inductor.

A widely used inductor model [3. 2] is depicted in Fig. 3. 1, where $R_s$ is the series resistance of the inductor $L_s$, $C_s$ represents the capacitance between each turn of the inductor, $C_{ox}$ is the capacitance between the inductor and the substrate, $R_{Si}$ and $C_{Si}$ model the lossy silicon substrate.

#### 3.2.1.1 *Q*-Factor

At low frequencies, the impedance of $C_{ox}$ is very high, and the inductor is isolated from the lossy substrate. Thus, the loss is mainly introduced by the metal layer. As the frequency of interests increases, the impedance of $C_{ox}$ decreases, and the resistive loss due to the lossy substrate becomes important. At a high frequency, $C_{ox}$ is virtually shorted and the substrate loss dominates. As a result, the loss of the on-chip inductor increases with the frequency, instead of remaining constant.

To gain more direct observation of the inductor quality, the quality factor ($Q$-factor) is fundamentally defined as

$$Q = \omega \frac{energy\ stored}{average\ power\ dissipated} \tag{3. 1}$$

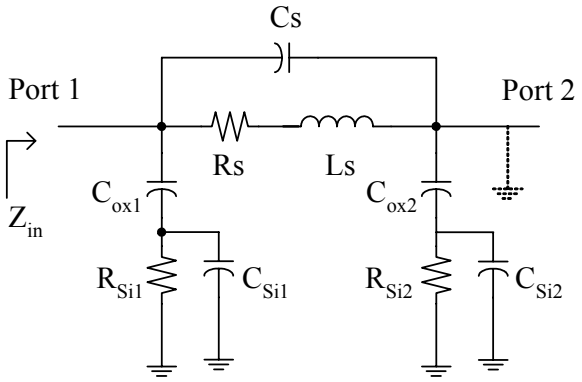



Fig. 3. 1 A complete inductor model

If the inductor is modeled as an ideal inductor in series with a parasitic resistor, i.e., a simplified model with $R_s$ and $L_s$ only in Fig. 3. 1, the $Q$-factor can be calculated as

$$Q = \omega \frac{(1/2) \cdot L_s I_{pk}^2}{(1/2) \cdot I_{pk}^2 R_s} = \frac{\omega L_s}{R_s} = \frac{\mathrm{Im}(Z_{in})}{\mathrm{Re}(Z_{in})} \tag{3. 2}$$

where $I_{pk}$ is the peak current flowing through the inductor, and $Z_{in}$ is the input impedance looking at port 1 or port 2 while the other end of the inductor is connected to ground.

(3. 2) is generalized and becomes the traditional approach of $Q$ extraction for all passive components [3. 3]:

$$Q = \frac{\mathrm{Im}(Z_{in})}{\mathrm{Re}(Z_{in})} = \frac{-\mathrm{Im}(Y_{in})}{\mathrm{Re}(Y_{in})} \tag{3. 3}$$

One of the critical limitations when integrating RF circuits on-chip is the lack of high $Q$ on-chip inductors. Unlike resistors and capacitors, whose values are well estimated except for the process variations (around 10%), on-chip inductor's $Q$ is still hard to predict accurately. $Q$ of around 6 ~ 8 is a typical value for on-chip inductors.

#### 3.2.1.2 Inductance

The inductance value of a single metal layer inductor is well estimated by the Greenhouse's formula [3. 4]. A simple estimate of the inductance can also be found to be [3. 5]:

$$L = \mu_0 n^2 r \tag{3. 4}$$

where $\mu_0$ is the permeability of the free space, $n$ is the number of turns, and $r$ is the radius of the inductor. To have a more accurate estimation of the inductance value, the inductor structure can be analyzed with a 3-D electromagnetic simulator like SONNET EM [3. 6]. However, it is usually time-consuming especially when the structure needs to be optimized in terms of metal spacing, metal width and total area.

As a compromise, other programs, such as ASITIC [3. 7], Agilent Momentum [3. 8], which run much faster, are employed in design and optimization of the inductors in this thesis. ASITIC is an interactive CAD tool that aids RF/microwave and high speed digital engineers to analyze, model, and optimize passive and interconnect metal structures residing on a lossy conductive substrate, it runs very fast and can predict inductance accurately. Momentum works together with Advanced Design System (ADS) to compute S-, Y-, and Z-parameters of general planar circuits, its simulation speed is slower than ASITIC, while it can predict the *Q*-factor more accurately. So ASITIC is usually used for initial inductor design, after the inductor structure is roughly fixed, Momentum is used to predict the inductor performance.

#### 3.2.1.3 Inductor Design

On-chip inductors are usually implemented using the topmost metal layer(s). In Taiwan Semiconductor Manufacturing Company (TSMC) 0.18-μm, 6-metal-layer CMOS process, the top metal layer M6 is 2-μm thick with sheet resistance 18-mΩ/sq, while the lower metal layers is 0.53-μm thick with sheet resistance 80-mΩ/sq, in

addition, the separation between the topmost layer and the lossy substrate is the largest, thus, best $Q$ of the inductor results.

Fig. 3. 2 shows a 3-turn square planar inductor. The main design parameters for a square inductor include the number of turns $n$, metal width $w$, and out dimension $L$. The spacing between turns $s$ is usually set to be the minimum space allowed, which increases the positive magnetic coupling between windings, causing an increase in the inductance and the $Q$-factor for a given layout area.

As the number of turns $n$ increases while fixing the total metal length, the spacing between opposite sides of the spiral shrinks, causing a drop in inductance because of negative mutual coupling between opposite sides. Thus, more metal is required to maintain a constant inductance value. It means the inner turns contribute little to the inductance but suffer from an increase in resistance. So the number of turns cannot be too large and it is necessary to use a hollow coil. The inner hole is usually at least 70-μm in dimension.

The $Q$-factor as a function of frequency for a 3-nH inductor is shown in Fig. 3. 3. There is a frequency for the peak $Q$-factor. At low frequency, the loss is almost constant while the imaginary part for $Z_{in}$ of the inductor in (3. 3) increases linearly with frequency, so the $Q$-factor also increases; at high frequency, the loss increases faster than the imaginary part of $Z_{in}$, $Q$-factor then begins to drop; at very high frequency, the parasitic capacitors in Fig. 3. 1 resonate with the inductor, and the spiral no longer works as an inductor. This frequency is called self-resonant frequency. As a rule of thumb, self-resonant frequency of the on-chip inductor should be at least 2 times the inductor operating frequency.

It is necessary to make the inductor peak-$Q$ frequency locate at inductor operating frequency to maximize the $Q$-factor. Metal width of the inductor will shift the

frequency where peak *Q*-factor is reached. Wider metal will lower low-frequency resistive loss but suffer from larger parasitic capacitor and magnetic loss at high-frequency. So a wider metal inductor will have peak *Q*-factor at lower frequency. By choosing the appropriate metal width, peak *Q* can be obtained at specific frequency.

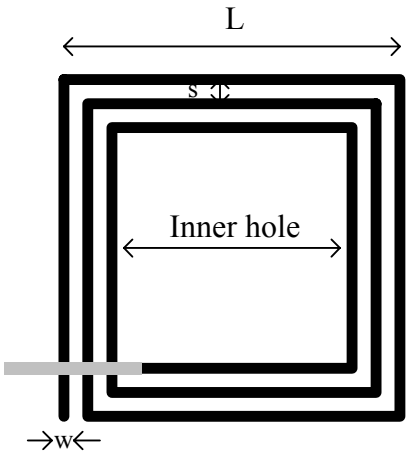



Fig. 3. 2 Suqare planar inducotor

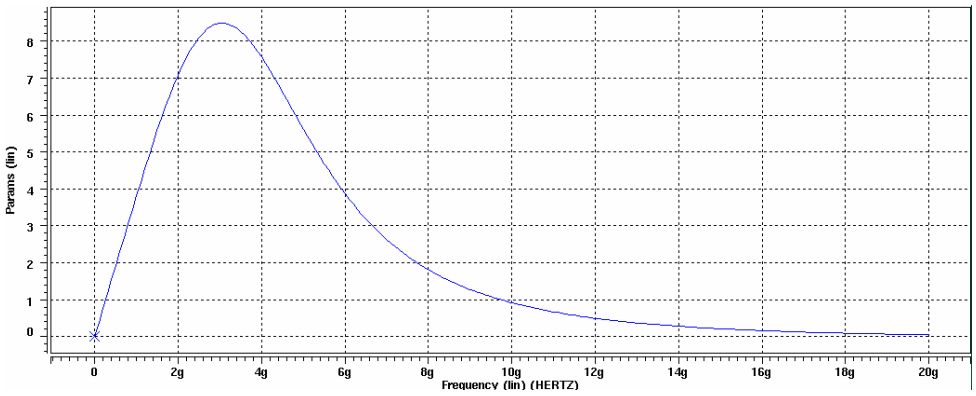
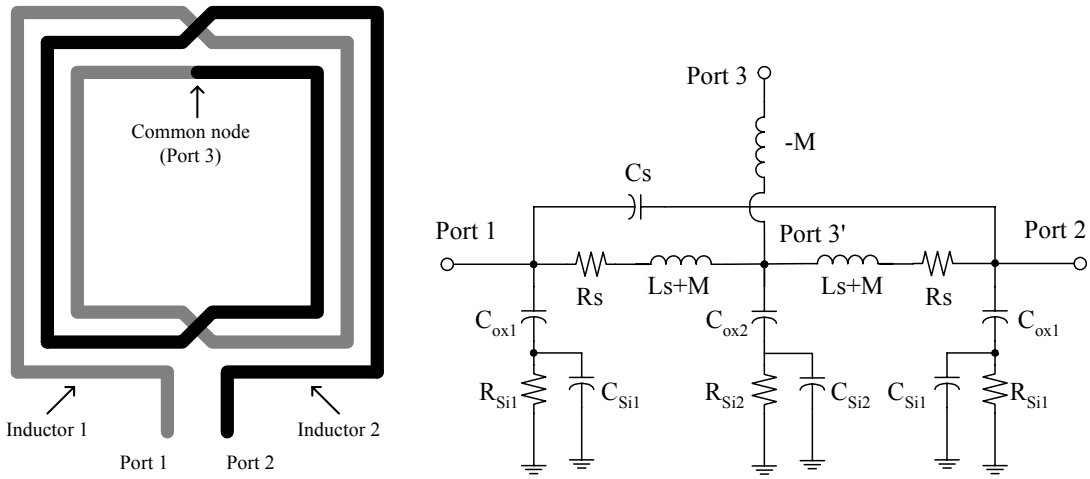



Fig. 3. 3 Simulated *Q*-factor of a 3-nH inductor

### 3.2.1.4 Center-Tapped Inductor [3. 9]

A symmetric inductor for differential circuits, called center-tapped inductor, is shown in Fig. 3. 4, together with its equivalent circuit model. One advantage of a center-tapped inductor is that the two separate spirals are replaced by a single coil which has both electrical and geometric symmetry. Thus, the chip area is greatly

reduced. The symmetry is important when locating the common-mode (a convenient bias point for active circuits), which separates the spiral into two inductances that have identical substrate parasitics at ports 1 and 2.




(a) (b)

Fig. 3. 4 Center-tapped inductor (a) layout, and (b) equivalent circuit model

Center-tapped inductor also improves the *Q*-factor by differential excitation. As shown in Fig. 3. 4 (b), due to virtual-ground nature of port 3', parasitic $C_{ox2}$, $R_{Si2}$, $C_{Si2}$ do not have any effect on inductor operation, thus the substrate loss is less severe than single-ended excitation. –M caused by mutual coupling between two inductors does not affect differential operation, either.

### 3.2.2 Switched-Capacitor Array (SCA)

Due to process variation in the inductors and capacitors, there's usually a frequency shift between the actual resonant frequency and the intended operating frequency for LC resonant tank circuits, and frequency tuning is desired.

Switched-capacitor array (SCA) is one of the candidates to be used in LC resonant tanks to tune the center frequency. The structure of the SCA is shown in Fig. 3. 5. Capacitors $C_0$, $C_1$, … $C_n$ can be binary-weighted, i.e., $C_n = C_0*2^n$. Transistors M0,

M1, … Mn work as switches and are also binary-weighted as capacitors, when gate bias $V_{bi}$ (*i* represents the *i*th bit) is high, transistor exhibits small turn-on resistance and the connected capacitor $C_i$ is enabled; when gate bias $V_{bi}$ is 0, transistor Mi is shut down and $C_i$ is opened in the circuit.

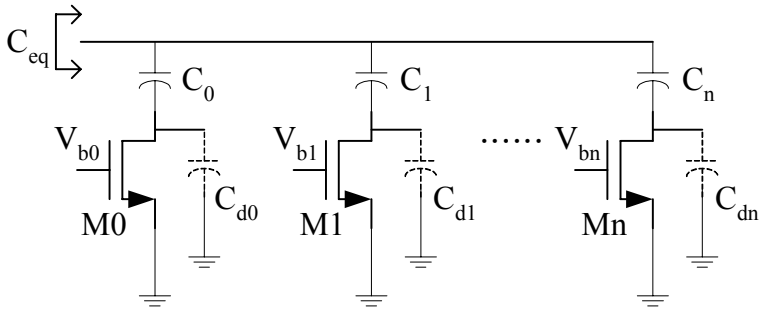



Fig. 3. 5 Schematic of switched-capacitor array

In real case, when the switch is opened, there is parasitic capacitor $C_{di}$ coming from Mi's drain parasitic capacitor and bottom-plate parasitic capacitor in $C_i$. Then $C_i$ is not completely opened, there is still finite equivalent capacitance $C_{eq}$ in an off-state ($C_{eq,off} = C_i C_{di} / (C_i + C_{di})$). If the switch transistors are small, the problem of none-zero turn-off capacitance is relaxed, at the expense of lager turn-on resistance $R_{on}$ ($R_{on} = 1/\left( \mu C_{ox} \frac{W}{L} (V_{GS} - V_T) \right)$) in the switches for on-state and degraded *Q*-factor in the SCA ($Q = 1/(\omega C_i R_{on,i})$). As seen, *Q*-factor degrades as frequency increases, thus, the tradeoff between *Q*-factor and nonzero turn-off capacitor is more severe at high-frequencies.

To maximize the tuning range of the SCA (maximize $C_{eq,on}/C_{eq,off}$) while not sacrificing the *Q*-factor, donut transistors [3. 11] are used in realizing the switches, as shown in Fig. 3. 6 (a). The target is to minimize the drain area for a given gate length. The drain area of the donut transistor is shared by the gates in all four sides, thus the parasitic drain capacitance is reduced.

In donut transistor, drain area is proportional to $w^2$, where $w$ is transistor width, so (drain area) / (transistor width) $\propto w$, it means that each donut transistor cell should be as small as possible to minimize drain area per unit width. For the case when donut transistor has $w$ = 0.72-μm * 4, the drain area is 0.3-μm$^2$, while for traditional 2-finger transistor with $w$ = 1.44-μm * 2 as shown in Fig. 3. 6 (b), the drain area is 0.78-μm$^2$.

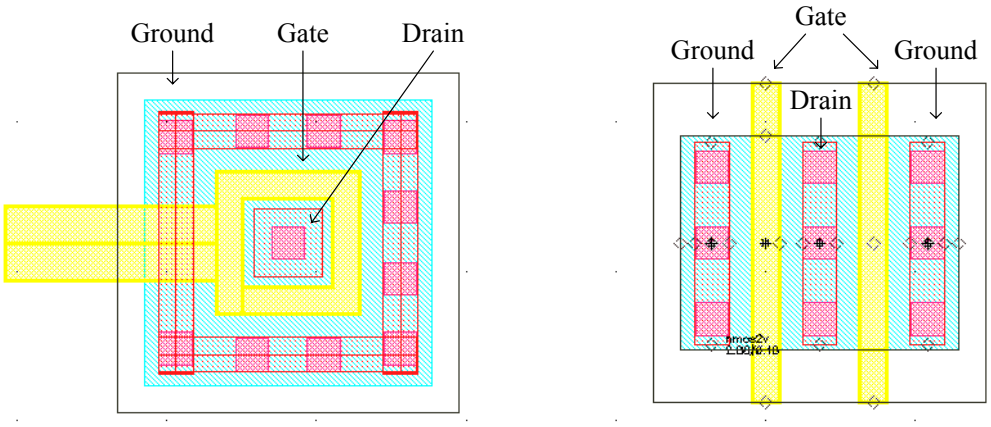



Fig. 3. 6 (a) donut transistor, and (b) traditional transistor

*Q*-factor of SCA should be larger than that of the inductor in an LC tank in order not to dominate the overall *Q* of the tank, usually SCA's *Q* should be around 20-30.

In SCA on-state, the capacitor is in series with a resistor formed by a MOSFET switch that is nonlinear. If SCA is used in the parallel LC tank as amplifier loading, linearity of the LNA is usually affected by this nonlinear loading. *Q*-factor of SCA has to be large enough so that $R_{on}$ of the switch is negligible to maintain a good linearity. This in turn limits the tuning range of SCA.

## 3. 3. Inductive Source Degeneration LNA Design

Wireless receivers in many standards involve narrowband LNA design, such as GSM [3. 12], Bluetooth [3. 13], WLAN receiver [3. 14], and RFID reader. Due to the narrowband nature, LC resonant tank is available in the design. Gain, input impedance matching, and NF only need to be optimized at the center frequency.

### 3.3.1 Noise Model for MOSFET

Circuit components, including active devices and resistors, will add extra noise to the received signal, thus deteriorate the SNR at circuit output.

A high-frequency thermal noise model for a short-channel metal-oxide-semiconductor-field-effect-transistor (MOSFET) device is depicted in Fig. 3. 7 [3. 15], where $R_{gpar}$ and $e_{npar}$ represent the parasitic gate resistance and its corresponding noise, $i_{ng}$ is the induced gate noise, $g_g$ is caused by the distributed channel resistance, and $R_s$ is the source resistance, $i_{nd}$ is the drain noise.

The mean-square noise current at the drain terminal due to thermal noise is the major noise source in MOSFET, which can be expressed as:

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \tag{3. 5}$$

where $g_{d0}$ is the drain-source conductance at zero $V_{DS}$, its value equals transconductance $g_m$ for long channel device, and $\Upsilon$ is an enhancement factor and is 2/3 for long-channel MOSFETs, and is around 1.2 for 0.18-μm process, when devices are biased at the saturation regime [3. 15], [3. 16].

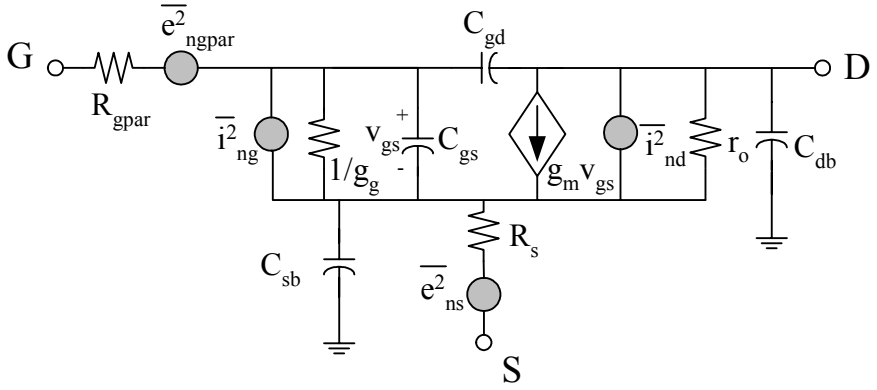



Fig. 3. 7 Equivalent small-signal noise model for the MOSFET

The induced gate noise is caused by fluctuating channel potential coupling capacitively into the gate terminal, inducing a noise gate current:

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f = 4kT\delta g_{d0}(\frac{2\pi f C_{gs}}{\sqrt{5} g_m / \alpha})^2 \Delta f \quad (3.6)$$

where $$g_g = \frac{\omega^2 C_{gs}^2}{5 g_{d0}}, \quad and \quad \alpha = \frac{g_m}{g_{d0}} \quad (3.7)$$

δ is gate noise coefficient and is typically assumed to be 2ϒ.

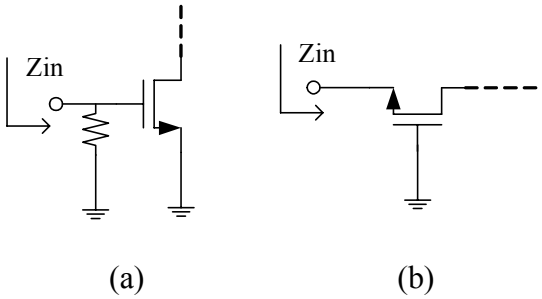
### 3.3.2 LNA Topologies

The general topology of any LNA can be broken down into two stages: an input matching network and the amplifier itself.

#### 3.3.2.1 Resistive Termination

A simple approach to providing a 50-Ω termination is to put a 50-Ω resistor across the input terminals of a common-source amplifier, or use a common-gate configuration, as shown in Fig. 3. 8. In case (b), ignoring the gate-source parasitic capacitor $C_{gs}$, the input impedance is

$$Z_{in} = 1/g_m \quad (3.8)$$




(a) (b)

Fig. 3. 8 (a) 50-Ω termination input stage; (b) common-gate input stage

Such matching is limited to low-gigahertz range when the impedance from input parasitic capacitor of the active device is significantly larger than 50-Ω. Despite the simple implementation, matching topology in Fig. 3. 8 suffers NF degradation from the presence of noisy resistance in the signal path (including channel resistance, as in case (b) of Fig. 3. 8). The minimum NF is 3 dB and 2.2 dB for case (a) and (b)

respectively, where only the noise from the matching part is considered. If noise contributed from other parts, such as loading, cascode device, is considered, NF will be even larger.

#### 3.3.2.2 Shunt-Series Feedback

The shunt-series amplifier, as shown in Fig. 3. 9, is another circuit that provides impedance matching. If ignoring $C_{gs}$ of the active device, the input impedance Zin can be written as:

$$Z_{in} = \frac{R_F}{1 + A_v} \tag{3. 9}$$

where $A_v$ is the voltage gain of the amplifier. Therefore, the shunt resistor $R_F$ is $(1+A_v)$ times larger than 50-Ω if the amplifier is impedance-matched, and its noise contribution is reduced by $(1+A_v)$ times as compared to Fig. 3. 8 (a). However, resistive component $R_F$ is still used for impedance matching, as a consequence, the overall amplifier's NF, while usually much better than that of Fig. 3. 8, is still not the minimum possible.

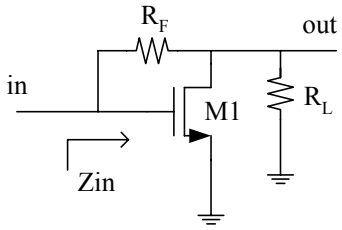



Fig. 3. 9 Shunt-series amplifier (biasing not shown)

Shunt-series feedback impedance matching finds applications in both wideband and narrowband circuits. If narrowband filtering is implemented in the feedback loop, e.g., $R_L$ is replaced by parallel LC tank, narrowband filtering is also achieved in input impedance matching. Despite the relatively large NF, shunt-series feedback is found in many LNA applications due to its simple topology and broadband capability.

### 3.3.2.3 Inductive Source Degeneration

A better method for narrowband low-noise application is to employ inductive source degeneration [3. 15], as shown in Fig. 3. 10. If we simplify the MOSFET model to include only a transconductance $g_{m1}$ and a gate-source capacitance $C_{gs}$, it can be shown that the input impedance has the following form:

$$Z_{in} = s(L_s + L_g) + \frac{1}{sC_{gs}} + \frac{g_{m1}}{C_{gs}} L_s \approx s(L_s + L_g) + \frac{1}{sC_{gs}} + \omega_T L_s \tag{3. 10}$$

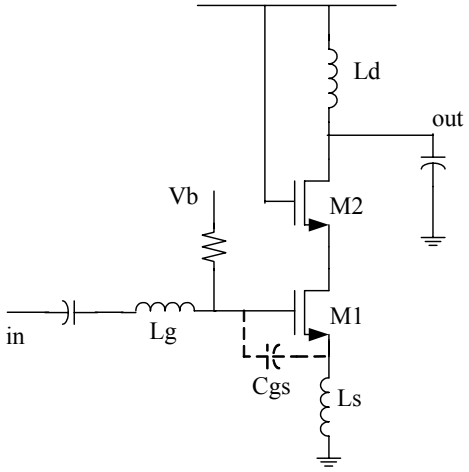



Fig. 3. 10 Schematic of inductive source degeneration LNA

The matching conditions occur when

$$\begin{cases} \omega_0^2 C_{gs}(L_g + L_s) = 1 \\ R_s = \dfrac{g_{m1}}{C_{gs}} L_s \end{cases} \tag{3. 11}$$

where $\omega_0$ is the operating frequency. The 1st equation makes sure that imaginary parts in (3. 10) cancel out ($L_g$ and $L_s$ resonate with $C_{gs}$ at operating frequency), the 2nd equation guarantees 50-Ω real input impedance. The matching is achieved without adding large noise source (resistor) in the signal path, so its NF is much smaller than the resistive termination or common-gate topology. The matching only happens at the resonance of the input network, so it can only provide a narrowband matching.

### 3.3.3 Noise Analysis of Inductive Source Degeneration LNA

#### 3.3.3.1 Noise Analysis of Input Active Device with Ideal Inductors [3. 17]

Noise of input device M1 in Fig. 3. 10 is usually the dominant part in inductive degeneration LNA because it interferes with the signal at the very beginning of the signal processing chain.

To analyze input device noise contribution, MOSFET noise model in Fig. 3. 7 will be used. To get a clearer picture of the main noise transfer and simplify the calculation, noise model will be simplified: $R_{gpar}$ and the associated $e_{ngpar}$ can be made negligible with proper layout; $g_g$ can be assumed to be negligible when operating frequency $\omega$ is much lower than transistor cut-off frequency $\omega_T$; $C_{sb}$, $C_{gd}$, and $C_{db}$ are removed to simplify calculation; $r_o$ is also ignored because loading impedance is usually much smaller than $r_o$. Thus, the main noise sources from a MOSFET are drain noise and induced gate noise.

For an amplifier as shown in Fig. 3. 11, the MOSFET is connected to a source impedance $Z_g$, and degenerated by $Z_{deg}$. The output noise current $i_{ndg}$ is found to be:

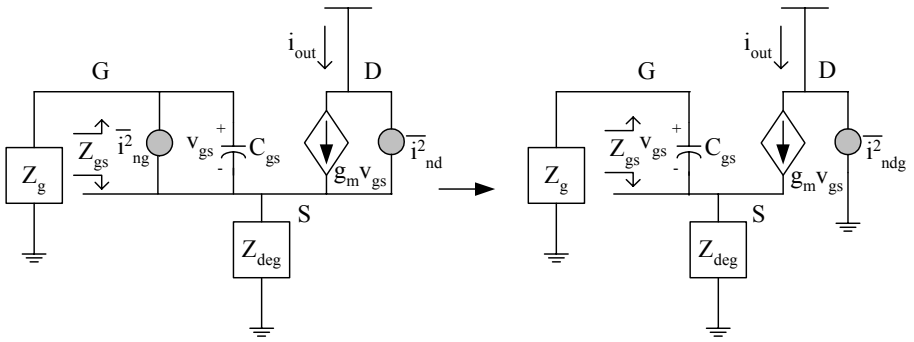



Fig. 3. 11 Noise calculation for a degenerated common-source amplifier

$$i_{out} = \overline{i_{ndg}} = \eta \overline{i_{nd}} + g_m Z_{gs} \overline{i_{ng}} \tag{3. 12}$$

where

$$\begin{cases} Z_{gs} = \dfrac{1}{sC_{gs}} // \dfrac{Z_{\deg} + Z_g}{1 + g_m Z_{\deg}} \\ \eta = 1 - (\dfrac{g_m Z_{\deg}}{Z_{\deg} + Z_g}) Z_{gs} \end{cases} \tag{3. 13}$$

To find noise variance:

$$\begin{aligned} \overline{i_{ndg}^2} &= \overline{i_{ndg}^* i_{ndg}} = \overline{(\eta^* i_{nd}^* + g_m Z_{gs}^* i_{ng}^*)(\eta i_{nd} + g_m Z_{gs} i_{ng})} \\ &= |\eta|^2 \overline{i_{nd}^* i_{nd}} + \overline{i_{nd}^* i_{ng}} g_m \eta^* Z_{gs} + \overline{i_{nd} i_{ng}^*} g_m \eta Z_{gs}^* + \overline{i_{ng}^* i_{ng}} |g_m Z_{gs}|^2 \\ &= |\eta|^2 \overline{i_{nd}^2} + 2\,\mathrm{Re}\{\overline{i_{nd}^* i_{ng}} g_m \eta^* Z_{gs}\} + \overline{i_{ng}^2} |g_m Z_{gs}|^2 \\ &= |\eta|^2 \overline{i_{nd}^2} + 2\,\mathrm{Re}\{\frac{\overline{i_{nd}^* i_{ng}}}{\sqrt{\overline{i_{nd}^2}\,\overline{i_{ng}^2}}} \sqrt{\overline{i_{nd}^2}\,\overline{i_{ng}^2}}\, g_m \eta^* Z_{gs}\} + \overline{i_{ng}^2} |g_m Z_{gs}|^2 \end{aligned} \tag{3. 14}$$

After defining the correlation coefficient *c* between $i_{ng}$ and $i_{nd}$,

$$c = \frac{\overline{i_{nd}^* i_{ng}}}{\sqrt{\overline{i_{nd}^2}\,\overline{i_{ng}^2}}} \tag{3. 15}$$

(3. 14) becomes

$$\overline{i_{ndg}^2} = \overline{i_{nd}^2} \left[ |\eta|^2 + 2\,\mathrm{Re}\left( c \sqrt{\frac{\overline{i_{ng}^2}}{\overline{i_{nd}^2}}} g_m \eta^* Z_{gs} \right) + \frac{\overline{i_{ng}^2}}{\overline{i_{nd}^2}} g_m^2 |Z_{gs}|^2 \right] \tag{3. 16}$$

If parameters $Z_{gsw}$ and $\chi_d$ are defined to be:

$$Z_{gsw} = \omega_0 C_{gs} Z_{gs}, \qquad \chi_d = \frac{g_m}{g_{d0}} \sqrt{\frac{\delta}{5\gamma}} \tag{3. 17}$$

combining with (3. 5) and (3. 6), (3. 16) becomes

$$\overline{i_{ndg}^2} = \overline{i_{nd}^2} \left[ |\eta|^2 + 2\,\mathrm{Re}\left( c \chi_d \eta^* Z_{gsw} \right) + \chi_d^2 |Z_{gsw}|^2 \right] \tag{3. 18}$$

The correlation coefficient *c* is purely imaginary, and *c*=-0.395j for long-channel devices [3. 15]. For 0.18-μm CMOS, we will assume the following:

c=-j0.55, γ=1.2, δ/γ=2, $g_m/g_{d0}$=1/2,

In the case $Z_g$ only has purely resistive component $R_s$, and $Z_{deg}$=0, from (3. 13),

$$\eta = 1, \quad Z_{gsw} = \omega_0 C_{gs}\left( R_s // \frac{1}{j\omega_0 C_{gs}} \right) = \frac{\omega_0 C_{gs} R_s}{1 + j\omega_0 C_{gs} R_s} \tag{3. 19}$$

Combining (3. 18) and (3. 19), we can see that, for $\omega_0 << 1/(R_s C_{gs})$,

$$Z_{gsw} \approx \omega_0 C_{gs} R_s \quad \Rightarrow \quad \frac{\overline{i_{ndg}^2}}{\Delta f} \approx \frac{\overline{i_{nd}^2}}{\Delta f}\left(1 + \chi_d^2 (\omega_0 C_{gs} R_s)^2\right) \tag{3. 20}$$

For $\omega_0 >> 1/(R_s C_{gs})$,

$$Z_{gsw} \approx 1/j \quad \Rightarrow \quad \frac{\overline{i_{ndg}^2}}{\Delta f} \approx \frac{\overline{i_{nd}^2}}{\Delta f}\left(1 - 2|c|\chi_d + \chi_d^2\right) \tag{3. 21}$$

The terms other than "1" in the brackets in (3. 20) and (3. 21) are inferred to be induced gate noise, which are much smaller than 1. Therefore, the gate noise has little effect on the common-source amplifier when source impedance is purely resistive, and drain thermal noise is the chief issue of concern.

In the case of inductive degenerated LNA, the schematic in Fig. 3. 11 is redrawn in Fig. 3. 12.

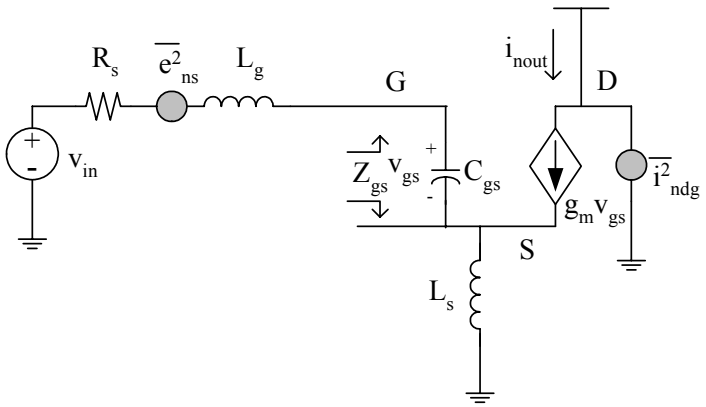



Fig. 3. 12 Noise calculation for inductive source degeneration amplifier

$L_g$ and $L_s$ are used for input impedance matching and their values need to satisfy (3. 11). We can get the following results:

$$\begin{cases} Z_{gs} = \dfrac{1}{SC_{gs}} // \dfrac{Z_{\deg} + Z_g}{1 + g_m Z_{\deg}} = \dfrac{1}{j\omega_0 C_{gs}} // \dfrac{j\omega_0(L_s + L_g) + R_s}{1 + g_m j\omega_0 L_s} = \dfrac{j\omega_0(L_s + L_g) + R_s}{j\omega_0(g_m L_s + R_s C_{gs})} \\ \eta = 1 - \left( \dfrac{g_m Z_{\deg}}{Z_{\deg} + Z_g} \right) Z_{gs} = 1 - \dfrac{g_m j\omega_0 L_s}{j\omega_0(L_s + L_g) + R_s} \dfrac{j\omega_0(L_s + L_g) + R_s}{j\omega_0(g_m L_s + R_s C_{gs})} \\ \quad = 1 - \dfrac{(g_m / C_{gs}) L_s}{(g_m / C_{gs}) L_s + R_s} = 1 - \dfrac{R_s}{R_s + R_s} = \dfrac{1}{2} \end{cases} \tag{3. 22}$$

At operating frequency, $L_g$ and $L_s$ resonate with $C_{gs}$, the *Q*-factor of the input matching network is found to be:

$$Q = \frac{1}{\omega_0 C_{gs} 2R_s} = \frac{\omega_0(L_g + L_s)}{2R_s} \tag{3. 23}$$

According to (3. 17),

$$\begin{aligned} Z_{gsw} &= \omega_0 C_{gs} \frac{j\omega_0(L_s + L_g) + R_s}{j\omega_0(g_m L_s + R_s C_{gs})} = \frac{j\omega_0^2 C_{gs}(L_s + L_g) + \omega_0 C_{gs} R_s}{j\omega_0(g_m L_s + R_s C_{gs})} \\ &= \frac{j + 1/(2Q)}{j\omega_0 C_{gs}\left[(g_m / C_{gs}) L_s + R_s\right]} = \frac{j + 1/(2Q)}{j\omega_0 C_{gs} 2R_s} = \frac{j + 1/(2Q)}{j/Q} = \frac{1}{2}(2Q - j) \end{aligned} \tag{3. 24}$$

Plug in values (3. 22) and (3.24) to (3. 18), we have:

$$\begin{aligned} \frac{\overline{i_{ndg}^2}}{\Delta f} &= \frac{\overline{i_{nd}^2}}{\Delta f}\left\{ \frac{1}{4} + 2\,\mathrm{Re}\left[ -j|c|\chi_d \frac{1}{4}(2Q - j) \right] + \chi_d^2 \frac{1}{4}|2Q - j|^2 \right\} \\ &= \frac{\overline{i_{nd}^2}}{\Delta f}\frac{1}{4}\left[1 - 2|c|\chi_d + \chi_d^2(4Q^2 + 1)\right] \end{aligned}$$

To calculate the NF of the input active device of the inductive degeneration LNA, we need to find out the output noise due to source resistor. As shown in Fig. 3. 12, due to noise source $e_{ns}$,

$$v_{gs} = \frac{\overline{e_{ns}}}{R_s + Z_{in}}\left( \frac{1}{j\omega_0 C_{gs}} \right) = \left( \frac{Q}{j} \right)\overline{e_{ns}} \quad \Rightarrow \quad \overline{i_{nout}^2} = (g_m Q)^2 \overline{e_{ns}^2} \tag{3. 25}$$

$$F = 1 + \frac{\overline{i_{ndg}^2}}{(g_m Q)^2 \overline{e_{ns}^2}} = 1 + \frac{4kT\gamma g_{d0}(1/4)\left[1 - 2|c|\chi_d + \chi_d^2(4Q^2+1)\right]}{(g_m Q)^2 4kTR_s}$$
$$= 1 + \left(\frac{1}{g_m Q R_s}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)\frac{1}{4Q}(1 - 2|c|\chi_d + \chi_d^2(4Q^2+1)) \quad (3.26)$$
$$= 1 + \left(\frac{\omega_0}{\omega_T}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)\frac{1}{2Q}(1 - 2|c|\chi_d + \chi_d^2(4Q^2+1))$$

From (3. 26), it is obvious the advantage of higher $\omega_T$ device in inductive source degeneration LNA implementation. It is also seen that there are two competing factors in $Q$ determining the noise factor $F$: $1/(2Q)$ accounts for the effect of the gain $v_{gs}/v_{in}$ (equals to $Q$) in the input matching network, when the gain before the MOSFET is larger, noise contributed by the MOSFET will be suppressed; $\chi_d^2(4Q^2+1)$ means that when $Q$-factor is larger, the induced gate noise will also be larger. Therefore, there exists an optimal $Q$ value for NF.

The optimal $Q$ value can be found by:

$$\frac{d}{dQ}\left[\frac{1}{2Q}(1 - 2|c|\chi_d + \chi_d^2(4Q^2+1))\right] = 0 \quad (3.27)$$

and it is solved to be $$Q = \frac{1}{2\chi_d}\sqrt{1 - 2|c|\chi_d + \chi_d^2} \quad (3.28)$$

In 0.18-μm CMOS process, assume that c = -0.55j, $\chi_d = 0.32$, the optimal $Q$ can be calculated to be 1.35, relating to the $Q$ definition in (3. 23), at $\omega_0$ = 5.25 GHz * 2π, $C_{gs}$ is calculated to be 225 fF, thus the input transistor size is determined with the fact that $C_{gs} \approx \frac{2}{3}WLC_{ox}$, where $W$ and $L$ are the transistor width and length, and $C_{ox}$ is gate oxide capacitance per unit area.

#### 3.3.3.2 Noise Optimization with On-Chip Inductors

Due to the low *Q*-factor of the on-chip inductor, the lossy components contribute significant portion of noise to the LNA. The effect needs to be taken into account in the design optimization.

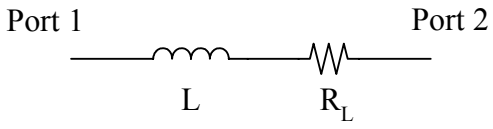



Fig. 3. 13 Simplified on-chip inductor model

At a narrow operating frequency, the usually used inductor π-model shown in Fig. 3. 1 is equivalent to a simple model composed of L and parasitic $R_L$ in series, as shown in Fig. 3. 13. All the noise from the inductor lossy components ($R_s$, $R_{Si1}$ and $R_{Si2}$ in Fig. 3. 1) can be represented by the noise of $R_L$. In inductive source degeneration LNA, the source inductor $L_s$ is usually small, around hundreds of pico henries, while gate inductor $L_g$ is in the range of a few nano henries, so the main loss is associated with $L_g$. $R_{Lg}$ and $R_{Ls}$ represent $L_g$ and $L_s$'s equivalent series resistors, respectively. Assume $L_g$ and $L_s$ have similar *Q*-factor $Q_{ind}$:

$$R_{Lg} + R_{Ls} = \frac{\omega_0 (L_g + L_s)}{Q_{ind}} = \frac{2R_s Q}{Q_{ind}} \tag{3. 29}$$

Then (3. 26) can be rewritten to be

$$\begin{aligned} F &= 1 + \frac{R_{Lg} + R_{Ls}}{R_s} + \left(\frac{\omega_0}{\omega_t}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)\frac{1}{2Q}(1 - 2|c|\chi_d + \chi_d^2(4Q^2 + 1)) \\ &= 1 + \frac{2Q}{Q_{ind}} + \left(\frac{\omega_0}{\omega_T}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)\frac{1}{2Q}(1 - 2|c|\chi_d + \chi_d^2(4Q^2 + 1)) \end{aligned} \tag{3. 30}$$

and optimal *Q* of input matching network for minimum *F* is re-calculated to be:

$$Q_{opt} = \frac{1}{2}\sqrt{\frac{\left(\frac{\omega_0}{\omega_T}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)(1 - 2|c|\chi_d + \chi_d^2)}{\frac{1}{Q_{ind}} + \left(\frac{\omega_0}{\omega_T}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)\chi_d^2}} \tag{3. 31}$$

Assume that the transistor is biased at 40μA/μm, and $\omega_T$ = 110 GHz, $\gamma$ = 1.2, $g_{d0}/g_m$ = 2, $Q_{ind}$ = 5, $c$ = -0.55j, $\chi_d$ = 0.32, and $\omega_0$ = 5.25 GHz * 2π. The minimum $F$ is achieved when $Q$ = 0.70. $Q$ is 48% smaller than the case without considering the noise contribution from the on-chip inductor, $C_{gs}$ now needs to be increased to 433 fF. $F$ is 2.48 dB (1.77). 15.8% noise is contributed by the lossy components in the on-chip inductor, whereas 27.7% noise is from the active input transistor.

#### 3.3.3.3 Noise Contribution of the Cascode Transistor

The cascode transistor M2 in Fig. 3. 10 is added as a current buffer to provide low impedance for M1's drain, which helps suppress the Miller effect of M1's $C_{gd}$, thus improves reverse isolation and stability.

The noise contribution of the cascode transistor is calculated based on the noise model in Fig. 3. 14. All parasitic capacitances of node 1, as well as junction capacitors and the gate-source capacitance of M2, are absorbed into $C_p$.

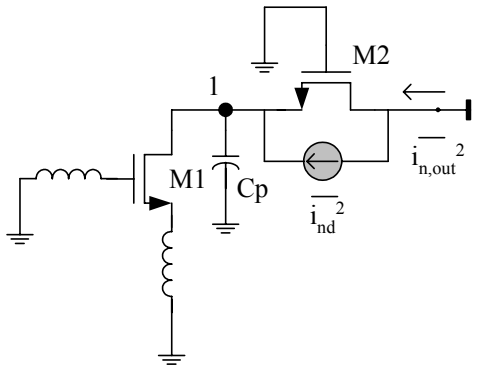



Fig. 3. 14 Noise calculation model for the cascode transistor

Noise appears at the output due to M2's thermal noise $i_{nd}$ is calculated to be

$$i_{n,out} = i_{nd} \frac{sC_p}{sC_p + g_{m2}} \tag{3. 32}$$

Refer $i_{n,out}$ to the input, the total NF of the LNA can be calculated to be

$$NF_{tot} = NF_1 + \frac{1}{4kTR_s G_m^2}\overline{i_{n,out}^2} = NF_1 + \frac{1}{4kTR_s}\frac{(\omega_0 C_{gs} 2R_s)^2}{g_{m1}^2} 4kT\gamma g_{d0,2}\frac{\omega_0^2}{\omega_0^2 + (g_{m2}/C_p)^2}$$
$$= NF_1 + 4R_s\gamma g_{d0,2}\left(\frac{\omega_0}{\omega_T}\right)^2 \frac{\omega_0^2}{\omega_0^2 + (g_{m2}/C_p)^2} \approx NF_1 + 4R_s\gamma g_{d0,2}\left(\frac{\omega_0}{\omega_T}\right)^2 \frac{C_p^2\omega_0^2}{g_{m2}^2}$$

(3. 33)

where $NF_1$ is the noise figure only considering the noise contribution from M1 and on-chip inductors as we calculated before. The approximation is made when $C_p$ is small so that $\omega_0^2 << (g_{m2}/C_p)^2$. It is seen that noise contribution from M2 is proportional to $C_p^2$ and $\omega_0^4$, $C_p$ needs to be minimized to reduce the noise contribution from M2.

If the size of M2 is smaller than that of M1, there is still finite gain at node 1 in Fig. 3. 14, miller effect is not effectively suppressed; if M2 is larger than M1, $C_p$ increases leading to worse NF and degraded frequency response. As a rule of thumb, M2 is chosen to have the same width as M1.

If we assume $g_{m2}^2/C_p^2 \approx \omega_T^2$, $\omega_T$ = 110 GHz, $\omega_0$ = 33 GHz, $g_{m2}$ = 40 mS, and $NF_1$ = 1.77 as calculated before, the 2nd term in (3. 33) is equal to 0.16, which is not negligible as compared to $NF_1$.

If M2 is equal-size with M1, (3. 33) becomes:

$$NF_{tot} \approx NF_1 + \frac{2}{Q}\frac{\gamma}{\alpha}\left(\frac{\omega_0}{\omega_T}\right)^3 \tag{3. 34}$$

where $Q$ is the $Q$-factor of the input matching stage. The optimal $Q$ for minimum $NF_{tot}$ in (3. 31) can be re-calculated as:

$$Q_{opt} = \frac{1}{2}\sqrt{\frac{\left(\frac{\omega_0}{\omega_T}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)(1-2|c|\chi_d + \chi_d^2) + 4\frac{\gamma}{\alpha}\left(\frac{\omega_0}{\omega_T}\right)^3}{\frac{1}{Q_{ind}} + \left(\frac{\omega_0}{\omega_T}\right)\gamma\left(\frac{g_{d0}}{g_m}\right)\chi_d^2}} \tag{3. 35}$$

Assume that the transistor is biased at 40μA/μm, and $\omega_T$ = 110 GHz, $\gamma$ = 1.2, $g_{d0}/g_m$ = 2, $Q_{ind}$ = 5, $c$ = -0.55j, $\chi_d$ = 0.32, and $\omega_0$ = 5.25 GHz * 2π. The minimum $NF_{tot}$ is achieved when $Q$ = 0.85. This is the optimal $Q$ after considering the cascode transistor, which is 21% larger than that calculated before. $NF_{tot}$ is now 2.87 dB (1.94).

Fig. 3. 15 shows the simulated NF of the LNA in Fig. 3. 14 for different $Q$-factor of the input matching network. The transistors are biased at 40μA/μm and the $Q$-factor of the on-chip inductors is assumed to be 5. It can be seen that the optimal condition for NF shift significantly between ideal inductor case and real inductor case, and the calculation above predicts the result well. $Q$-factor of 0.85 corresponds to a input capacitor of 370 fF, or 140-μm/0.18-μm transistor size in 0.18-μm CMOS process.

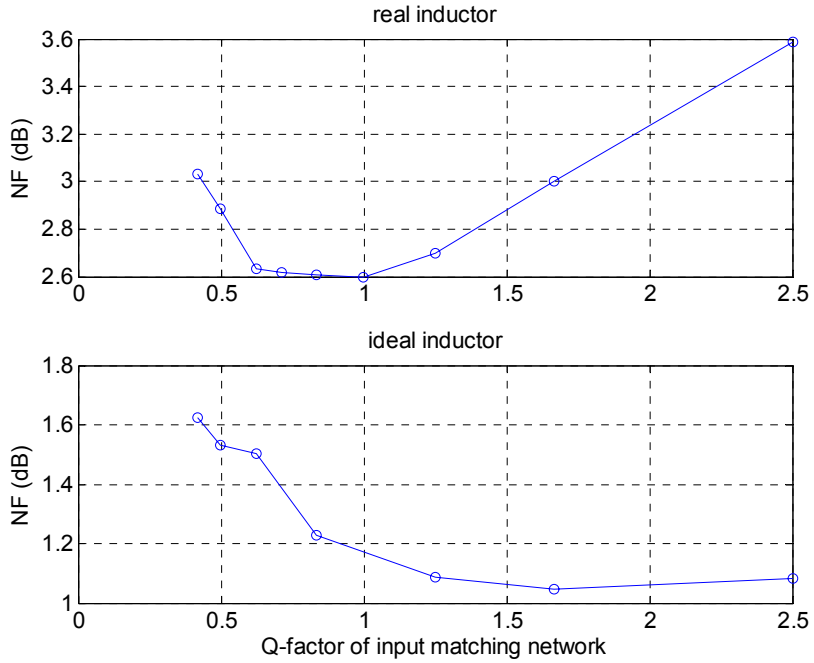



Fig. 3. 15 NF as a function of $Q$ of input matching network

### 3.3.3.4 Proposed Noise Optimization Procedure

In (3. 30), $\omega_T$ ($g_m/C_{gs}$) is bias-dependent parameters on active device, on the other hand, there usually exists power budget on the LNA. Therefore, design of minimum NF with constrained power consumption is of great importance.

The design procedure for minimizing NF with constrained power is summarized as follows (assuming 0.18-μm CMOS process):

**Step 1)** Set the bias of input transistor to a current density *J*, e.g., 40-μA/μm, which corresponds to $V_{gs}$ 700 mV. For 1-V low-voltage design in cascode topology, this is the maximum allowed gate-source voltage, which allows 300 mV $V_{ds}$ for the input transistor to comfortably work in the saturation region.

**Step 2)** Find the optimal *Q* for input matching network based on (3. 35), and then calculate the corresponding $C_{gs}$ using (3. 23) and transistor width $W_{opt}$.

**Step 3)** With $W_{opt}$ and current density *J*, if current consumption exceeds the limit *I*, a new current density is calculated to be $J' = I/W_{opt}$; if current consumption does not exceed the limit, go to step 5.

**Step 4)** Repeat steps 1 and 2 with new current density *J'*. The updated $W_{opt}'$ is usually smaller than $W_{opt}$ (because with smaller current density *J'*, noise is larger from active device, and *Q* of input matching network needs to be increased for optimal NF). Therefore, the current consumption is now smaller than the constraint ($J'W_{opt}' < J'W_{opt} = I$). Better NF can be achieved if we update the current density $J'' = I/W_{opt}'$, and repeat steps 1 and 2, the process will finally converge; or if the NF is already satisfactory, go to step 5.

**Step 5)** Add output loading to achieve desired gain.

**Step 6)** Run simulation with more complete noise model for MOSFET (such as BSIM3v3 model) and real inductor model, fine tuning is done to fix input transistor size and input matching network.

### 3.3.4 Design of Inductive Degeneration LNA for WLAN

A fully integrated differential CMOS LNA for WLAN 802.11a is designed using inductive degeneration topology, and integrated in a WLAN transceiver [3. 18].

**Specifications**

The IEEE 802.11a standard for wireless LAN uses the frequency bands of 5.15-5.25 GHz, 5.25-5.35 GHz, and 5.725-5.825 GHz. In our WLAN receiver, only the lower two bands are covered.

Wireless LAN systems require receiver architectures with wide dynamic range. When a transmitter and receiver are close to each other, the received signal strength can be as high as -11 dBm. A highly linear receiver is needed to accommodate such strong signals. On the other hand, the received signal can be quite weak due to fading. The receiver must be sensitive enough to detect signals as small as -148 dBm/Hz. (i.e., -74 dBm for a 24 MHz bandwidth signal). To have a SNR of at least 12 dB, the overall NF of the receiver must be better than

$$NF = -148\ \text{dBm/Hz} - 12\ \text{dB} - (-174\ \text{dBm/Hz}) = 14\ \text{dB} \quad (3.36)$$

where -174 dBm/Hz is the available noise power of the source. Based on the blocking level of IEEE 802.11a standard, receiver IIP3 is derived to be -28.7 dBm [3. 18]. In order to allow for some margin, the target value of the minimum IIP3 is 20 dBm.

Based on receiver specification, LNA requirement is assigned [3. 18], as listed in TABLE 3. 1.

TABLE 3. 1 SPECIFICATION OF WLAN LNA

| Frequency/GHz | Voltage gain/dB | IIP3/dBm | NF/dB | S11/dB |
|---|---|---|---|---|
| 5.15-5.35 | 18-25 | >-14 | 4 | <-10 |

LNA gain needs to be large to suppress the noise contribution of the mixer, but it also should not be too large so as to ensure that mixer is not saturated and the linearity

of the whole receiver is good enough. A gain range of around 7 dB is included to accommodate different levels of input power so that the linearity of the mixer can be relaxed.

Based on the frequency plan of the receiver, an image signal at around 3.15 GHz for the mixer exists. LNA is required to reject it by at least 25 dB.

**Circuit Design**

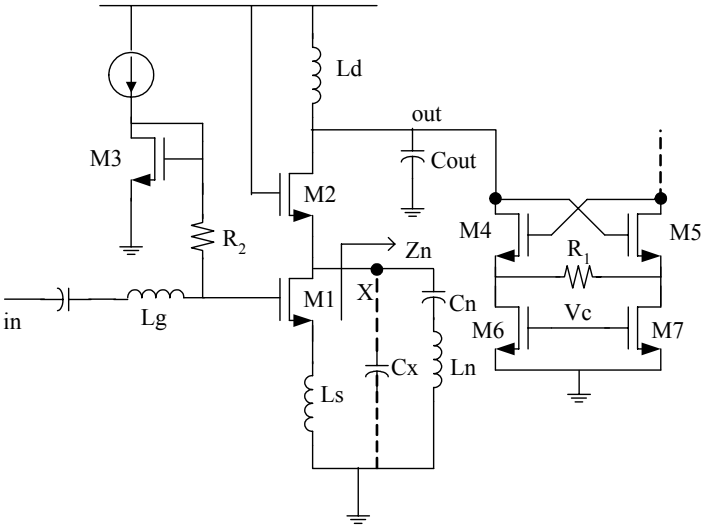



Fig. 3. 16 Inductive source degeneration LNA with notch filter

The schematic of the LNA is shown in Fig. 3. 16 (only single-ended schematic is shown). It is operating at low supply voltage of 1 V, and designed in 0.18-μm CMOS process. M1 is the input amplification device, and M2 works as cascode device. LC parallel tank $L_d$ and $C_{out}$ form the output loading. Negative $g_m$-cell, composed of M4, M5 and M6, is adopted to realize gain tuning function. $C_n$, $C_x$ and $L_n$ form notch filter to get the desired image rejection.

Differential topology is adopted, because balanced and symmetrical design presents a circuit designer with the advantages of keeping the common-mode substrate noise at a minimum due to the inherent properties of high common-mode rejection. In addition, due to inevitable bondwire connecting the circuit and printed

circuit board (PCB) for grounding and power supply, circuit's input impedance and frequency response will be altered by the bondwire inductance because $L_s$ and $L_d$ are changed. The length of the bondwire is hard to control accurately during operation so its inductance is also hard to predict. Differential topology is usually used so that source degenerating inductances $L_s$ and loading inductances $L_d$ return to virtual ground for differential signals, therefore, bondwire's effect can be eliminated.

The LNA input transistor M1's size is optimized for the noise performance, following the proposed procedure, to be 140-μm/0.18-μm. Cascode transistor M2 has the same size, to reduce the miller effect from $C_{gd}$ of M1 and improve reverse isolation. Another point for reverse isolation is to tie all the transistors' substrate to circuit ground instead of their source terminals [3. 15].

**Notch Filter**

Notch filter is adopted to get improved image rejection. It is merged with the LNA to save area and power [3. 19]. The filter comprises an inductor $L_n$ and a capacitor $C_n$, together with the parasitic capacitor $C_x$ at node X in Fig. 3. 16. The filter has imaginary zero at $\omega_z = \frac{1}{\sqrt{L_n C_n}}$ (when series $L_n$ and $C_n$ resonate) and imaginary pole at $\omega_p = \sqrt{\frac{C_x + C_n}{L_n C_n C_x}}$. By locating $\omega_z$ at image frequency and $\omega_p$ at operating frequency, the notch filter thus not only boosts image rejection but also diminishes the effect of the parasitic capacitance at node X, leading to better noise performance of M2. However, considering the on-chip inductor with finite *Q*, the lossy components of $L_n$ contribute noise to the circuit, leading to degraded NF. Nevertheless, we still need notch filter for better image rejection. Taking into consideration the noise from notch filter, the NF is written as:

$$NF = NF_1 + \frac{4R_s}{\omega_0 L_n Q}\left(\frac{\omega_0}{\omega_T}\right)^2 \tag{3. 37}$$

where $Q$ is the quality factor of $L_n$, $NF_1$ is the noise figure from input transistor and matching network. It is better using large $L_n$ for smaller added noise. $L_n$ can be calculated as:

$$L_n = 1/\left[(\omega_p^2 - \omega_z^2)C_x\right] \tag{3. 38}$$

which also shows that to make $L_n$ large, M2 size should not be too big with a large $C_x$.

**Gain Tuning**

The loading of the LNA is a parallel LC tank, as shown in Fig. 3. 17, before compensation negative resistor $-R_c$ is added, at resonant frequency, tank impedance $Z_{in}$ is limited by $R_p$, which originates from the loss of $L_p$.

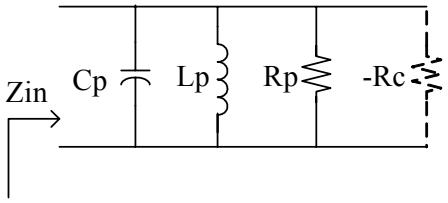



Fig. 3. 17 Parallel LC tank with impedance compensation

After adding negative resistance $-R_c$, at resonant frequency,

$$Z_{in} = R_p //(-R_c) = \frac{R_c R_p}{R_c - R_p} \tag{3. 39}$$

Negative $g_m$-cell is known as negative resistance, the cross-coupled connection is shown in Fig. 3. 16 as M4 and M5. The negative resistance $R_c$ is equal to $1/g_{m4}$, which can also be varied by changing the bias current through M4 and M5. Therefore, LNA loading impedance can be varied, and the corresponding gain is also varied. As $g_{m4}$ increases, gain will also increase.

Using negative $g_m$-cell to increase gain is power-efficient, because we only need to compensate the loss of $R_p$, i.e., $1/g_{m4} = R_p$, the gain can be made infinite. In contrast, if we want to get infinite gain by increasing $g_m$ of input transistor M1, $g_{m1}$ also needs to be infinite. However, gain tuning by connecting negative $g_m$-cell to the loading greatly affects IIP3, because negative $g_m$-cell is quite nonlinear.

Resistive degeneration is used to improve its linearity, $R_1$ is equal to 150 Ω. The negative $g_m$-cell's transistors M4 and M5 are sized to get just enough gain while not adding too much capacitive load, which is 15-μm/0.18-μm.

The NF contribution from the loading is represented in the 2nd term in the following equation:

$$NF = NF_1 + 4R_s\left(\frac{1}{\omega_0 L_d Q} + \gamma\frac{g_{m,4}}{\alpha}\right)\left(\frac{\omega_0}{\omega_T}\right)^2 \qquad (3.40)$$

where $NF_1$ is the NF from input transistor and matching network. As gain increases, NF will be larger due to larger current noise from negative $g_m$-cell. Loading inductor $L_d$ should be maximized to reduce its noise contribution.

2-bit binary-weighted SCAs are added in the loading tank and notch filter, to tune the frequency response in case process variation happens. The SCA has *Q*-factor of 15, and the frequency tuning range is ±8%.

Main design parameters of the LNA are listed in TABLE 3. 2. Simulated NF is 4.1 dB at 5.25 GHz, noise contribution from each part is shown in TABLE 3. 3 when negative $g_m$-cell is turned off.

TABLE 3. 2 MAIN DESIGN PARAMETERS IN THE WLAN LNA

| | |
|---|---|
| Input transistor size | 140-μm/0.18-μm |
| Cascode transistor size | 140-μm/0.18-μm |
| Gate inductor $L_g$ | 3.3 nH |
| Source inductor $L_s$ | 0.69 nH |

| | |
|---|---|
| Drain inductor $L_d$ | 1.2 nH |
| Notch inductor $L_n$ | 1.5 nH |
| Negative $g_m$ transistor size | 15-μm/0.18-μm |
| Degeneration resistor in negative $g_m$-cell | 150 Ω |

TABLE 3. 3 NOISE CONTRIBUTION FROM EACH PART OF THE LNA

| | |
|---|---|
| Input source resistance | 38.8% |
| Loss of $L_g$ | 17.5% |
| Input transistor | 12.8% |
| Loss of $L_d$ | 10.9% |
| Loss of $L_n$ | 9.7% |
| Cascode transistor | 7.5% |
| Loss of $L_s$ | 1.7% |
| Bias resistor $R_2$ | 1.1% |
| Total | 100% |

### 3.3.5 Measurement Setup and Results

Fabricated in TSCM 0.18-μm CMOS process, the LNA occupies area of 1mm*1.1mm. Fig. 3. 18 shows the micrograph of the LNA with each inductor labeled on it. Except $L_g$. inductors $L_s$, $L_d$, $L_n$ are all center-tapped inductors to save chip area and improve *Q*-factor.

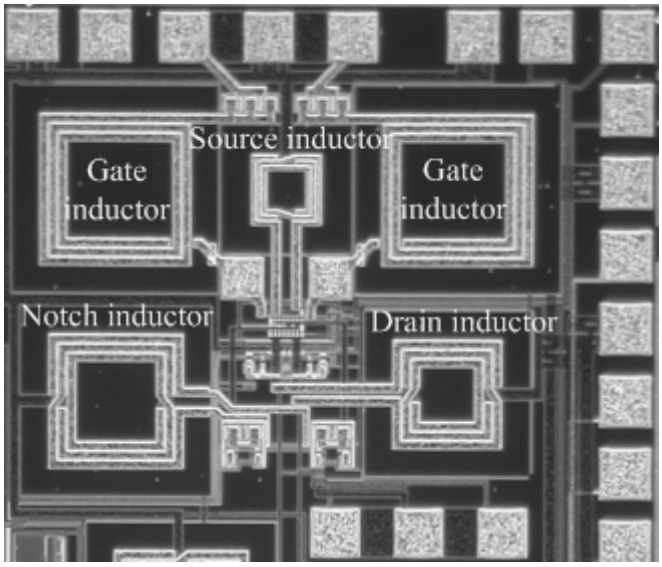



Fig. 3. 18 Micrograph of the WLAN LNA

**Testing Setup**

Due to the importance of $L_g$ on both NF and input impedance matching, a separate testing structure is fabricated and measured. The *Q* of $L_g$ is around 5, and measured data is close to the expectation (the expected *Q* is adjusted to be 5 by fitting $R_s$ in the inductor model, Momentum simulation predicts *Q* of 10), as listed in TABLE 3. 4.

TABLE 3. 4 MEASURED DATA OF THE GATE INDUCTOR

| | L/nH | $R_s$/Ω | $C_{ox1}$/fF | $R_{Si1}$/Ω | $C_{ox2}$/fF | $R_{Si2}$/Ω | *Q* |
|---|---|---|---|---|---|---|---|
| Expected | 3.3 | 4.11 | 89 | 163 | 88 | 210 | 5.09 |
| Sample1 | 3.63 | 7.42 | 94.7 | 85.0 | 103 | 128 | 4.66 |
| Sample2 | 3.62 | 5.91 | 94.2 | 90.6 | 99.5 | 109 | 5.04 |
| Sample3 | 3.62 | 6.08 | 96.0 | 86.8 | 101 | 110 | 4.92 |

*A. Impedance and Gain Measurement*

Impedance matching (S11) and gain can be measured through S parameters using network analyzer.

The circuit is differential-input differential-output (4 ports), whereas network analyzer has only 2 ports. We need to do single-ended to differential conversion at circuit input using power splitter, and differential to single-ended conversion at circuit output using power combiner. The setup is shown in Fig. 3. 19. Port 1 of the network analyzer is connected to a power splitter with $180^o$ phase difference to convert the single-ended signal to differential format. Two bias-T's are used to apply a DC voltage to the input signal so that it is suitable for the LNA input devices. A high-speed SGS probe is used to pick up the output signal of the LNA after a 50-Ω buffer. Two bias-T's are used after the probe to remove the DC voltage from the output signal. A power combiner is used to convert the differential signal to a single-ended signal. The output signal is finally connected to the port 2 of the network analyzer.

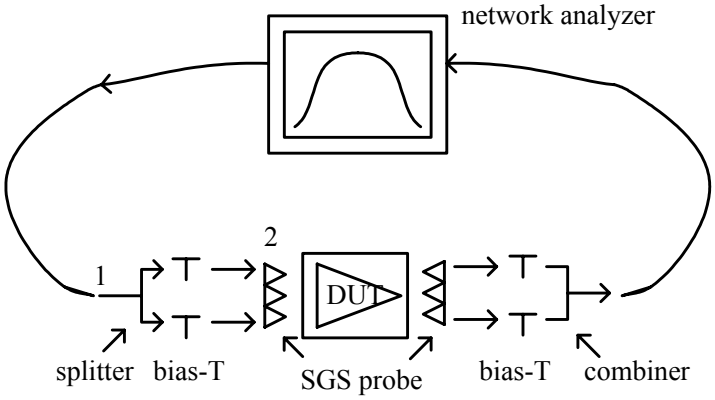



Fig. 3. 19 Impedance and gain measurement

With the setup in Fig. 3. 19, the voltage gain can be calculated from the measured full 2-port S parameters with the help of ADS [3. 20], its shape is close to S21 but the magnitude may be different.

*B. Two-Tone Measurement*

The setup for two-tone measurement is shown in Fig. 3. 20. Two signal generators are used at the input of the LNA to apply two signals at different frequencies. The two signals are combined by a power combiner, and connected to the LNA with a power splitter, two bias-T's and a high-speed SGS probe. The output signal of the device-under-test (DUT) is measured by a spectrum analyzer which is connected to the LNA output by a high-speed probe, two bias-T's and a power combiner.

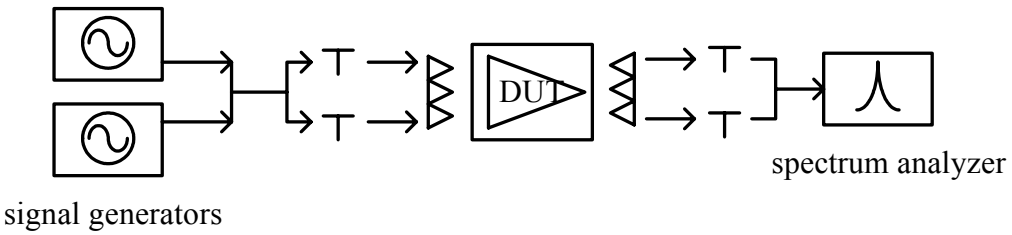



Fig. 3. 20 Two-tone measurement

*C. Noise Figure Measurement*

The setup for the NF measurement is shown in Fig. 3. 21. The setup is similar to that for the frequency response measurements. However, the network analyzer is replaced with a noise figure meter. An accurate noise source is used at the input of the LNA. Because calibration can only be done before the power splitter (due to the fact that noise source cannot be differentially connected after power splitter), the components in the dashed circle are all included in the measurement.

The overall NF at different frequency can be directly read from the NF meter, and NF of DUT needs to be de-embedded from the overall noise figure using Friis equation introduced in Chapter 2. To improve the accuracy of the de-embedded result, the losses in front of the DUT have to be minimized because it directly affects the overall NF, while the losses after DUT are less severe because its influence will be suppressed by the gain of the DUT.

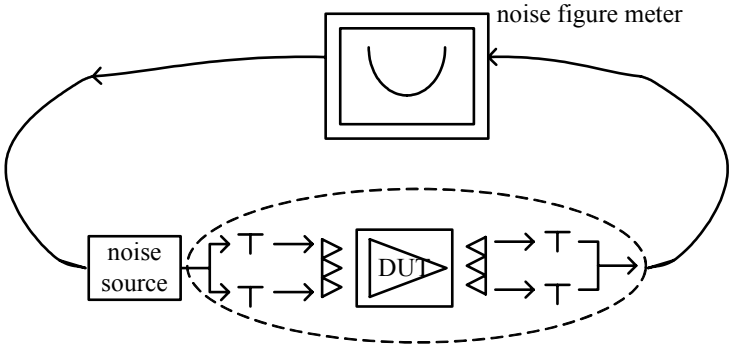



Fig. 3. 21 Noise figure measurement

**Experimental Results**

The WLAN LNA consumes around 10 mA from 1-V supply. The measured input matching is shown in Fig. 3. 22. Inductors $L_s$ and $L_g$ are close to the expected value, but $C_{gs}$ from input transistor seems smaller than the predicted value, so the best impedance matching shifts to higher frequency. However, S11 is still better than -13 dB within 5.15 – 5.35 GHz. The overall frequency response of the LNA is measured

and plotted in Fig. 3. 23. The gain of the LNA is tunable from 17.0 dB to 25.2 dB at 5.25 GHz. Rejection for 3.15 GHz image is at least 27 dB, which is high enough to eliminate an off-chip image-rejection RF filter. The two-tone measurement result is shown in Fig. 3. 24. Two input signals are located at 5.2 GHz with 20 MHz away from each other. The IIP3 is measured to be -3.2 dBm for 17 dB gain setting.

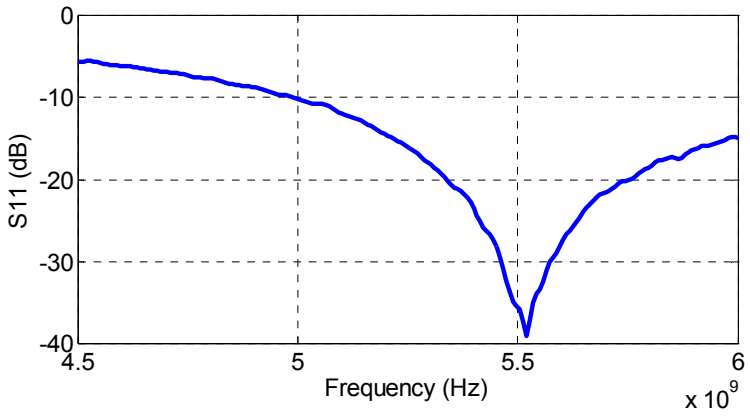



Fig. 3. 22 S11 of the LNA

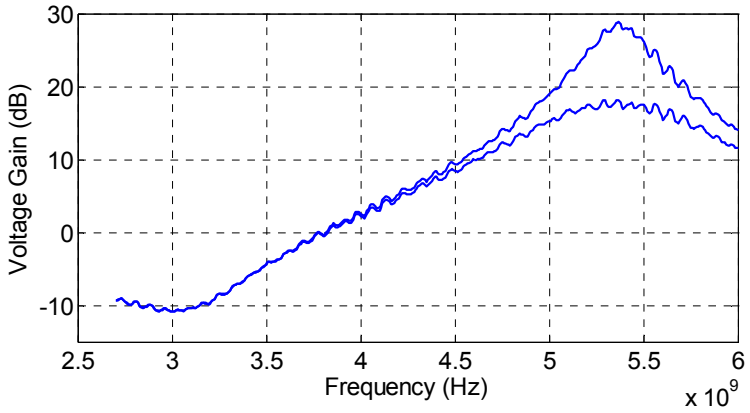



Fig. 3. 23 Frequency response of the LNA

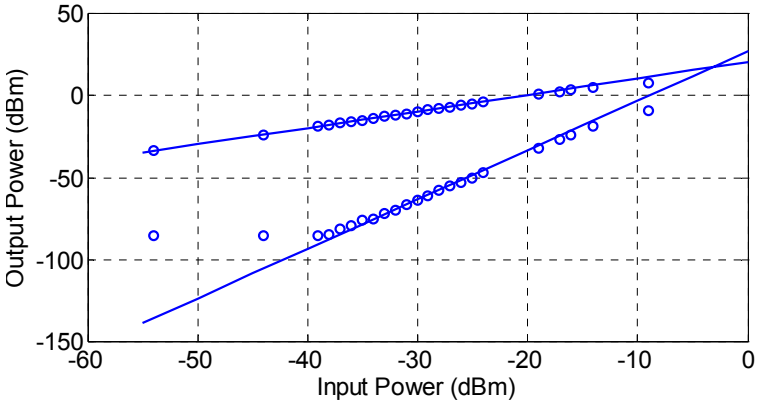



Fig. 3. 24 IIP3 of the LNA

The measured performance is summarized in TABLE 3. 5.

TABLE 3. 5 MEASURED PERFORMANCE OF THE WLAN LNA

| | Voltage Gain/dB | S11/dB | NF/dB | IIP3/dBm | Power /mW |
|---|---|---|---|---|---|
| Low Gain | 17.0 | -13.0 | 4.7 | -3.2 | 9.7 |
| High Gain | 25.2 | -10.6 | 4.9 | -19.9 | 11.7 |

## 3. 4. High Linearity LNA Design for RFID Reader

### 3.4.1 Introduction

Linearity is an important characteristic of a receiver. As introduced in Chapter 2, third order nonlinearity causes weak signals to be blocked in the presence of strong signals. For example, a pair of strong signals at 1.00 and 1.01 GHz can interfere with the reception of a weak signal at 1.02 GHz through third order nonlinearity, despite the fact that all three signals are on different frequency channels.

Radio frequency identification (RFID) is drawing more and more attention nowadays. There are several hardware challenges for the RFID reader, one of which is the linearity of the receiver front-end. Because there may be a large echo signal from the transmitter antenna to the receiver, RFID reader LNA has to be linear enough in order not to generate large intermodulation (IM) signals.

Transconductance and output conductance are two dominant nonlinear sources of MOSFETs. When biased in saturation region, and with small signal operation, nonlinear factors, such as input impedance, conductive substrate and drain-source-substrate junction diode capacitance, have significantly smaller effect than transconductance and output conductance blow 6 GHz [3. 21]. For low load impedance, output conductance will also not be a dominant nonlinear source.

To get better linearity, many linearization techniques are proposed [3. 23]-[3. 30], we will review briefly the existing solutions in the following section, to see their effectiveness and tradeoffs.

### 3.4.2 Prior Linearization Methods

**Linear Feedback**

Feedback is the most widely known linearization technique, based on feeding back a linearly scaled version of the output signal and subtracting it from the input. The block diagram of the method is shown in Fig. 3. 25. The open-loop transfer function of the amplifier is described by the following power series:

$$y(t) = a_1 e(t) + a_2 e^2(t) + a_3 e^3(t) + \cdots \tag{3. 41}$$

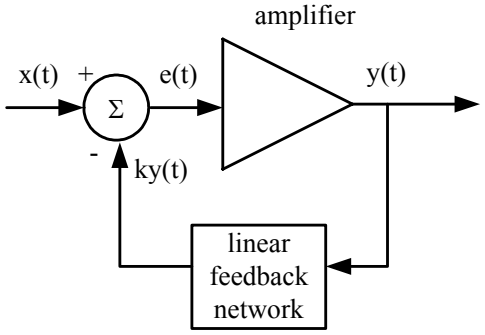



Fig. 3. 25 Linear feedback method

where $a_1$ is the open-loop small-signal gain of the amplifier, and the higher-order coefficients ($a_2$, $a_3$, …) characterize the nonlinearities of the open-loop transfer function, e(t) is the error signal, given by

$$e(t) = x(t) - ky(t) \tag{3. 42}$$

where $k$ is the feedback factor. The closed-loop transfer function can be represented by a power series as

$$y(t) = c_1 x(t) + c_2 x^2(t) + c_3 x^3(t) + \cdots \tag{3. 43}$$

The coefficients $c_n$'s are functions of $a_n$'s and $k$. Their derivations can be found in [3. 22]. The two important coefficients are

$$\begin{aligned} c_1 &= \frac{a_1}{1+T} \\ c_3 &= \frac{a_3}{(1+T)^4} - \frac{2a_2^2}{a_1}\frac{T}{(1+T)^5} = \frac{a_3}{(1+T)^4}\left(1 - \frac{2a_2^2}{a_1 a_3}\frac{T}{1+T}\right) \end{aligned} \tag{3. 44}$$

where $T=a_1k$ is the loop gain. As expected, the negative feedback reduces the small-signal gain of the amplifier by a factor of $(1+T)$. The closed-loop 3rd order nonlinearity, represented by $c_3$, has two contributions: that of an open-loop 3rd order nonlinearity, reduced by a factor of $(1+T)^4$, and that contributed from 2nd order nonlinearity. For small value of the loop gain or small $a_2$,

$$c_3 \approx \frac{a_3}{(1+T)^4} \tag{3. 45}$$

and thus,

$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{c_1}{c_3}\right|} \approx \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|(1+T)^3} \tag{3. 46}$$

That is, $A_{IP3}$ is increased by a factor of $(1+T)^{3/2}$ in comparison with the open-loop case. From (3. 44), if $a_1$ and $a_3$ have the same sign, $c_3$ and thus, the 3rd order IM distortion (IM3) can be made zero by properly selecting the loop gain $T$. This IM3

cancellation has rarely been used in practical analog circuits because the $c_3$ null is very narrow and very difficult to maintain over a wide range of operating conditions and process parameters [3. 22].

There are two main approaches to apply a negative feedback in an amplifier: a source or emitter degeneration, a parallel resistive feedback. Inductive degeneration in LNAs not only brings the conjugate input impedance matching, but also provides good linearity. The main drawback of the negative feedback method is the reduced gain.

**Feedforward**

Feedforward technique is based on splitting the input into two signals, amplifying them by two amplifiers with different transfer characteristics, as shown in Fig. 3. 26, such that upon combining their output signals, their distortions cancel each other.

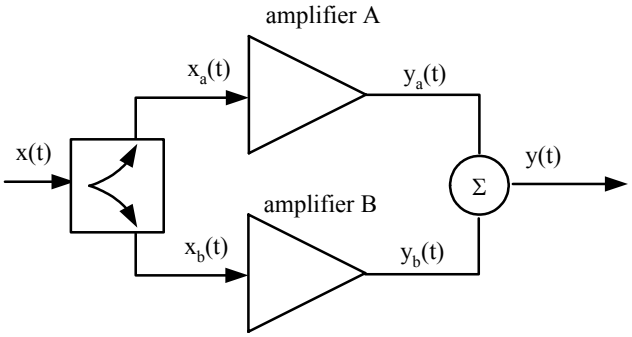



Fig. 3. 26 Feedforward linearization technique

One of the well known implementations of this technique is the multi-tanh method [3. 23]. It is shown in Fig. 3. 27. The differential pair formed by M1 and M2 can be viewed as the main amplifier, whereas the pair formed by M3 and M4 is an auxiliary amplifier, whose purpose is to cancel IM3 of the main amplifier. Assuming the square-law characteristics of the FETs, the combined differential output current is given by

$$i_{out} = v_{in}\sqrt{KI_0}\sqrt{1-\frac{Kv_{in}^2}{4I_0}} - v_{in}\sqrt{mKnI_0}\sqrt{1-\frac{mKv_{in}^2}{4nI_0}} \tag{3. 47}$$

where $v_{in}$ is the differential input voltage, $K$ is the transconductance parameter of M1 and M2, $m < 1$ and $n < 1$ are the scaling ratios. The corresponding power series coefficients are:

$$\begin{aligned} g_1 &= \sqrt{KI_0}(1-\sqrt{mn}) \\ g_3 &= -\frac{1}{8}\sqrt{\frac{K^3}{I_0}}\left(1-\sqrt{\frac{m^3}{n}}\right) \end{aligned} \tag{3. 48}$$

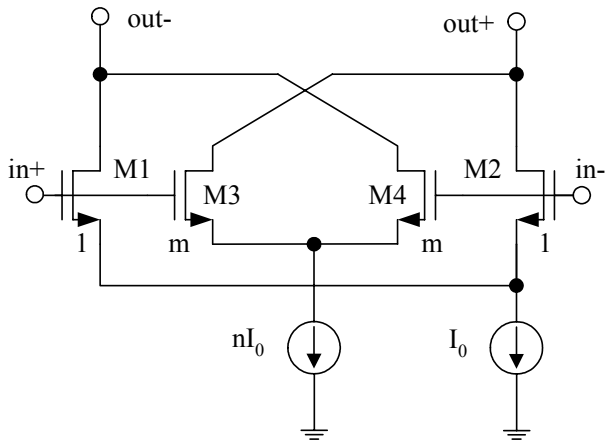



Fig. 3. 27 Cross-coupled CMOS differential pairs

If $n$ is designed to be equal to $m^3$, $g_3$ is zero, and the transconductance is degraded by $(1-m^2)$. To reduce the degradation, $m$ should be as small as possible. For example, for $m = 0.5$, the gain is degraded by 2.5 dB.

Another approach similar to the one shown in Fig. 3. 27 is used to linearize the differential CMOS LNA in [3. 24]. In this LNA, both the tail current and the differential FETs of the auxiliary amplifier are scaled by the same ratio $\alpha$. The input signal of the main amplifier is attenuated $\beta$ times that of the auxiliary amplifier. Using the block diagram in Fig. 3. 26, with amplifier A being the main amplifier and amplifier B being the auxiliary amplifier, we can write their transfer functions as

$$y_a(t) = a_1\beta x(t) + a_3(\beta x(t))^3$$
$$y_b(t) = \alpha[a_1 x(t) + a_3 x^3(t)] \quad (3.\ 49)$$

where we have neglected the 2nd order terms for simplicity. After subtracting $y_b(t)$ from $y_a(t)$, we get

$$y(t) = y_a(t) - y_b(t) = (\beta - \alpha)a_1 x(t) + (\beta^3 - \alpha)a_3 x^3(t) \quad (3.\ 50)$$

If $\alpha$ is designed to be equal to $\beta^3$, the IM3 is cancelled, and the fundamental signal is attenuated by $\beta(1-\beta^2)$. The optimum value of $\beta$ for maximizing gain is $1/\sqrt{3}$, and the overall gain is reduced by $2\sqrt{3}/9$ or 8.3 dB relative to the gain of the main amplifier.

This implementation of feedforward linearization suffers from several drawbacks. First, the overall gain of the composite amplifier is significantly degraded. Second, NF is high due to the fact that the noise power of the two amplifiers adds, while their desired signals subtract. Third, splitting the signal with a well controlled attenuation value is technically challenging without using bulk coaxial assemblies.

The multiple gated transistor linearization [3. 25] utilizes the characteristic of $180^o$ IM3 phase difference between two differently gate-biased transistors. As shown in Fig. 3. 28 of simulated nonlinearity coefficients for a 60-μm/0.18-μm nMOSFET, when $V_{gs}$ increases, $g_3$ will change from positive to negative sign. By combining the two transistors' drain current with anti-phase and equal-amplitude in IM3, we can get IM3 cancellation. This linearization depends on the device physical characteristic, large input power level may affect the characteristic and degrade the cancellation.

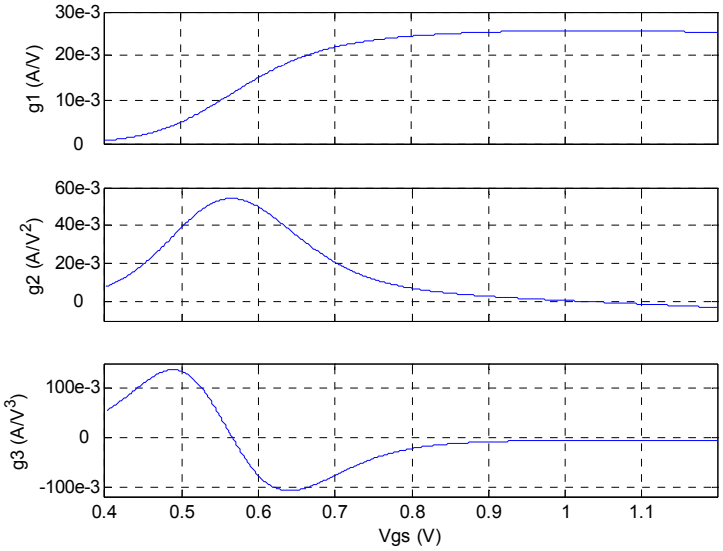



Fig. 3. 28 Coefficient of $g_1$, $g_2$ and $g_3$ as a function of gate bias for an nMOSFET

**Predistortion and postdistortion**

The predistortion method [3. 26] [3. 27] adds a nonlinear element (also called linearizer) prior to an amplifier such that the combined characteristic of the two devices is linear, while postdistortion [3. 28] [3. 29] uses a linearizer after the amplifier. In practice, it is impossible to cancel all orders of nonlinearity simultaneously, therefore, the linearizer is usually designed to cancel the nonlinearity of a certain order. The cancellation of the 3$^{rd}$ order nonlinearity is more common because it controls IM3 and the gain compression or expansion of an amplifier. If the amplifier exhibits a gain compression, the linearizer is designed to have a gain expansion, and vice versa. The linearizer can be either shunt or series, active or passive.

The main challenge of open-loop predistortion and postdistortion is to design a practical linearizer with the desired transfer function. Variations in the amplifier transfer function, caused by tolerances of the manufacturing process, require manual tuning of the linearizer from part to part, making these methods costly and ill-suited

for high-volume production. An adaptive feedback is often added to overcome this drawback, but it makes the circuit rather complex.

**Adaptive bias**

In [3. 30], an adaptive bias scheme is used in which the bias current is designed to be increased as a function of the input signal in order to compensate for the nonlinearity in a square-law device. However, the technique is only suitable for device linearization at dc or low frequencies.

**Second-Order Harmonic/Intermodulation Injection**

It has been reported to reduce the 3rd order inter-modulation (IM) using the 2nd order IM components [3. 31]. However, the implementation is complicated with discrete components, phase shifter, filter and combiner are usually required [3. 32] [3. 33] to control the phase and amplitude of the injected signal. A typical block diagram is shown in Fig. 3. 29. Its complication makes it difficult to be fully integrated on-chip and the cost and power consumption will be high.

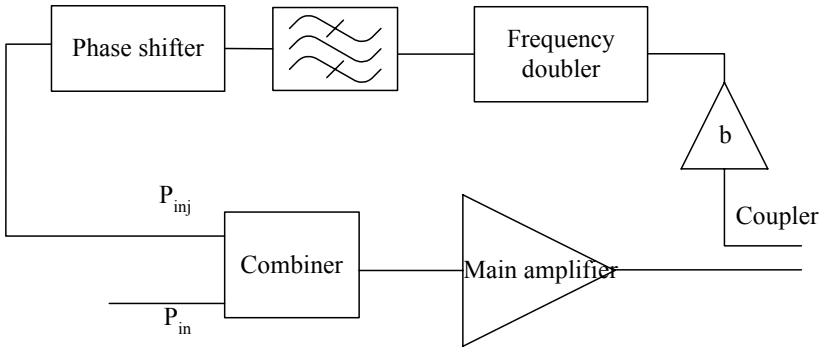



Fig. 3. 29 Block diagram of 2nd harmonic injection scheme

In summary, existing solutions linearize the circuit at the cost of larger power consumption, larger NF, or smaller gain. Sometimes such tradeoff cannot be afforded, e.g., the noise in LNA should not be sacrificed, linearization scheme with less tradeoff is highly desired.

We propose a linearization technique with low-frequency 2nd IM product (IM2) injected to the current source of the differential pair, getting suppressed IM3, without affecting gain, NF and consuming little power with simple extra circuitry.

### 3.4.3 Proposed IM2 Injection Linearization

For a differential pair, IM3 mainly originates from the nonlinearity of the devices' I-V characteristic. The small-signal output current around the quiescent bias point can be expressed by a Taylor series expansion as

$$i_p = g_1(v_g - v_s) + g_2(v_g - v_s)^2 + g_3(v_g - v_s)^3 + \cdots \quad (3.51)$$

where $g_i$ represents the i$^{th}$-order transconductance coefficient of the input device.

As shown in Fig. 3. 30, a linearization technique is proposed by injecting a low-frequency IM2 signal to the common source node of the differential pair. The injected IM2 signal, with proper phase and amplitude, will mix with the fundamental tones through the 2nd term $g_2(v_g\text{-}v_s)^2$ in (3. 51) and generate a 3rd-order output current having equal amplitude and anti-phase with the transconductor's intrinsic IM3, which can be used for IM3 suppression and linearization.

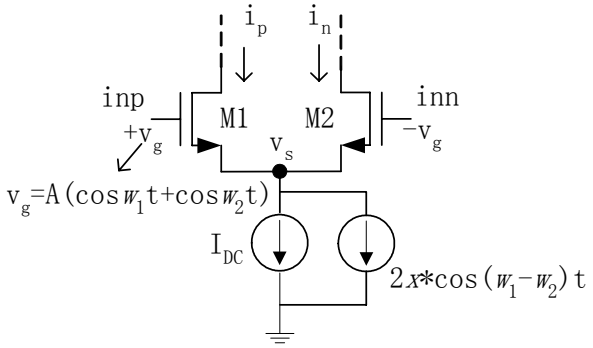



Fig. 3. 30 Schematic of a differential pair with the developed linearization technique

For simplicity, assume that two input signals at $\omega_1$ and $\omega_2$ with equal amplitude, $A\cos(\omega_1 t)$ and $A\cos(\omega_2 t)$, are applied to the differential pair's input. With an IM2 current, $2x\cdot\cos(\omega_1-\omega_2)t$, injected to the bias current source, the output current becomes:

$$i_p + i_n = 2x \cdot \cos(\omega_1 - \omega_2)t \tag{3. 52}$$

By combining (3. 51) and (3. 52), the small-signal voltage $v_s$ at the common-source node at frequency $(\omega_1-\omega_2)$ is derived to be:

$$v_s\big|_{\omega 1-\omega 2} = \frac{g_2 A^2 - x}{g_1}\cos(\omega_1 - \omega_2)t \tag{3. 53}$$

Other 2nd-order frequency components in $v_s$ located at frequencies ($\omega_1+\omega_2$, $2\omega_1$, and $2\omega_2$) have amplitudes of (1, 1/2, 1/2)* $\frac{g_2 A^2}{g_1}$, respectively. Combining (3. 53) with (3. 51),

$$\begin{aligned} i_p\big|_{3rd} = &-2g_2 A(\cos\omega_1 t + \cos\omega_2 t)\frac{g_2 A^2 - x}{g_1}\cos(\omega_1 - \omega_2)t + g_3 A^3(\cos\omega_1 t + \cos\omega_2 t)^3 \\ &-2g_2 A(\cos\omega_1 t + \cos\omega_2 t)\frac{g_2 A^2}{g_1}[\cos(\omega_1 + \omega_2)t + \frac{1}{2}\cos 2\omega_1 t + \frac{1}{2}\cos 2\omega_2 t] \end{aligned} \tag{3. 54}$$

The IM3 component of the output current at $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$ has an amplitude of $-g_2 A\frac{3g_2 A^2 - 2x}{2g_1} + \frac{3g_3 A^3}{4}$, which can be cancelled with a proper value of $x$ shown in (3. 55):

$$x = (-\frac{3g_1 g_3}{4g_2} + \frac{3}{2}g_2)A^2 \tag{3. 55}$$

(3. 55) shows the requirement on the amplitude of the injected low-frequency IM2 current. To determine the phase requirement of the injected signal, for the two input tones $A\cdot\cos(\omega_1 t)$ and $A\cdot\cos(\omega_2 t)$, the intrinsic IM3 generated by term $g_3(v_g-v_s)^3$ in (3. 51) at the frequencies $2\omega_1-\omega_2$ and $\omega_1-2\omega_2$ have a zero phase. With a low-frequency

IM2 signal $2x\cdot\cos[(\omega_1-\omega_2)t+\varphi]$ injected at $v_s$, where $\varphi$ is an initial phase, the IM3 generated at $2\omega_1-\omega_2$ and $\omega_1-2\omega_2$ through term $g_2(v_g-v_s)^2$ in (3. 51) for linearization also has a phase of $\varphi$. Therefore, the proper value for $\varphi$ of the injected IM2 signal is 0 or $\pi$ (depending on the sign of $x$ in (3. 55)), at the time when input two tones have phase 0, which is equivalent to saying that the injected signal should be in phase with the envelope of the input two tones. It can be further shown that the IM3 suppression is proportional to $|\sin\varphi|$.

By mixing injected IM2 with the fundamental tones, there are also other terms generated at $\omega_1$ and $\omega_2$. However, under the condition in (3. 55), their amplitudes have the same order of magnitude with the un-compensated IM3 of the transconductor, which is several tens of dB smaller than the fundamental tones, and therefore the fundamental tone is not affected by the linearization.

If the IM2 injection current is changed to be $2x\cdot\cos(\omega_1+\omega_2)t$, only IM3 at frequencies $2\omega_1+\omega_2$ and $2\omega_2+\omega_1$ can be cancelled. Even though the injection current at $2\omega_1$ together with $2\omega_2$ can also be used to cancel IM3 at $2\omega_1-\omega_2$ and $2\omega_2-\omega_1$, the phase shift for these high-frequency tones is hard to be controlled accurately, which in turn limits the achievable IM3 suppression due to the phase inaccuracy.

The IM2 injection technique is applicable for various RF circuits because the injected signal is not related to the RF frequency but the channel spacing ($\omega_1-\omega_2$). Furthermore, effective cancellation over wide range of input power may be achieved because the injected signal is tracking with the input signal as shown in (3. 55).

### 3.4.4 RFID LNA Implementation with Linearization

The linearization technique is applied to the LNA in an RF receiver front-end covering a frequency band from 860 MHz to 960 MHz for RFID reader to avoid large intermodulation caused by strong echo signals from the reader's transmitter antenna

to the reader's receiver. The schematic of the proposed RFID LNA is shown in Fig. 3. 31.

**Squaring Circuit to Generate Desired IM2 Tone**

A simple squaring circuit, composed of M8, M9 and R3, is used to generate the IM2 tones for linearization. Its input signals are obtained directly from the LNA input signals whereas its output voltage is applied to the bias transistor M7. If the input signal to the squaring circuit is $A\cdot\cos(\omega_1 t)+A\cdot\cos(\omega_2 t)$, and if $C_1$ and $R_4$ are large enough to form ideal AC coupling, the injected IM2 current can be expressed as

$$i_{injected,2nd} = [A\cos(\omega_1 t) + A\cos(\omega_2 t)]^2 (-2g_{2,M8})[R_3 // (sC_p)^{-1}]g_{1,M7} \quad (3.\ 56)$$

where $g_{2,M8}$ is the 2nd transconductance of M8, and $C_p$ is the parasitic capacitor at the output node of the squaring circuit. At the low frequency of the IM2 tones, the parasitic capacitor $C_p$ can be ignored, and the IM2 injected current becomes $A^2\cdot\cos(\omega_1-\omega_2)t\cdot(-2g_{2,M8})\cdot R_3\cdot g_{1,M7}$. The PMOS transistors have negative $g_2$, so $\varphi=0$ for the injected IM2 signal, and the phase requirement is inherently satisfied. The output of the squaring circuit also contains frequency components at $2\omega_1$, $2\omega_2$, and $(\omega_1+\omega_2)$. These 3 high-frequency tones experience larger phase shift than $(\omega_1-\omega_2)$ due to parasitic $C_p$, as discussed earlier, which would limit the achievable IM3 cancellation. Large R3 of 3-kΩ is used to filter out these tones and to produce the required low-frequency IM2 amplitude with less bias current in squaring circuit. The transistor size in the squaring circuit is designed to be 6-μm/0.18-μm in order to achieve the required IM2 amplitude for maximum IM3 cancellation. In Fig. 3. 31, $C_1$, $C_p$, $R_3$ and $R_4$ form a band-pass filter, and the phase shift will be zero at frequency $\omega = 1/\sqrt{R_3 C_p R_4 C_1}$, for which tone-spacing the linearization works best.

In this design, the cut-off frequency for the squaring circuit is designed to be around 80 MHz, and the IM2 amplitude can be tuned by changing DC bias of M8 and M9. If IM3 suppression for wide tone spacing is desired, R3 would need to be smaller to extend the bandwidth for injection signals at the cost of larger power consumption in the squaring circuit to maintain the same IM2 output.

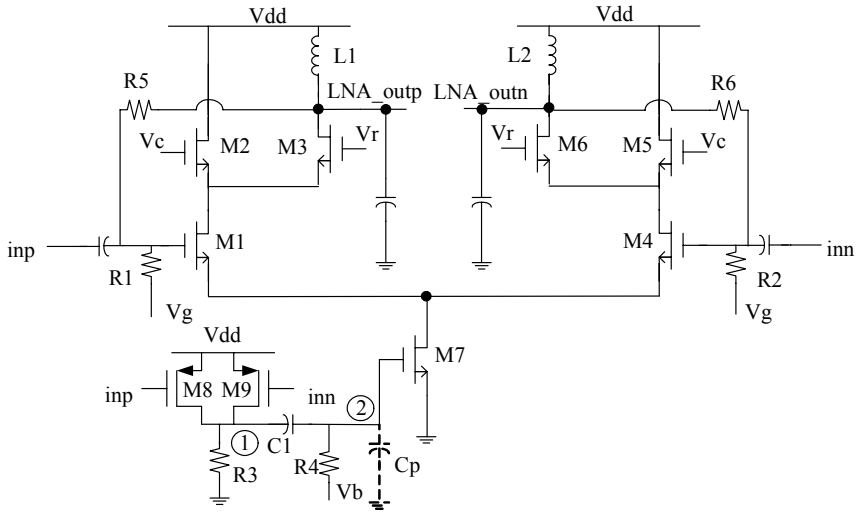



Fig. 3. 31 Schematic of the RFID LNA with linearization

The desired amplitude and phase of the low-frequency IM2 injection signal at the gate of M7 is shown in Fig. 3. 32 when -22-dBm two tones with 5-MHz tone spacing are applied at the input of the LNA. The required injection signal is 1mVp, much weaker than the input signal 34mVp for each tone, which further confirms that the normal operation of the transconductor is not affected.

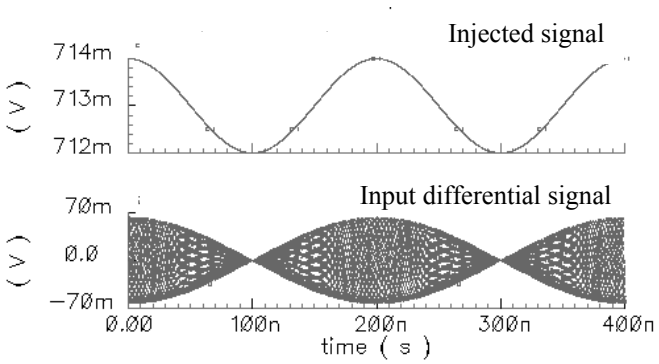


Fig. 3. 32 Simulated transient responses of the IM2 injected signal at the gate of M7 in comparison with the input signal

All the noise generated by the squaring circuit is injected at the common-source node and is rejected by the differential pair's high common-mode rejection ratio.

One potential problem with the proposed technique is that the injected IM2 may leak to the output and degrade IIP2 in the presence of device mismatches. However, this is typically not a problem for LNAs in narrow-band applications or for mixers with relatively high output frequency, in which IM2 tones are located completely outside of the signal bands.

**LNA Design**

As shown in Fig. 3. 31, the LNA input matching is realized by the feedback resistors $R_5$ and $R_6$ instead of inductive source degeneration because the latter would require four ~10-nH inductors for the differential pair, which occupy very large chip area. Assuming that the parallel reactive components from the gate capacitors of the differential pair can be neglected due to relatively low operation frequency of 900 MHz, the input impedance is given by:

$$Z_{in} = \frac{R_5}{(1 + A_v)} \tag{3. 57}$$

where $A_v$ is the voltage gain of the amplifier. By putting the matching resistors in the feedback loop, the resistance value can be increased by the voltage gain $A_v$, and the NF can be improved.

The variable gain tuning is realized with current steering pair M2 and M3. M2 is directly connected to Vdd to avoid its distortion contribution at low-gain setting.

### 3.4.5 Negative $g_m$-Cell Design with Linearization

To further demonstrate its potential, the linearization technique is also applied to a negative $g_m$-cell formed by M5 and M6 connected to the output of an amplifier prototype as shown in Fig. 3. 33. The squaring circuit is composed of M3, M4 and R3.

Main design parameters of RFID LNA and the amplifier with negative $g_m$-cell are summarized in TABLE 3. 6.

TABLE 3. 6 DESIGN PARAMETERS FOR THE TWO AMPLIFIER PROTOTYPES

| | RFID LNA | Amplifer with negative $g_m$ |
|---|---|---|
| Input transistor size | 120-μm/0.18-μm | 40-μm/0.18-μm |
| Cascode trnasistor size | 120-μm/0.18-μm | - |
| Load resistor | - | 200-Ω |
| Load inductor | 15-nH | - |
| Feedback resistor | 500-Ω | - |
| Negativ $g_m$ transistor size | - | 10-μm/0.18-μm |
| Suqaring transistor size | 6-μm/0.18-μm | 2-μm/0.18-μm |
| Suqaring circuit resistor | 3-kΩ | 2.55-kΩ |

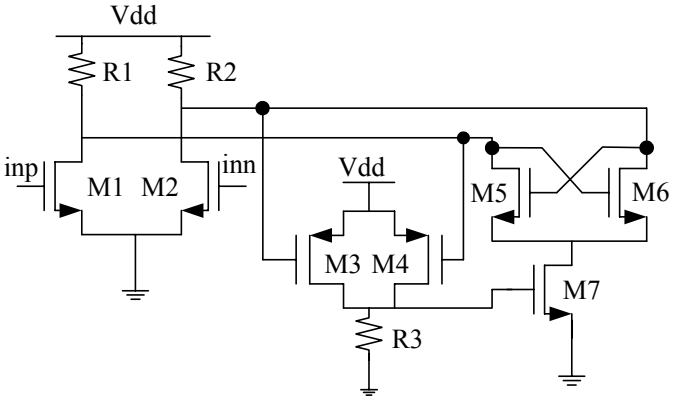
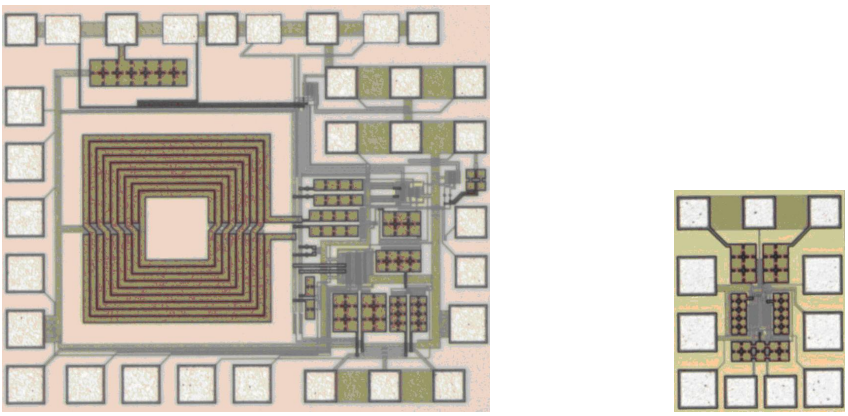



Fig. 3. 33 Linearization applied to a negative $g_m$-cell at the output of an amplifier

### 3.4.6 Experimental Results

The two design prototypes are fabricated in TSMC 0.18-μm CMOS process. The chip micrograph is shown in Fig. 3. 34.



(a) (b)

Fig. 3. 34 Die micrographs of (a) RFID LNA, and (b) amplifier with negative $g_m$-cell

For comparison, the squaring circuit is turned on and off through its gate bias to enable and disable the IM2 injection, respectively. The RFID LNA in Fig. 3. 31 consumes a total current of 10.4-mA from 1.8-V power supply. Fig. 3. 35 shows the measured voltage gain of the RFID LNA, Fig. 3. 36 shows the measured IIP3. IM3 is suppressed by more than 17 dB with 0.1-mA extra current for the squaring circuit while both the gain and NF are not affected by the linearization. Fig. 3. 37 shows the output spectrum.

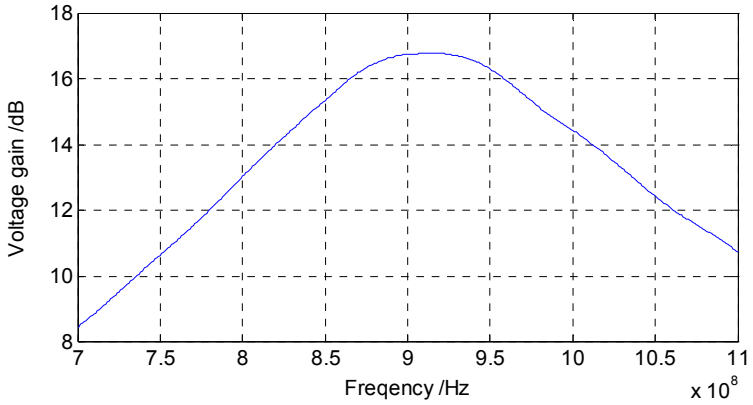



Fig. 3. 35 Measured voltage gain of the RFID LNA

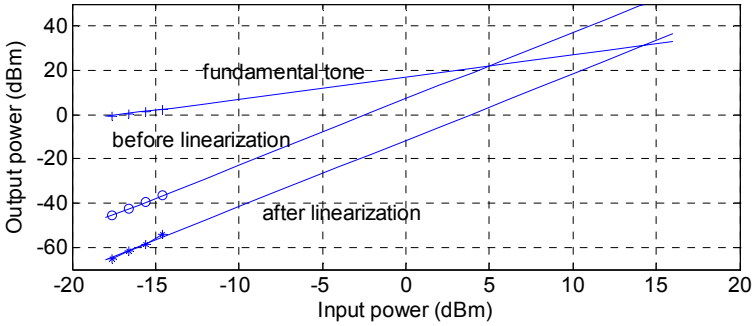
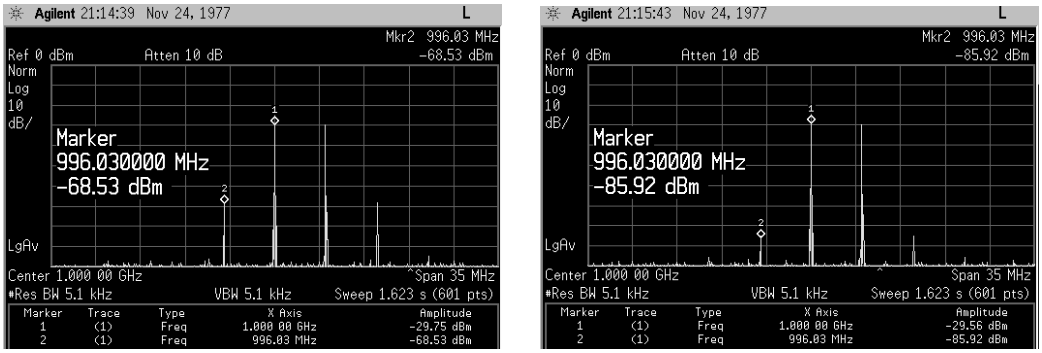



Fig. 3. 36 Measured IIP3 of the RFID LNA




(a) (b)

Fig. 3. 37 Output spectrum of the RFID LNA (a) without, and (b) with the linearization circuit being turned on

The second amplifier prototype delivers 16 dB gain with 7.3 dB contributed by the negative $g_m$-cell and consumes 5.7-mA current from 1.8-V supply. The IIP3 is measured to be -8.3 dBm without linearization. With the linearization, IM3 suppression of 20 dB is achieved in this case, which demonstrates the effectiveness of linearization for the negative $g_m$-cell.

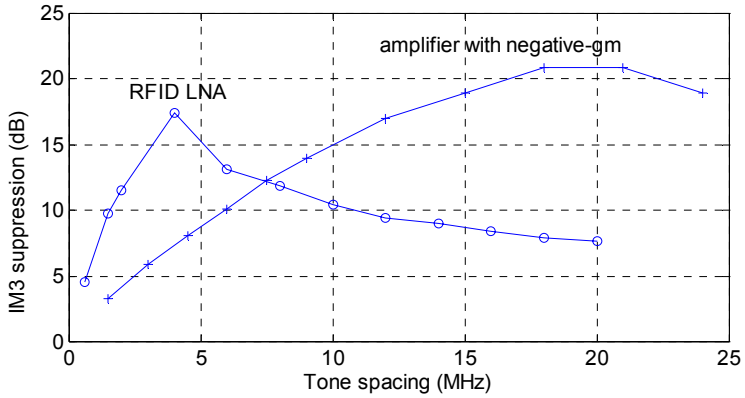



Fig. 3. 38 IM3 suppression as function of two-tone spacing

TABLE 3. 7 PERFORMANCE SUMMARY OF THE TWO AMPLIFIER PROTOTYPES

| | RFID LNA | Amplifier with negative-$g_m$ cell |
|---|---|---|
| Voltage Gain (dB) | 16.8 | 16 |
| S11 (dB) | -14 | - |
| NF (dB) | 4.1 | - |
| IIP3 without / with linearization (dBm) | 5.4 / 14.3 | -8.3 / 1.7 |
| Input $P_{-1dB}$ (dBm) | -4.5 | - |
| Supply Voltage (V) | 1.8 | 1.8 |
| Power (mW) | 18.7 | 10.3 |
| Area ($mm^2$) | 1.0 | 0.19 |
| Process | 0.18-μm CMOS | |

Fig. 3. 38 shows the IM3 suppression as a function of two-tone spacing. If good IM3 suppression is needed for small tone spacing, AC coupling $R_4$ and $C_1$ in Fig. 3. 31 need to be removed, if wideband IM3 suppression is desired, R3 can be smaller to extend the bandwidth for injection signal, at the cost of larger power consumption in squaring circuit. TABLE 3. 7 summarizes the measured performance of the two amplifiers.

## REFERENCES

[3. 1] IEEE Std. 802.11a-1999 (R2003), http://standards.ieee.org/getieee802/802.11.html
[3. 2] C. P. Yue, C. Ryu, J. Lau, T. H. Lee, and S. S. Wong, "A physical model for planar spiral inductors on silicon," *International Electron Devices Meeting (IEDM)*, pp. 155-158, Dec. 1996.
[3. 3] A. M. Niknejad, and R. G. Meyer, "Analysis, design, and optimization of spiral inductors and transformers for Si RF IC's," *IEEE J. Solid-State Circuits*, Vol. 33, No. 10, pp. 1470-1481, Oct. 1998.
[3. 4] H. M. Greenhouse, "Design of planar Rectangular micoroelectronic inductors," *IEEE Trans. Parts. Hybrids, and Pachaging*, pp. 101-109, Vol. 10, Issue 2, Jun. 1974.
[3. 5] T. H. Lee, *the Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.
[3. 6] Sonnet User's Manual, *Sonnet Software*, 1996.
[3. 7] A. M. Niknejad, and R. G. Meyer, "Analysis, design and optimization of spiral inductors and transformers for Si RF IC's," *IEEE J. Solid-State Circuits*, pp. 1470-1481, Vol. 33, Oct. 1998.
[3. 8] *Momentum User's Manual*, Agilent Technologies, Aug. 2005.
[3. 9] M. Danesh, and J. R. Long, "Differentially driven symmetric microstrip inductors," *IEEE Trans. Microwave Theory and Techniques*, Vol. 50, No. 1, pp. 332-341, Jan. 2002.
[3. 10] S. Galal, and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *IEEE J. Solid-State Circuits*, pp. 2334-2340, Dec. 2003.
[3. 11] Chi-Wa Lo, and Howard C. Luong, "A 1.5-V 900-MHz monolithic CMOS fast-switching frequency synthesizer for wireless appplication," *Symp. VLSI Circuits*, Jun. 2000.
[3. 12] C. B. Guo, C. W. Lo, T. Choi, I. Hsu, D. Leung, T. Kan, A. Chan, H. C. Luong, "A 900-MHz fully-integrated CMOS wireless receiver with on-chip RF and IF filters and 79-dB image rejction," *IEEE J. Solid-State Circuits*, Vol. 37, No. 8, pp. 1084-1089, Aug. 2002.
[3. 13] V. Cheung, and H. C. Luong, "A 1-V 10-mW monolithic Bluetooth rceiver in a 0.35-μm CMOS process," *European Solid-State Circuits Conference (ESSCIRC)*, Sep. 2003.
[3. 14] L. Leung, D. Lau, S. Lou, A. Ng, R. Wang, G. Wong, P. Wu, H. Zheng, V. Cheung, and H. C. Luong, "A 1-V 86-mW-Rx 53-mW-TX single-chip CMOS transceiver for WLAN IEEE 802.11a," *IEEE J. Solid-State Circuits*, to appear.
[3. 15] D. K. Shaeffer, and T. H. Lee, "A 1.5-V, !.5-GHz CMOS low noise amplifer," *IEEE J. Solid-State Circuits*, Vol. 32, No. 5, pp. 745-759, May 1997.
[3. 16] K. Han, J. Gil, S. Song, J. Han, H. Shin, C. Kim, and K. Lee, "Complete high-frequency thermal noise modeling of short-channel MOSFETs and design of 5.2-GHz low noise amplifier," *IEEE J. Solid-State Circuits*, Vol. 40, No. 3, pp. 726-735, Mar. 2005.
[3. 17] M. Perrott, *High Speed Communication Circuits and Systems*, MIT Open Courseware, Lecture 7, 2003.
[3. 18] Lai Kan, Leung, "Design and integration of a single-chip 1-V CMOS IEEE 802.11a Transceiver," Ph.D.dissertation, Hong Kong University of Science and Technology, 2006.
[3. 19] H. Samavati, H. R. Rategh, and T. H. Lee, "A 5-GHz CMOS wireless LAN receiver front end," *IEEE J. Solid-State Circuits*, Vol. 35, No. 5, pp. 765-772, May, 2000.
[3. 20] *Expressions, Measurements, and Simulation Data Processing*, Agilent Technologies, May 2003.
[3. 21] S. Kang, B. Choi, and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Trans. Microwave Theory and Techniques*, pp. 972-977, Mar. 2003.
[3. 22] W. Sansen, "Distortion in elementary transistor circuits," *IEEE Trans. Circ. & Syst., Part II, Analog and Digital Signal Processing*, Vol. 46, No. 3, pp. 315-325, Mar. 1999.

[3. 23] K. Kimura, “Some circuit design techniques using two cross-coupled, emitter-coupled pairs,” *IEEE Trans. Circ. & Syst., Part I*, Vol. 41, No. 5, pp. 411-423, May 1994.

[3. 24] Y. Ding, and R. Harjani, “A +18dBm LNA in 0.35μm CMOS,” *IEEE International Solid-State Circuits Conference*, Feb. 2001.

[3. 25] 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*, pp. 223-229, Jan. 2004.

[3. 26] K. Yamauchi, K. Mori, M. Nakayama, Y. Mitsui, and O. Ishida, “A novel series diode linearizer for mobile radio power amplifiers,” *IEEE MTT-S Int. Microwave Symp. Dig.*, pp 831-834, 1996.

[3. 27] M. Bao, Y. Li, and A. Cathelin, “A 23 GHz active mixer with integrated diode linearizer in SiGe BiCMOS thchnology,” *33rd European Microwave Conf.*, Vol. 1, pp. 391-393, Oct. 2003.

[3. 28] Y.-J, Jeon, H.-W. Kim, M.-S. Kim, Y.-S. Ahn, J.-W. Kim, J.-Y. Choi, D.-C. Jung, and J.-H. Shin, “Improved HBT linearity with a “post-distortion”-type collector linearizer,” *IEEE Microwave and Wireless Components Lett.*, Vol. 13, No. 3, pp. 102-104,Mar. 2003.

[3. 29] C. H. Kim, C. S. Kim, H. K. Yu, M. Park, and D. Y. Kim, “RF active balun circuit for improveing small-signal linearity,” U.S. Patent 6 473 595, Oct. 29, 2002.

[3. 30] C. Toumazou, F. J. Lidgey, and D. G. Haigh, *Analogue IC Design: the Current-Mode Approach*, pp. 193-205, Peter Peregrinus Ltd., London, UK, 1990.

[3. 31] Y. Hu, J. C. Mollier, J. Obregon, “A new method of third-order intermodulation reduction in nonlinear microwave systems,” *IEEE Transactions on Microwave Theory and Techniques*, pp. 245-250, Feb. 1986.

[3. 32] K. K. Jeon, E. K. Kim, Y. Kim, “An analog linearizer using second harmonic signal mixing operation,” *Asia-Pacific Microwave Conference*, pp. 1-4, Vol. 2, Dec. 2005.

[3. 33] T. Nesimoglu, R. J. Wilkinson, C. N. Canagarajah, J. P. McGeehan, “Second harmonic zone injection for amplifier linearisation,” *Vehicular Technology Conference*, pp. 2353-2357, Vol. 3, 1999.

# Chapter 4 Wideband LNA Design

## 4. 1. Introduction

Besides narrowband LNA, wideband LNAs are used in a large variety of applications, including wireless systems, instrumentation, and optical communication. One example of wideband application is cable network [4. 1]. The down-stream from the service provider to the customer covers a wide frequency band of 50-864 MHz. In down-stream, analog television signals are in 6 MHz channels and there are approximately 130 channels.

Wideband LNAs also find application in the recently introduced ultra-wide band (UWB) systems [4. 2]. Over the last few years, these systems, initially developed for wireless personal area network (WPAN) application, have received significant attention from industry, media and academia. Theoretically, UWB systems support data rates from 110MB/s at a distance of 10 meters to 480Mb/s at a distance of 2 meters, while consuming little power. The allocated frequency band for the UWB system is 3.1-10.6 GHz. Therefore, the design of a wideband LNA covering the entire band of interest is of major concern in the development of a UWB receiver. Other applications of wideband LNAs include satellite (950-2150 MHz) and terrestrial digital (450-850 MHz) video broadcasting. Transceivers used in optical links with a high number of channels also need a wideband LNA at the front-end.

In this chapter, wideband LNA design challenges and issues will be discussed. Wideband output loading using inductive peaking is proposed to extend the bandwidth. Two wideband LNAs are designed for cable TV tuner and UWB system to demonstrate the idea.

### 4.1.1 Design Challenges

*A. Low Noise Figure*

In wideband amplifier design, noise optimization for all the frequency is hard. As introduced in Chapter 2, there exists optimum source admittance for minimum NF of a noisy device. For MOSFET device, the optimum admittance $Y_s = G_s + jB_s$ can be calculated as [4. 3]:

$$\begin{aligned} G_{s,opt} &= \alpha\omega C_{gs}\sqrt{\frac{\delta}{5\gamma}(1-|c|^2)} \\ B_{s,opt} &= -\omega C_{gs}\left(1+\alpha|c|\sqrt{\frac{\delta}{5\gamma}}\right) \end{aligned} \tag{4. 1}$$

The optimum source susceptance $B_{s,opt}$ is inductive in character, except that it has the wrong frequency behavior. Hence, achieving a broadband noise match is difficult.

*B. Wideband Input Matching*

Wideband input impedance matching network can be either lossless or lossy. For a lossless matching, the matching network is composed of reactive components, and energy is ideally conserved. As a result, there exists a limit on the maximum bandwidth over which an arbitrarily good impedance matching can be obtained [4. 4]. For a lossless network to match a parallel-RC load impedance, the Bode-Fano criterion states that

$$\int_0^\infty \ln\frac{1}{|\Gamma(\omega)|}d\omega \le \frac{\pi}{RC} \tag{4. 2}$$

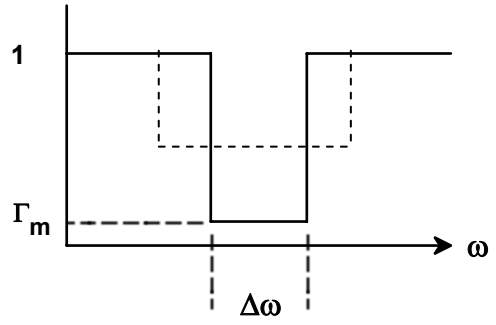



Fig. 4. 1 Illustrating the Bode-Fano criterion

As depicted in Fig. 4. 1, where Γ(ω) is the reflection coefficient looking into the lossless matching network, Bode-Fano limit implies that the bandwidth of impedance matching is limited (by $\pi/RC$), one should not waste any match out-of-band. The best in-band match is obtained with Chebychev network rather than maximally flat networks. There are similar limitations on other forms of complex load impedance, but the general nature of the limitation is the same.

On the other hand, for lossy matching, due to their resistive components, the energy in the matching network is not conserved, and there exists no limit for the good wideband matching. The matching network can also be much simpler with better matching over process variation. Two lossy matching networks are shown in Fig. 4. 2. The matching networks can match the complex loading to $R_s$ over all the frequencies. However, NF is large for lossy matching because resistive components significantly contribute to noise.

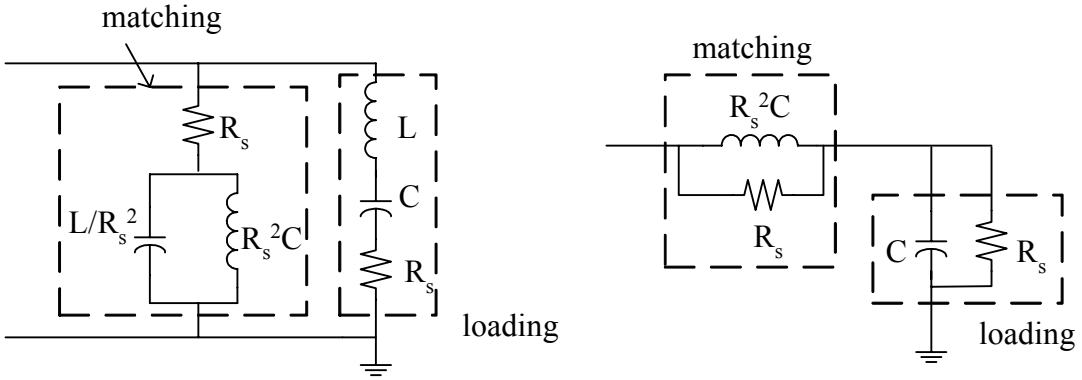



Fig. 4. 2 Two lossy matching networks

*C. Wideband Amplifier Output Loading*

The inherent parasitic capacitors of MOSFET devices and wiring in layout are the main cause of bandwidth limitation in wideband amplifiers. To achieve large gain-bandwidth product (GBW), wideband loading is required for bandwidth enhancement.

### 4.1.2 Wideband LNA Topologies and Techniques

The topology for wideband LNA is mainly based on input impedance matching consideration. Except resistive termination, common-gate and resistive feedback topologies introduced in Chapter 3 that can be used in either narrowband or wideband, LC ladder filter matching is another popular topology.

#### 4.1.2.1 LC Ladder Filter Matching

LC ladder filter matching is one kind of lossless matching, as shown in Fig. 4. 3. It is the wideband version of inductive degeneration LNA. Degeneration inductor $L_s$ is used to provide real part of input impedance $(g_{m1}/C_{gs})L_s$. In the pass-band frequency of ladder filter, the input impedance is matched to source impedance.

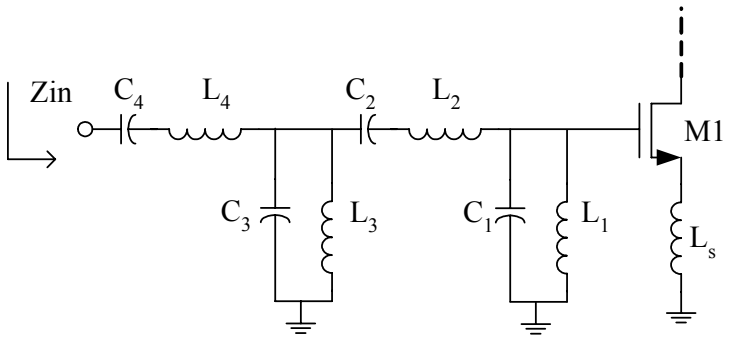



Fig. 4. 3 Schematic of LC ladder filter matching LNA

This kind of matching is implemented in [4. 5] to simultaneously achieve wideband matching, low NF, and low power consumption. However, source-degeneration inductors need to be employed, which would lead to gain degradation and rapid NF increase at high frequencies [4. 6]. Moreover, many on-chip inductors in the matching network would occupy very large chip area.

#### 4.1.2.2 Capacitive Cross-Coupling Common-Gate LNA [4. 6]

An improved version of common-gate input stage is proposed in [4. 6], as shown in Fig. 4. 4. The two differential common-gate transistors M1 and M2 are cross-coupled between source and gate. If the coupling capacitors $C_1$ and $C_2$ are much

larger than $C_{gs}$ of M1 and M2, they work as short at operating frequency. Therefore, driven by differential input signal, M1 and M2 get double input signal across source and gate compared to simple common-gate topology, which makes effective transconductance doubled.

The capacitive cross-coupling common-gate topology is effective in reducing noise and power consumption from active devices, because of transconductance doubling. In addition, common-gate stage in general is immune to induced gate noise, and noise performance is stable across large frequency range, whereas for inductive degeneration topology, NF can be optimized to low level at certain frequency, but will increase rapidly ($\propto \omega_0 / \omega_T$) at high frequency, induced gate noise contributes significant noise so $Q$ of input matching network needs to be designed carefully to tackle this problem [4. 6]. Thus, capacitive cross-coupling common-gate stage is a good candidate for high-frequency wideband operation.

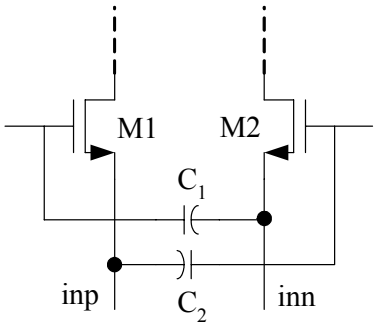



Fig. 4. 4 Capacitive cross-coupling common-gate input stage

#### 4.1.2.3 Noise Canceling LNA [4. 7]

Thermal noise canceling technique was first proposed in 2002. Some circuit topologies are found to be able to cancel noise from a specific device, or an auxiliary noise canceling path is added to cancel particular device's noise.

One circuit topology is shown in Fig. 4. 5. Consider thermal current drain noise of M1, due to feedback through resistor R, it causes two instantaneous noise voltages at nodes X and Y, which have equal signs. On the other hand, the signal voltage at nodes X and Y have opposite signs because the gain is negative. This difference in sign for noise and signal makes it possible to cancel the noise of M1, while simultaneously adding the signal contributions constructively. The adding function is implemented with M2 and M3.

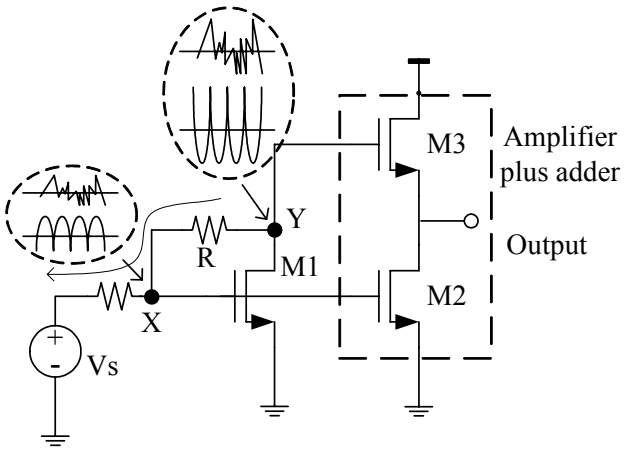



Fig. 4. 5 Wideband LNA exploiting noise canceling

Another possible noise canceling topology is shown in Fig. 4. 6. M1 is a common-gate input matching transistor, and its thermal noise can be canceled if the requirement $R_1 = g_{m2} R_2 R_s$ is met.

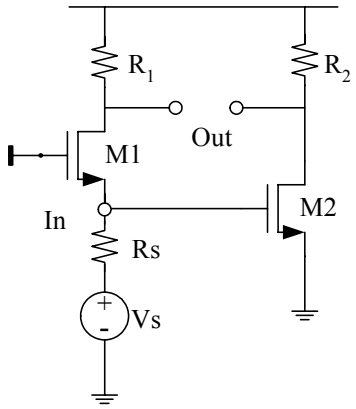



Fig. 4. 6 Alternative noise canceling implementation

There are a few limitations for noise canceling scheme. The noise can be canceled for only one specific device, because noises from different noise sources are uncorrelated and impossible to cancel; more devices are added to cancel the main noise, but the added devices introduce new noise and consume power; gain and phase responses for main and auxiliary paths need to be matched to get better cancellation in wideband application, which is difficult and requires large power consumption for wideband auxiliary path.

#### 4.1.2.4 Distributed Amplifier

Distributed amplifiers (DA) (also known as traveling wave amplifiers) employ an architecture in which several active devices are connected in parallel [4. 8]. The output current of individual amplifiers combines in an additive fashion, and this dictates a relatively low gain for this architecture. The advantage of DA comes from the fact that the input capacitances of these amplifiers are absorbed in an LC network which forms an artificial transmission line and allows the realization of amplifiers with large bandwidth.

The large power consumption of most of the DAs is a major problem, which makes it difficult to be integrated for low-power wideband applications.

## 4. 2. Wideband LNA Design for Cable TV Tuner

The OpenCable standard allocates up to 137 analog and digital channels with channel spacing of 6 MHz and a bandwidth of approximately 4.2 MHz per channel. The required tuner needs to handle signals with input frequencies from 50 MHz to 864 MHz. The signal level per channel is specified to be between -25 and +5 dBmV. Consequently, a wideband LNA is required at the front-end to amplify the signals and to suppress the noise contribution from the latter stages with acceptable linearity performance.

The key design challenge for the LNA is to provide sufficiently high and reasonably flat gain within the required bandwidth, the gain needs to be tunable by 20 dB for different levels of input signal power. Moreover, the input impedance must be well matched to 75-Ω over the whole frequency range with reasonable NF. Lastly, the LNA must have good linearity in order to accommodate many strong interference signals without much distortion. The specification of cable TV LNA is summarized in TABLE 4. 1 [4. 9].

TABLE 4. 1 SPECIFICATION OF CABLE TV LNA

| | |
|---|---|
| Gain/dB | 0~22 |
| NF/dB | <3 @ 22 dB gain setting |
| IIP3/dBm | 5 @ 10 dB gain settting |
| S11/dB | <-10 |
| Power/mW | 40 |

The frequency band of cable TV is very wide ($f_U/f_L$ = 864 MHz/50 MHz ≈ 17), but the whole frequency range is low (below 900 MHz), based on these observations, one good topology for impedance matching is resistive termination or with resistive termination in feedback loop. For common-gate input stage, because $g_m$ is set by input impedance requirement, gain and impedance cannot be independently adjusted, which is not convenient in real application, so resistive feedback is finally adopted in the LNA design.

### 4.2.1 Active Common-Gate Feedback LNA

Active common-gate feedback [4. 10] utilizes active device as resistive component, as shown in Fig. 4. 7. M4 works as common-gate feedback transistor to provide real part of input impedance. M1 is the main amplification transistor, and M3 provides the biasing for M4.

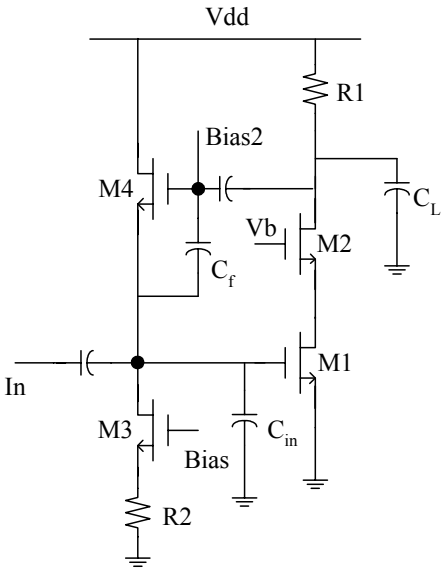



Fig. 4. 7 Schematic of active feedback impedance matching

*A. Input Impedance*

We can calculate input impedance of Fig. 4. 7 from feedback theory:

Input loading of the feedback network: $(1/g_{m4})$ // $(1/sC_f)$.

Output loading of feedback network: $1/sC_f$.

Feedback factor: $-(g_{m4}+sC_f)$.

The open-loop input impedance can then be written as:

$$Z_{in}^{'} = R_s //(1/g_{m4})//(1/(sC_f + sC_{in})) \tag{4. 3}$$

while the close-loop input impedance is:

$$Z_{incl} = \frac{R_s //(1/g_{m4})//(1/(sC_f + sC_{in}))}{1+(g_{m4}+sC_f)g_{m1}\left[R_s //(1/g_{m4})//(1/(sC_f + sC_{in}))\right]\cdot\left[R_1 //(1/(sC_f + sC_L)\right]} \tag{4. 4}$$

Eliminating the contribution of source impedance $R_s$ to input impedance, input impedance can be written as:

$$Z_{in} = \frac{R_s \cdot Z_{incl}}{R_s - Z_{incl}}$$

$$= \frac{R_s}{R_s\left\{1/\left[R_s //(1/g_{m4})//(1/(sC_f + sC_{in}))\right] + (g_{m4} + sC_f)g_{m1}\left[R_1 //(1/(sC_f + sC_L)\right]\right\} - 1} \quad (4.5)$$

$$= \frac{1}{g_{m4} + s(C_f + C_{in}) + (g_{m4} + sC_f)g_{m1}\dfrac{R_1}{1 + sR_1(C_f + C_L)}}$$

If we ignore the effect of parasitic capacitors, (4. 5) becomes:

$$Z_{in} = \frac{1}{g_{m4}(1 + g_{m1}R_1)} \quad (4.6)$$

which coincides with our intuitive feeling that effective transconductance of M4 is boosted by the feedback loop, therefore, the required M4 transconductance for impedance matching is effectively reduced, and so is the noise contribution from M4.

If the amplifier gain ($g_{m1}R_1$) is large, from (4. 5),

$$Z_{in} \approx \frac{1 + sR_1(C_f + C_L)}{g_{m4}g_{m1}R_1} \quad (4.7)$$

which means, to maintain good impedance matching with small imaginary part at high frequency, one needs to minimize $R_1C_L$.

From the above calculation, it is realized that the loading effect of the feedback to the amplifier output is purely capacitive, as compared to resistive feedback, MOSFET active feedback will not reduce the amplifier voltage gain (from *in* to *out* in Fig. 4. 7) by loading the output impedance, but only affects the output bandwidth, and this effect can be minimized by controlling the feedback transistor size. In addition, feedback transistor's transconductance (and thus equivalent resistance) can be tuned conveniently, which well suits the purpose of having a variable feedback resistor for different gain settings to maintain a fixed input impedance. The drawback of implementing resistor with active device is the active device's nonlinearity.

*B. Gain Tuning*

Gain tuning is required in LNA to handle large range of input power. In the LNA design, the nominal gain is set at 17.5 dB. Negative $g_m$-cell is connected at the output to increase gain to 22 dB by increasing the output impedance, the working principle is similar to that discussed in Chapter 3. Current steering is used to reduce the gain in discrete steps to 15 dB, 10 dB, and 5 dB. A bypass switch, implemented by NMOS transistor, is connected in parallel between LNA input and output to bypass it when input signal is large. If the nominal gain is set at maximum gain of 22 dB, the loading resistor $R_1$ needs to be large, which is not good for both output bandwidth and input impedance matching, as explained in (4. 7).

The current steering schematic is shown in Fig. 4. 8. M1 is the amplification transistor, M2 works as cascode transistor and M3 steering transistor. The drain of M3 is connected to Vdd, so part of the signal from M1 through M3 does appear at output, thus the gain is reduced. The gain of the amplifier can be written as:

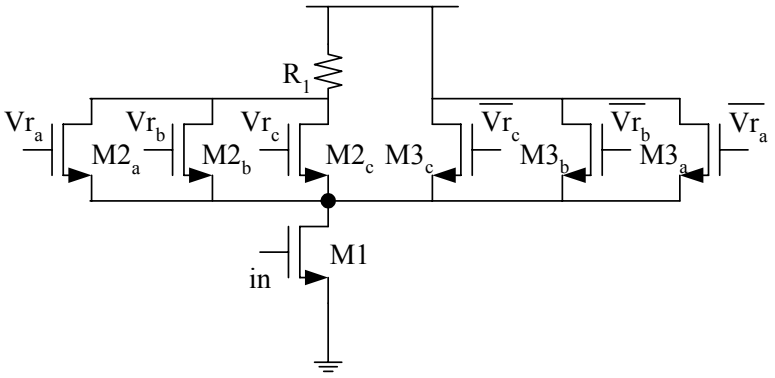



Fig. 4. 8 Current steering pair M2 and M3

$$A_v = \frac{g_{m2}}{g_{m2} + g_{m3}} g_{m1} R_1 \tag{4. 8}$$

$g_{m2}$ and $g_{m3}$ can be varied by changing M2 and M3's bias voltage $V_{gs}$, or changing transistor size of M2 and M3 while fixing their $V_{gs}$. In the first approach, gain tuning is continuous, but linearity is bad at low-gain setting because cascode transistor M2 is

not biased well and introduces significant nonlinearity. With steering transistor M3 being turned on, the current redistributes between M2 and M3. It may be that only little current flows through the cascode transistor M2. For example, if M2 and M3 are equally sized, and we want to reduce gain by 12 dB (1/4), following (4. 8), $g_{m3}$ = 3*$g_{m2}$, and $I_{M3}$ = 9*$I_{M2}$ (if the square-law I-V characteristic is assumed). Only 10% of the total current flows through M2 which makes it near cutoff region, so the scheme of changing $V_{gs}$ of M2 and M3 cannot guarantee good linearity if gain tuning range is large.

The alternative way of varying $g_m$ is to change the transistor size, as shown in Fig. 4. 8. $V_r$ and $\overline{V_r}$ are the complementary control voltage for M2 and M3, their values can be either $|V_r|$ or 0 to turn on or off the controlled transistor. $M2_a$ and $M3_a$, $M2_b$ and $M3_b$, … are equal-size. In this $g_m$ tuning scheme, the transistor width ratio $W_{M2} : W_{M3}$ can be programmably controlled while the current density through M2 and M3 remains fixed, because the overall transistor width of M2 and M3 that is turned on is fixed. Therefore, linearity of the amplifier will not be limited by the cascode transistor M2 due to small current density through M2 in low-gain setting. This scheme can only achieve discrete-step gain-tuning and is adopted in the LNA design.

*C. Noise Figure*

The feedback loop in Fig. 4. 7 forms a noise suppression loop which partially suppresses the thermal noise from M1. If we ignore the effect of parasitic capacitors and refer each part's noise to LNA input, the noise contributed by M1 is:

$$F_{M1} = \frac{4kT\gamma g_{d0,1}}{4kTR_s}\left(\frac{1+g_{m4}R_s}{(g_{m4}R_s)^2+g_{m4}R_s+1}\right)^2\frac{4}{g_{m1}^2} \qquad (4.\ 9)$$

When the amplifier gain is large so that the required $g_{m4}$ is small based on (4. 6), the noise suppression effect on M1 becomes negligible because feedback factor ($g_{m4}$) is small.

Noise contributed by feedback transistor M4 is:

$$F_{M4} = \frac{4kT\gamma g_{d0,4}}{4kTR_s}\left(\frac{1}{1/R_s + (g_{m1}R_1 + 1)g_{m4}}\right)^2 * 4 = \gamma g_{d0,4}R_s \qquad (4.\ 10)$$

Due to feedback, transconductance of M4 is effectively reduced, so is its noise contribution. Noise contributed by the current source composed of R2 and M3 is:

$$F_{R2,M3} = 4\left(\frac{4kT\gamma g_{d0,3}}{4kTR_s}\left(\frac{R_s/2}{1+g_{m3}R_2}\right)^2 + \frac{4kT}{4kTR_sR_2}\left(\frac{g_{m3}R_2R_s/2}{1+g_{m3}R_2}\right)^2\right) = \frac{\gamma g_{d0,3}R_s + g_{m3}^2R_2R_s}{(1+g_{m3}R_2)^2} \qquad (4.\ 11)$$

(4. 11) shows that larger R2 helps to reduce noise from current source part (M3 and R2). The overall NF of the LNA can then be written as:

$$F = 1 + \frac{4\gamma g_{d0,1}}{R_s g_{m1}^2}\left(\frac{1+g_{m4}R_s}{(g_{m4}R_s)^2 + g_{m4}R_s + 1}\right)^2 + \gamma g_{d0,4}R_s + \frac{\gamma g_{d0,3}R_s + g_{m3}^2R_2R_s}{(1+g_{m3}R_2)^2} + \frac{4}{R_1R_sg_{m1}^2} \qquad (4.\ 12)$$

### 4.2.2 Circuit Implementation

As shown in Fig. 4. 9, the LNA is a common-source amplifier with active feedback for input matching, and with inductive-peaking output load for wide bandwidth. M1 is the main amplification transistor. Transconductance of M4 sets the input impedance of the LNA, as analyzed before.

M4 is very small, which contributes only 25 fF capacitive load to the output and draws only 0.3 mA current at nominal gain setting (17.5 dB). Compared to resistive feedback, the loading from the active device is capacitive and is extremely large at DC (and even at 900 MHz), while resistive feedback significantly reduces gain due to excessive loading effect, in addition, it is convenient to control input impedance by

tuning $g_{m4}$ in order to compensate for the process variation, or to satisfy the matching condition for different gain settings, which cannot be done with resistive feedback.

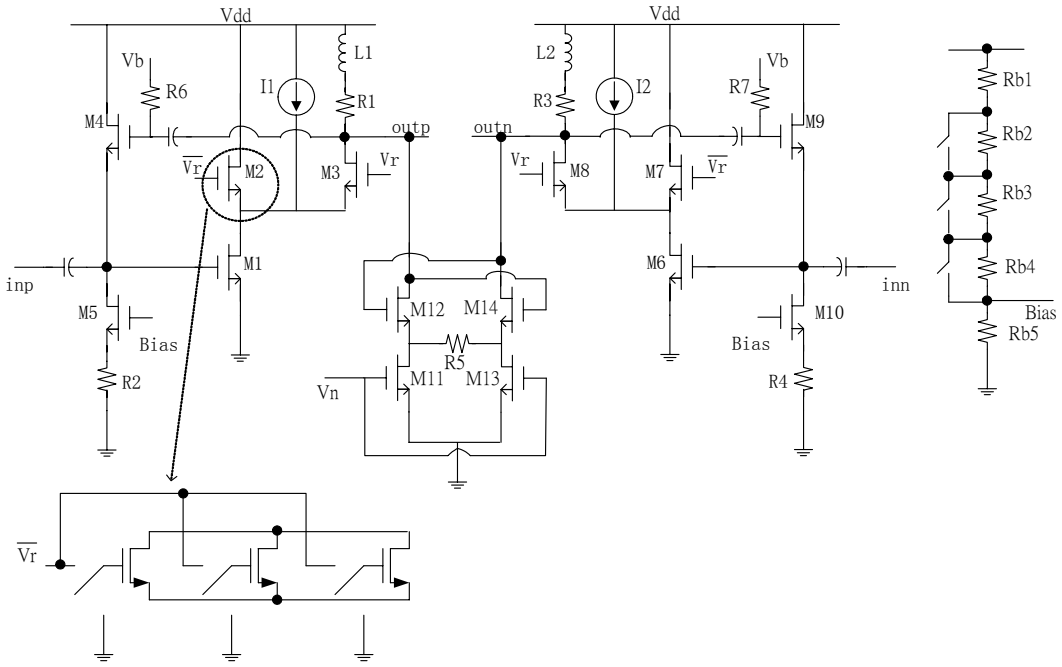



Fig. 4. 9 Schematic of the cable TV tuner LNA

The required transconductance for the feedback transistor M4 is reduced by factor $(1+A_v)$ compared to a common-gate configuration. Therefore, the noise contribution of the feedback transistor is effectively reduced. Transistor M5 together with resistor R2 acts as a bias current source for the feedback transistor M4. Large R2 would reduce noise current generated by the bias current source, as shown in (4. 11). However, too large R2 would also reduce the voltage headroom for M5 and make it in linear region with more noise contribution. As a compromise, a 200-Ω resistor is chosen in this design.

After the noise from impedance matching component M4 is suppressed, the noise from M1 becomes significant, and needs large $g_m$ (large current) to reduce its noise, as suggested by (4. 12). However, due to limited voltage headroom, large current can

not be accepted. Current sources I1 and I2 of 5 mA are added to share part of the current through the loading resistor.

The load resistor R1 is 140-Ω, and the total capacitive load (from the next mixer stage and from the parasitic) is around 1 pF. The -3 dB bandwidth would be limited to around 1.1 GHz, which would result in large gain variation within the bandwidth. In addition, the gain variation would cause large variation in the input impedance, and result in poor impedance matching within the whole frequency band. To enhance the LNA bandwidth to reduce the gain variation, inductive shunt peaking [4. 3] is employed by connecting a 7 nH inductor L1 in series with the resistive load R1. The gain flatness is reduced to within 0.7dB for the low and moderate gain settings and to around 1.4dB for the maximum gain setting.

The quality factor of the peaking inductor L1 is not important so its metal width, 5-μm is set just enough for maximum current density, and as a result, the layout is very compact even with a very large inductance value. Moreover, the two differential inductors at the two differential output nodes are combined into one center-tapped coil to reduce the chip area.

Programmable current steering pairs M2 and M3 are used for gain reduction (only M2 is programmable). The transistors used for the differential negative $g_m$-cell are sized to get just enough gain without adding much capacitive load. A 120-Ω degeneration resistor R5 is included in the negative gm cell to improve its linearity. Too large degeneration resistor will limit the gain tuning range and degrade NF. Bypass switches are added between LNA input and output to facilitate bypass-mode (not shown in Fig. 4. 9), and the LNA is turned-off in this mode. To achieve input impedance matching at different gain settings, a resistive network composed of Rb1, … Rb5 is implemented. Its output voltage will switch according to the gain

switching to set the bias current for the feedback transistor M4 and therefore controls its transconductance $g_{m4}$.

The main design parameters are summarized in TABLE 4. 2, and simulated noise contribution from each part at nominal gain setting is summarized in TABLE 3. 3.

TABLE 4. 2 MAIN DESIGN PARAMETERS IN THE CABLE TV LNA

| | |
|---|---|
| Input transistor size | 250-μm/0.18-μm |
| Cascode transistor size | 50-μm/0.18-μm |
| Steering transistor size | 30-μm/0.18-μm |
| | 80-μm/0.18-μm |
| | 180-μm/0.18-μm |
| Feedback transistor size | 10.8-μm/0.18-μm |
| Bypass switch transistor | 50-μm/0.18-μm |
| Current source M5 | 10.8-μm/0.18-μm |
| R2 | 200-Ω |
| Loading resistor | 140-Ω |
| Current sources I1 and I2 | 5 mA |
| Negative $g_m$ transistor size | 16.2-μm/0.18-μm |
| Degeneration resistor in negative $g_m$-cell | 120-Ω |
| Shunt-peaking inductor | 7 nH |

TABLE 4. 3 NOISE CONTRIBUTION FROM EACH PART OF THE LNA AT 500 MHZ

| | |
|---|---|
| Input source resistance | 70.1% |
| Input transistor | 14.0% |
| Feedback transistor | 6.3% |
| Current source M5 | 3.5% |
| Loading resistor R1 | 2.3% |
| R2 | 1.8% |
| Current source I1 | 1.2% |
| Cascode transistor | 0.4% |
| Others | 0.4% |
| Total | 100% |

### 4.2.3 Measurement Results

Fabricated in 0.18-μm CMOS process, the micrograph of the LNA is shown in Fig. 4. 10. It occupies chip area of 0.57*0.97 $mm^2$.

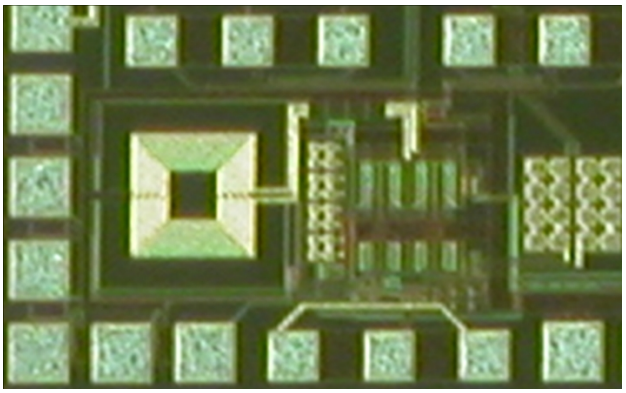


Fig. 4. 10 Micrograph of cable TV LNA

The input impedance of the LNA matches well to 75-Ω over the whole frequency range from 50 MHz to 864 MHz with the gain tunable from 0 dB to 22 dB, as shown in Fig. 4. 11 and Fig. 4. 12. Fig. 4. 13 shows the measured NF plot. The NF is achieved to be 2.8 dB at 22 dB gain setting and IIP3 is 5 dBm at 11 dB gain setting. The amplifier consumes a total current of 23 mA from a 1.8-V supply. A performance comparison with other wideband LNAs is listed in TABLE 4. 4.

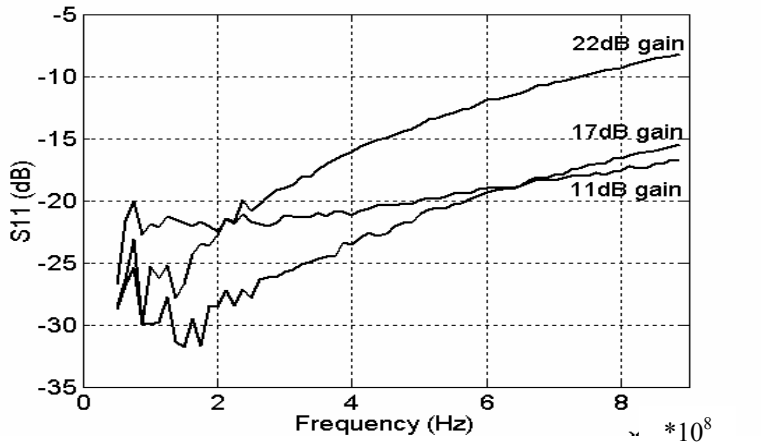



Fig. 4. 11 Measured input impedance matching

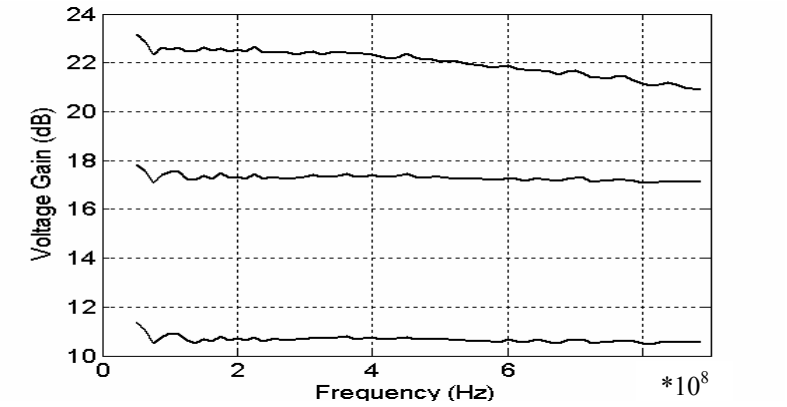



Fig. 4. 12 Measured variable gain

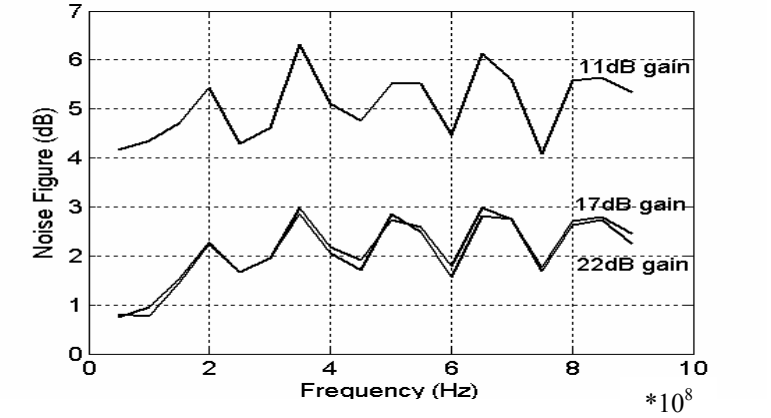



Fig. 4. 13 Measured NF at different gain settings

TABLE 4. 4 COMPARISON WITH OTHER WIDEBAND LNAS

| | [4. 11] | [4. 10] | [4. 12] | This work |
|---|---|---|---|---|
| Input Freq./MHz | 48~860 | 1000~5000 | 48~860 | 50~860 |
| NF/dB | 6 | 3.3@1GHz | NA | 4.2~6@ 11dB gain |
| Gain/dB | 6 | 13.1 | -10~14 | 0~22 |
| IIP3/dBm | 21.25 | 4.7@2.45GHz | 11@2dB gain | 4.3~5.0@ 11dB gain |
| Power/mW | NA | 75 | NA | 41.4 |
| Process | 0.5μm BiCMOS | 0.18μm CMOS | 0.35μm CMOS | 0.18μm CMOS |
| Supply Voltage/V | 3.3 | NA | 5 | 1.8 |
| Topology | NA | Transistor feedback | Common-gate | Transistor feedback |

## 4. 3. Wideband LNA Design for UWB Receiver

Ultra-wideband (UWB) systems are a new wireless technology capable of transmitting data over a wide spectrum of frequency bands with very low power and high data rates. Among the possible applications, UWB technology may be used for imaging systems, vehicular and ground-penetrating radars, and communication systems. In particular, it is envisioned to replace almost every cable at home or in an office with a wireless connection that features hundreds of megabits of data per second [4. 13].

The UWB system utilizes the unlicensed 3.1 – 10.6 GHz frequency band with a transmit output power below the FCC limit of -41.25 dBm/MHz. According to Multi-Band OFDM Alliance SIG proposal, the UWB spectrum is divided into 14 bands, each with a bandwidth of 528 MHz [4. 2]. Five band groups are defined, consisting of four band groups of three bands each and one band group of two bands, as shown in Fig. 4. 14. Support for the first band group shall be mandatory.

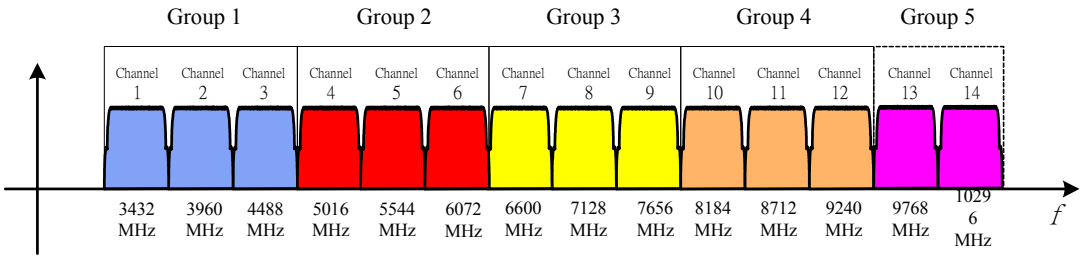



Fig. 4. 14 UWB band groups allocation

A wideband receiver front-end is designed to cover the first 9 UWB bands, which requires a wideband LNA operating from 3.1 GHz to 8 GHz. The LNA must feature wideband input matching to an impedance of 50-Ω for maximum power transfer and for out-of-band interferer filtering. Moreover, it must have a relatively flat gain over the entire bandwidth while achieving minimum NF and low power consumption. To

suppress the noise contribution from the latter stages, the LNA is required to provide maximum gain of larger than 22 dB. On the other hand, at maximum input signals, the LNA with reduced gain is preferred to relax the linearity requirement of the following building blocks. As a consequence, the LNA is also expected to have a variable gain from 12 dB to more than 22 dB. In addition, since the LNA's NF dominates the receiver's NF, its NF must be smaller than 5.5 dB. Wideband circuit techniques with T-coils and inductive series peaking are employed in the LNA to extend the bandwidth up to 8.0 GHz.

As the LNA is designed to operate over a wide frequency band from 3.1 to 8 GHz, parasitic capacitors impose great challenges in achieving both wideband input matching and large gain-bandwidth product (GBW) while maintaining reasonably small NF and low power consumption.

### 4.3.1 Circuit Implementation

*A. Input Impedance Matching*

To eliminate source-degeneration inductors, as introduced before, a lossy matching input stage, capacitive cross-coupling common-gate stage [4. 6], is adopted, as shown in Fig. 4. 15. The input differential pair is cross-coupled. Large capacitors $C_1$ and $C_2$ work as short at RF frequency, and by doing so, the effective transconductance becomes $2*g_m$, where $g_m$ is the transconductance of $M_1$ ($M_2$). For 50-Ω matching requirement, $g_m$ is set to be 10 mS, thus for the first stage, the power consumption is reduced by half. On-chip RF choke inductors $L_5$ and $L_6$ of 4 nH provide DC path for $M_1$ and $M_2$, and resonate with parasitic capacitors $C_p$ at input terminal at high-frequency band, leading to better impedance matching.

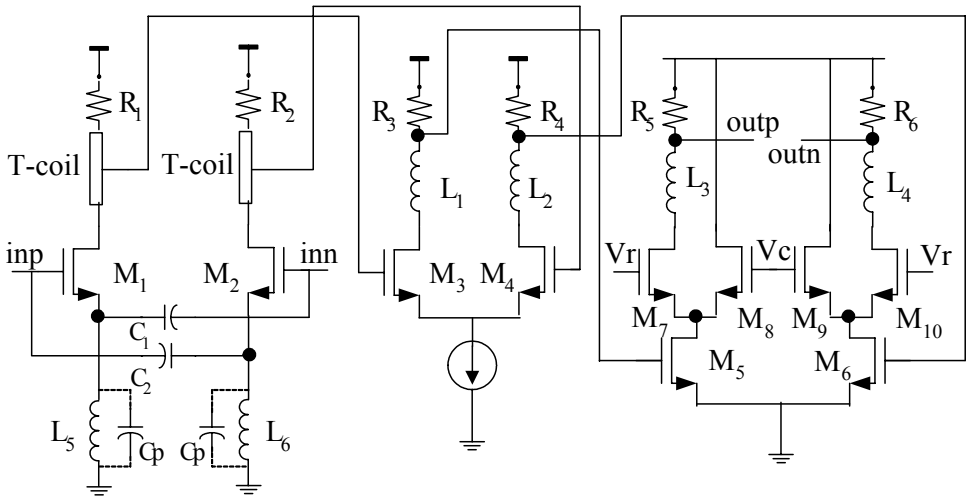
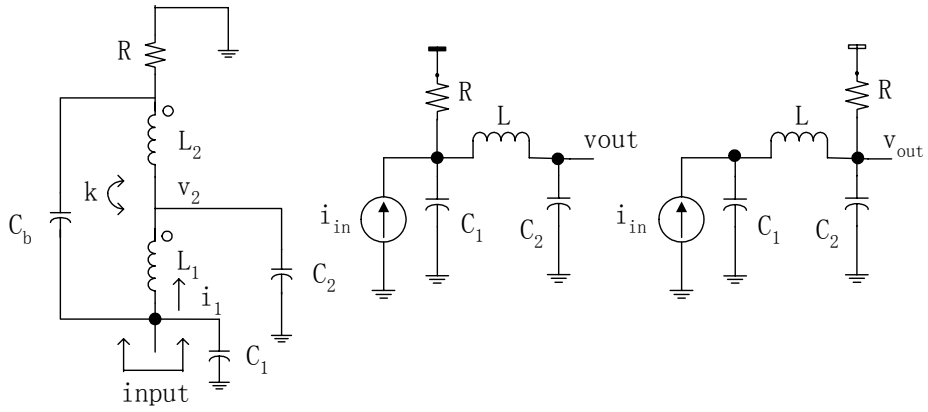



Fig. 4. 15 Schematic of the proposed 3-stage wideband LNA with variable gain

*B. Wideband Output Loading*

To enable the LNA to operate with a wide bandwidth, T-coil and inductive series peaking are employed as the output loadings. Fig. 4. 16 (a) shows the schematic of the T-coil. If the load capacitor $C_2$ is much larger than the parasitic capacitance $C_1$ at the T-coil's input port, the T-coil is capable of almost tripling the bandwidth [4. 3].




(a) (b) (c)

Fig. 4. 16 Schematics of (a) T-coil, (b) Type I inductive series peaking, (c) Type II inductive series peaking

For T-coils without parasitic capacitor $C_1$ as in Fig. 4. 16 (a), the design parameters are given by [4. 14]:

$$L_{1,2} = \frac{R^2 C_2}{4}(1 + \frac{1}{4\zeta^2}) \tag{4. 13}$$

$$coupling\ coefficient = \frac{4\zeta^2 - 1}{4\zeta^2 + 1} \tag{4. 14}$$

$$C_b = \frac{C_2}{16\zeta^2} \tag{4. 15}$$

where $\zeta$ is the damping factor of $v_2/i_1$.

$$\frac{v_2}{i_1} = \frac{R\omega_0^2}{s^2 + 2\zeta\omega_0 s + w_0^2}, \text{ with } \omega_0 = \frac{4\zeta}{RC_2} \tag{4. 16}$$

T-coil's transimpedance with parasitic capacitor $C_1$ can be expressed as

$$Z = \frac{R\omega_0^2}{s^2 + 2\zeta\omega_0 + \omega_0^2}\frac{1}{1 + sRC_1} \tag{4. 17}$$

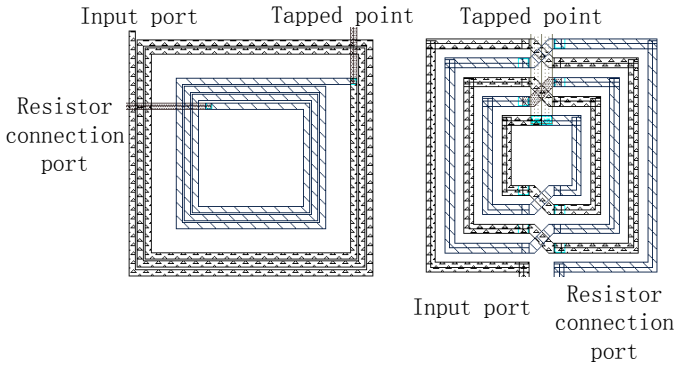
Set $|Z(s)/Z(0)| = \sqrt{2}/2$ and $C_1 = kC_2,\ C_2 = 1,\ R = 1,\ w = \omega^2$ to calculate -3-dB frequency, we have

$$k^2 w^3 + (64\zeta^4 k^2 - 32\zeta^2 k^2 + 1)w^2 + 32\zeta^2(2\zeta^2 + 8\zeta^2 k^2 - 1)w - 256\zeta^4 = 0 \tag{4. 18}$$

Plotting the curve of $w$ vs. $\zeta$ according to (4. 18) in MATLAB for a given capacitor ratio $k$, we can find maximum bandwidth $\omega$ and the corresponding $\zeta$ ($L_1$, $L_2$, $C_b$ and coupling coefficient).

The quality factor of the T-coil is not important due to termination resistor R, but the parasitic capacitance from the coil to the substrate and the coupling capacitance between the turns are. Consequently, in the layout, metal width is chosen to be thin, and the overlapping between the turns needs to be minimized. The proposed layout of the T-coil is shown in Fig. 4. 17 (a). Such a layout for the T-coil is more suitable for low coupling coefficient $k$ and adds less parasitic as compared to the conventional layout in[4. 15]. In addition, it becomes more convenient to lay out the T-coil so that $L_1 > L_2$, which is helpful for T-coil with large input parasitic capacitance $C_1$.

In conventional T-coil layout, $L_1$ and $L_2$ are interleaving. The spacing between turns needs to be large for low *k*, and $L_1$ and $L_2$ themselves cannot take advantage of the positive coupling between turns. As a consequence, the total metal trace needs to be longer, which occupies a larger area and contributes more parasitic capacitance. The proposed T-coil layout uses the minimum spacing within one inductor and large spacing between two inductors to make *k* low. As an example, for $L_1$=$L_2$=1.4-nH, *k*=0.5, with metal width w=4-μm, the proposed layout structure requires an outer dimension of 146-μm with metal trace of 1800-μm long whereas a conventional layout needs an outer dimension of 200-μm with metal trace of 2700-μm long.




(a) (b)

Fig. 4. 17 T-coil layout, (a) proposed; (b) conventional

T-coil occupies large chip area, and in practice, the parasitic $C_1$ at the T-coil's input port can be comparable to the load capacitor $C_2$, in which case inductive series peaking is a better ways to extend the bandwidth. Fig. 4. 16 (b) and (c) illustrate the two possible applications of series peaking as a wideband output loads. Parasitic capacitors $C_1$ and $C_2$ are separated by a peaking inductor L, and the load resistor R can be placed on either side of the inductor, which results in two different configurations referred to as Type I and Type II series peaking, respectively.

Type I series peaking is very useful for bandwidth extension only when $C_1$ is comparable or larger than $C_2$. For a loss-compensation L-section loading with $C_2$ being larger than $C_1$, the load resistor is moved to the larger capacitor side for better bandwidth extension [4. 16]. This idea is applied to the series peaking, yielding Type II series peaking. If $k$ is defined as the capacitor ratio $C_1/C_2$, it can be shown that the transimpedance $v_{out}/i_{in}$ of Type II series peaking is completely symmetrical with Type I if $k$ is replaced with $1/k$, therefore, different peaking types should be used for different values of $k$, and the bandwidth extension is better if the load resistor is put at the larger capacitor side. Intuitively, considering Type I peaking as an example, the inductor L and the smaller capacitor $C_2$ form a series LC tank with a resonant frequency higher than the uncompensated -3-dB bandwidth of the network, and most current flows through the low-impedance LC series path rather than the load R and the larger capacitor $C_1$. Therefore, a large voltage drops across the smaller capacitor $C_2$ with larger impedance and results in a peak in the transimpedance at high frequencies, which effectively extends the bandwidth.

For Type I peaking in Fig. 4. 16 (b), the trans-impedance $Z = v_{out}/i_{in}$ is given as:

$$Z = \frac{R}{1 + sR(C_1 + C_2) + s^2 LC_2 + s^3 RLC_1C_2} \tag{4. 19}$$

To calculate the -3-dB bandwidth, set $\left|Z(s)/Z(0)\right| = \sqrt{2}/2$ and $C_1 = kC_2, L = mR^2C_2, C_2 = 1, R = 1, w = \omega^2$ , where $m$ is a coefficient to be determined. Consequently, we have

$$m^2k^2w^3 + [-2mk(1+k) + m^2]w^2 + [(1+k)^2 - 2m]w - 1 = 0 \tag{4. 20}$$

This is a third-order function with three roots, and the maximum -3-dB bandwidth is achieved when two of the three roots merge, as illustrated in Fig. 4. 18, in which case the two merged roots satisfy (4. 21):

$$\frac{d}{dw}\{m^2k^2w^3+[-2mk(1+k)+m^2]w^2+[(1+k)^2-2m]w-1\}=0 \qquad (4.\ 21)$$

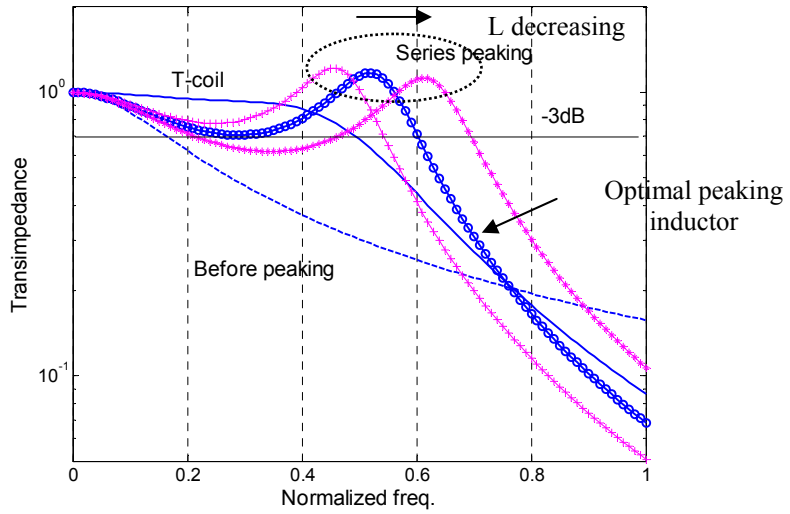



Fig. 4. 18 Frequency responses for T-coil and inductive series peaking with different peaking inductor values (the parasitic capacitor ratio $k$ is 1 for T-coil and series peaking)

Combining (4. 20) and (4. 21), the optimal value of $m$ for the maximum bandwidth extension and the corresponding bandwidth $\omega$ can be found. However, it is hard to obtain closed-form solutions. As an alternative, for a certain parasitic capacitor ratio $k$, bandwidth $\omega$ as a function of $m$ (series inductor) can be plot in MATLAB based on (4. 20), and maximum $\omega$ and corresponding $m$ can thus be found.

Fig. 4. 19 shows the maximum bandwidth extension as a function of the parasitic capacitors ratio $k$ for T-coil and two types of inductive series peaking. The maximum bandwidth extension for Type I series peaking is higher when $C_1>C_2$ and is lower when $C_1<C_2$, so it is always better to put the load resistor on the larger capacitor side. Fig. 4. 20 shows the peaking inductor values for maximum bandwidth extension. In practice, $C_1$ is usually smaller than $C_2$, because $C_1$ is from the drain parasitic capacitor and $C_2$ is from gate parasitic capacitor. However, considering that layout wiring and series peaking inductor contribute nearly equal amount of parasitic capacitors to both

$C_1$ and $C_2$, the $C_1/C_2$ ratio will not be far from 1. From Fig. 4. 19 and Fig. 4. 20, it is seen that 3.6~3.7 times bandwidth extension is readily achieved, and the corresponding peaking inductor is relatively constant over *k* range of 0.7 to 1.4. In other words, the series peaking technique if properly used can achieve good bandwidth extension with robust performance as function of *k* and with only one extra inductor.

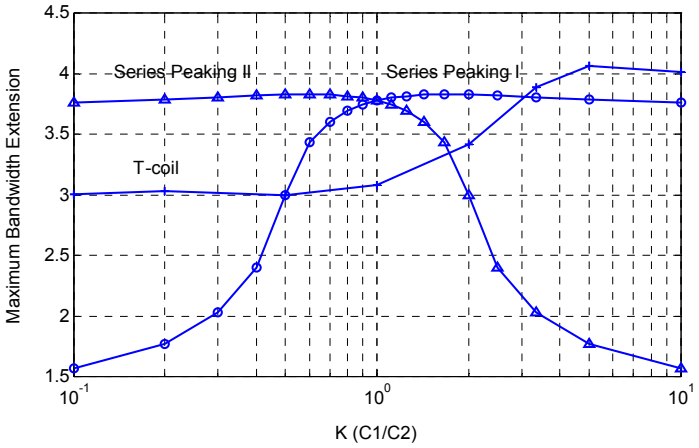



Fig. 4. 19 Bandwidth extension factor vs. parasitic capacitors ratio for T-coil and series peaking

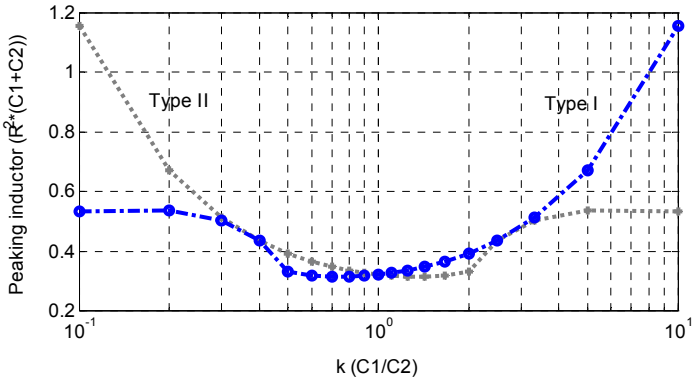



Fig. 4. 20 Peaking inductor values in series peaking networks for maximum -3-dB bandwidth extension

Compared with the T-coil, the series-peaking networks not only are simpler and easier to be modeled but also occupy less area and actually achieve larger bandwidth

extension when the two capacitors at the two ports are comparable. On the other hand, they will suffer from some peaking at high frequencies.

For Type I series peaking, the peaking in frequency response will be larger for larger capacitor ratio $k$, but the peaking would be reduced for a series inductor with finite quality factor $Q$. For example, when $k$=10, the peaking at high frequency is as large as 20.5 dB for ideal inductor. With an inductor $Q$ equal to 5 at the peaking frequency, the peaking level reduces to 2.5 dB with little reduction in bandwidth extension, as shown in Fig. 4. 21.

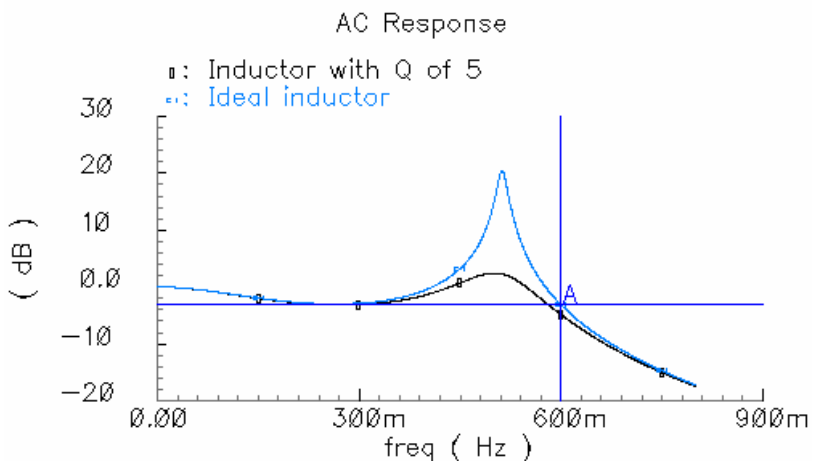



Fig. 4. 21 Series-peaking frequency response with ideal inductor and real inductor

*C. Three-Stage LNA Design*

Due to high-frequency and wideband operation, the maximum gain of a single-stage amplifier is very limited. Several gain stages can be cascaded to achieve larger *GBW* [4. 3]. Due to the fact that each stage's *GBW* is limited by the process and the power consumption, for a given overall gain $G_{overall}$, there exists an optimum number of stages $n$ to maximize the overall bandwidth. For a cascade of $n$ identical single-pole stages, the overall $GBW_{overall}$ becomes:

$$GBW_{overall} = G^n \cdot \omega\sqrt{2^{1/n}-1} = \frac{G_{overall}}{G_{overall}^{1/n}}\sqrt{2^{1/n}-1}\cdot GBW \tag{4. 22}$$

where $G$ and $\omega$ are each stage's gain and bandwidth. The optimum value of $n$ can be found when $\frac{\sqrt{2^{1/n}-1}}{G_{overall}^{1/n}}$ is maximized. Fig. 4. 22 shows normalized $GBW_{overall}$ to the single-stage $GBW$ as a function of $n$. $n$ equal to 5 is best for 22 dB gain requirement. However, a 5-stage design only achieves 1.06 times larger $GBW_{overall}$ compared to a 3-stage design, as a consequence, a 3-stage LNA is proposed as shown in Fig. 4. 15. In addition, a variable gain of more than 10 dB from 12 dB to 22 dB is implemented at the 3rd stage by current steering to accommodate for low gain requirement and thus to relax the linearity requirement for large input signals.

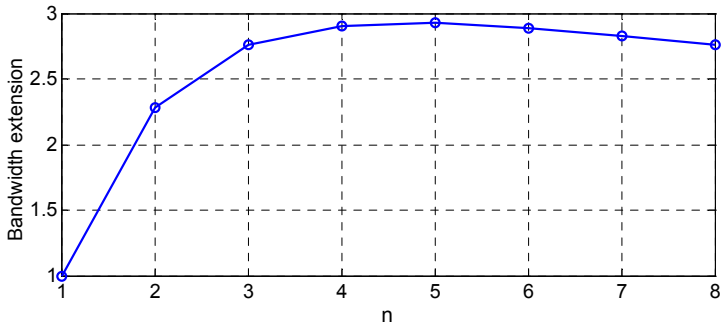



Fig. 4. 22 Bandwidth extension for $n$ single-pole identical stages relative to single-stage at 22 dB overall gain requirement

In 3-stage LNA design, each stage can only tolerate 1 dB gain drop in order to achieve a 3 dB drop for the whole LNA. With a similar procedure outlined by (4. 19) and (4. 20), the -1 dB bandwidth extension for each stage and the corresponding inductor values for Type I series peaking can be determined as plotted in Fig. 4. 23. From Fig. 4. 23, if 3 stages are cascaded together with series peaking, the -3 dB bandwidth extension is ~ 5, which is larger than that of a single stage (~3.6) and demonstrates the effectiveness of series peaking in multistage designs.

Because the first-stage LNA's transistor size is half of those in the last two stages, the parasitic capacitance at its output node is relatively small. In addition, the gain

flatness is important in order to suppress the noise contribution of the latter stages over the whole band. As a consequence, the first LNA stage uses a T-coil as its load while the last two stages use inductive series peaking instead.

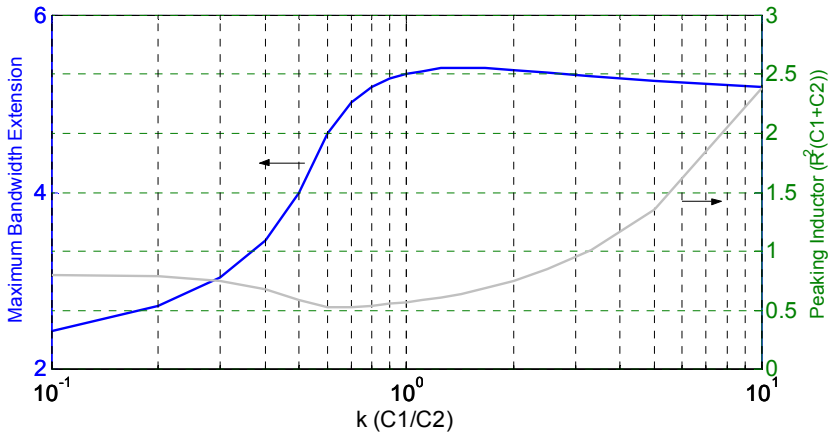



Fig. 4. 23 Maximum -1-dB bandwidth extension and the corresponding peaking inductance for Type I series peaking

Because each stage has low gain around 8 dB and because the signal phase difference at amplification transistor's gate and drain nodes deviates from 180° at high frequencies, the Miller effect of the gate-drain capacitor $C_{gd}$ is not significant. In addition, cascode transistors add parasitic capacitors at internal nodes, which cannot be compensated by peaking at output nodes. As such, cascode transistors do not help with bandwidth but would inevitably deteriorate the NF and the linearity. Consequently, cascode transistors are not adopted in all the three stages.

For a fixed gain, the load resistors in each stage are maximized while the amplifying transistors are reduced in size to maintain the operation frequency and to lower power consumption. There exists a maximum value of on-chip inductance for series peaking due to its limited self-resonant frequency, which results in a maximum value of the load resistors R due to the square relationship between the peaking inductor and the load resistor $L \propto R^2$. The series peaking inductors $L_1$, $L_2$, $L_3$, and $L_4$ in

Fig. 4. 15 are 4.2 nH with *Q* of 5 at 8 GHz and self-resonant frequency at 14.8 GHz. The load resistors for each stage are 150-Ω.

The *Q*-factor of the T-coil is not important due to the reasonably large termination resistor R, but its capacitive parasitics from the coil to the substrate and that between the turns are. Consequently, in the layout, the metal width is chosen to be as thin as 5-μm, and the overlapping between the turns is minimized.

*D. Noise Figure*

Let's first calculate NF at low-frequency, and assume RF choke inductors are ideally open for the AC signal, and AC coupling capacitors are ideally short. Under the condition $2g_{m1}R_s = 1$ (impedance matching)

$$F = 1 + \frac{\gamma}{2\alpha} + \frac{4}{2g_{m1}R_1} + \frac{4}{\alpha g_{m3} 2g_{m1}R_1^2} \tag{4. 23}$$

where the 2nd, 3rd and 4th terms represent the noise contributed from input transistor M1, 1st stage loading resistor R1, and 2nd stage input transistor M3, respectively. If we assume $\alpha=g_m/g_{d0}=3/4$, $\gamma=1.2$, $g_{m1}=10$ mS, R1=150 Ω, and $g_{m3}=20$ mS, the NF is calculated to be 3.72 (5.71 dB), noise contributed by M1, R1 and M3 is 21.5%, 35.8%, 15.9%, respectively. Loading resistor R1 contributes significant noise.

Because the LNA operates at high-frequency of 8 GHz, the RF choke inductors L5 and L6 cannot be made large due to limited self-resonant frequency. So at low-frequency of 3 GHz, RF choke inductor does not present extremely high impedance as compared to 50 Ω, e.g., 4 nH inductor has impedance 75.4 Ω at 3 GHz. Taking into account finite RF choke impedance, (4. 23) can be modified to be:

$$F = 1 + \left( \frac{\gamma}{2\alpha}\left(1 + \frac{3R_s^2}{R_s^2 + 4\omega^2 L_5^2}\right) + \frac{4}{2g_{m1}R_1} + \frac{4}{\alpha g_{m3} 2g_{m1}R_1^2} \right)\left(1 + R_s^2/(2\omega L_5)^2\right) \tag{4. 24}$$

If we assume $\left(\frac{\gamma}{2\alpha}(1+\frac{3R_s^2}{R_s^2+4\omega^2L_5^2})+\frac{4}{2g_{m1}R_1}+\frac{4}{\alpha g_{m3}2g_{m1}R_1^2}\right)=2.5$, NF is plotted in Fig. 4. 24. Without the factor $\left(1+R_s^2/(2\omega L_5)^2\right)$, the NF is 3.5 (5.44 dB), with the effect of finite RF choke inductance, NF is worse by 1 dB at low-frequency end.

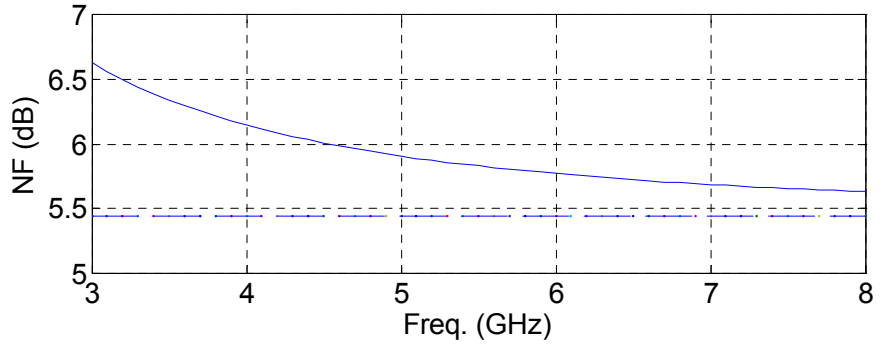



Fig. 4. 24 NF taking into accout the effect of finite RF choke inductance

As the loading resistor R1 contributes significant noise, its noise performance deserves careful investigation. Consider the loading resistor in type II series peaking, as shown in Fig. 4. 25, the noise from loading resistor R referred to input is:

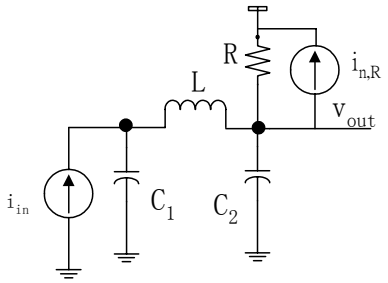



Fig. 4. 25 Loading resistor noise transfer in inductive series peaking

$$\begin{aligned}\overline{i_{n,i}^2} &= \overline{i_{n,R}^2}\cdot\left(R//\frac{1}{sC_2}//(sL+\frac{1}{sC_1})\right)^2/\left(\left[(R//\frac{1}{sC_2})+sL\right]//\frac{1}{sC_1}\right)^2\left(\frac{R//(1/sC_2)^{-1}+sL}{R//(1/sC_2)^{-1}}\right)^2 \\ &= \overline{i_{n,R}^2}\cdot\left(R//\frac{1}{sC_2}//(sL+\frac{1}{sC_1})\right)^2\cdot\left(sC_1+\frac{1+s^2LC_1}{R//(1/sC_2)^{-1}}\right)^2\end{aligned} \quad (4.25)$$

The term $\left( R // \frac{1}{sC_2} // (sL + \frac{1}{sC_1}) \right)$ represents output current to output voltage transfer $v_{out}/i_{out}$, the term $\left( sC_1 + \frac{1+s^2LC_1}{R//(1/sC_2)^{-1}} \right)$ represent output voltage to input current transfer $i_{in}/v_{out}$. The output noise current can hardly be converted to output noise voltage when L resonates with C1, which is the peaking frequency in series peaking response. Further, the transimpedance $v_{out}/i_{in}$ is bandwidth extended, i.e.,

$$\left| \left( R // \frac{1}{s(C_1+C_2)} \right) \cdot \left( sC_1 + \frac{1+s^2LC_1}{R//(1/sC_2)^{-1}} \right) \right| = \left| \frac{(1+s^2LC_1)(1+sRC_2)+sRC_1}{1+sR(C_1+C_2)} \right| < 1$$

within the operating bandwidth. Therefore, the loading resistor noise current referred to inductive peaking network input is much reduced. Fig. 4. 26 shows the plots on type II series peaking with capacitor ration $k$=1, R=1Ω, and $C_1$+$C_2$=1F, that the output referred noise power of the loading resistor reduces to below 0.5 of the original in 59.3% band ((0.5037-0.1474)/0.6012) of -3-dB bandwidth, showing wideband noise suppression.

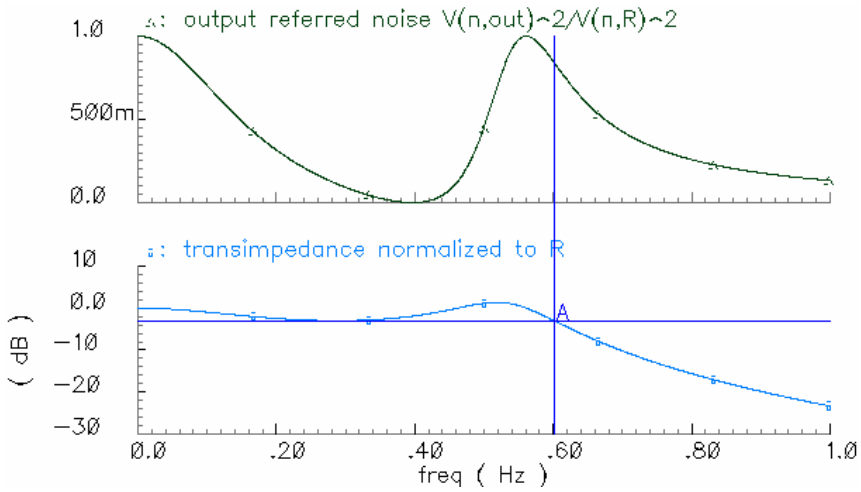



Fig. 4. 26 Frequency response of output referred noise from loading resistor, and transimpedance of type II series peaking

Similar noise suppression of loading resistor happens to T-coil, but not type I series peaking, in which case the noise current of the resistor appears at the input port of the peaking network, being bandwidth extended the same amount as the signal (noise transfer and signal transfer are the same). Therefore, in terms of noise, type II series peaking and T-coil are more favorable than type I peaking. As shown in Fig. 4. 27 for an ideal T-coil with R=1Ω and $C_2$=1F, in 47.8% of the whole -3-dB bandwidth, noise power from loading resistor reduces to below half of the original noise.

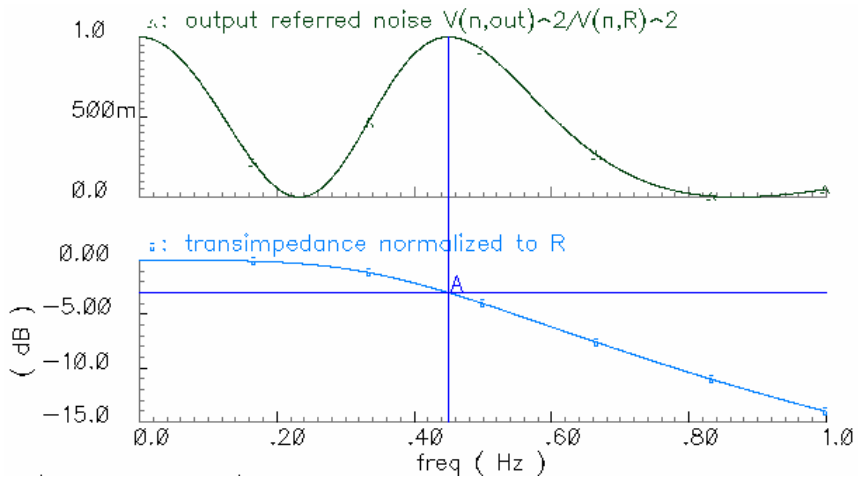



Fig. 4. 27 Frequency response of output referred noise from loading resistor, and transimpedance of T-coil

The simulated overall NF of the UWB LNA is shown in Fig. 4. 28. Due to larger gain, better RF choke inductor impedance, and noise filtering of loading resistor at high-frequency, the NF improves at high-frequency. TABLE 4. 5 summarizes noise contribution from each part at 5 GHz, TABLE 4. 6 summarizes the main design parameters of the LNA.

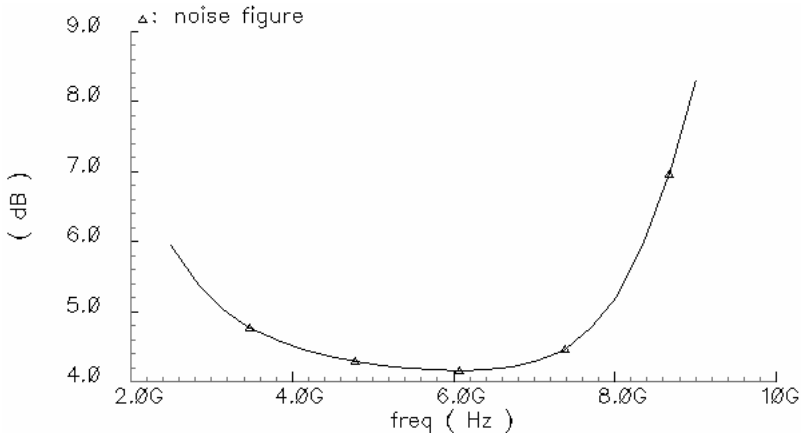



Fig. 4. 28 Simulated NF of the UWB LNA

TABLE 4. 5 NOISE CONTRIBUTION FROM EACH PART OF THE LNA

| | |
|---|---|
| Input source resistance | 37.6% |
| 1st stage input transistor | 20.8% |
| 2nd stage input transistor | 14.7% |
| 1st stage loading resistor | 11.9% |
| 3rd stage input transistor | 4.9% |
| 2nd stage loading resistor | 2.7% |
| Loss in RF choke inductor | 2.5% |
| Loss in T-coil | 1.9% |
| 3rd stage cascode transistor | 1.0% |
| 3rd stage loading resistor | 0.5% |
| Others | 1.5% |
| Total | 100% |

TABLE 4. 6 MAIN DESIGN PARAMETERS OF THE UWB LNA

| | |
|---|---|
| 1st stage Input transistor size | 30-μm/0.18-μm |
| 1st stage loading resistor | 150-Ω |
| 2nd stage transistor size | 70-μm/0.18-μm |
| 2nd stage loading resistor | 160-Ω |
| 3rd stage transistor size | 60-μm/0.18-μm |
| 3rd stage loading resistor | 190-Ω |
| 3rd stage cascode transistor | 60-μm/0.18-μm |
| Cross-coupling capacitor | 2 pF |
| RF choke inductor | 3.95 nH |
| Series peaking inductor | 4.22 nH |

*E. Consideration for the Common-Mode Rejection*

Due to the wideband operation, common-mode rejection is also important to avoid $2^{nd}$ order intermodulation products to corrupt the desired signal. Typical IIP2 achieved for the UWB receiver is around 20 dBm [4. 5] [4. 17], this is not a very stringent requirement. Whenever there is enough voltage headroom (in $2^{nd}$ stage), tail current source is added to help reject the common-mode tones; when voltage headroom is limited (in $3^{rd}$ stage), current source is removed from the differential pair to allow larger $V_{ds}$ for the pair to improve its $3^{rd}$ order linearity, while the $2^{nd}$ order linearity is still acceptable for the UWB application.

### 4.3.2 Measurement Results

The proposed UWB LNA is fabricated in TSMC 0.18-μm CMOS process ($V_{Tn}$ = 0.52 V, $V_{Tp}$ = - 0.54 V) with 6 metal layers. Fig. 4. 29 shows the photograph of the chip, which occupies the area of $1.21 \times 0.75$ $mm^2$.

As shown in Fig. 4. 30, the proposed LNA measures the S11 of better than -13 dB within the whole 9 frequency bands from 3.1 GHz to 8.0 GHz. Fig. 4. 31 shows the measured voltage gain of the LNA over the frequency range. It achieves a variable gain from 12 dB to 22.4 dB while consuming 15.5 mA from a 1.5-V supply.

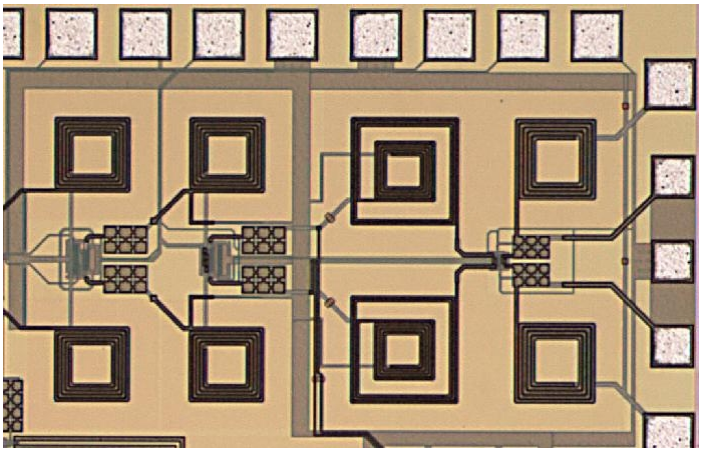


Fig. 4. 29 Micrograph of the UWB LNA

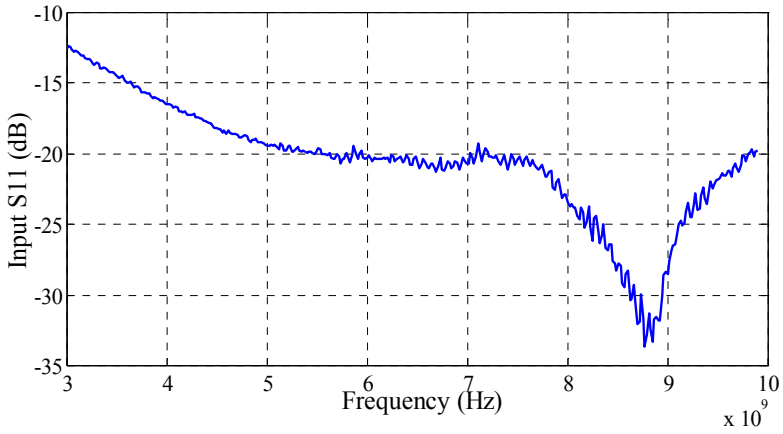



Fig. 4. 30 Measured S11 of the LNA

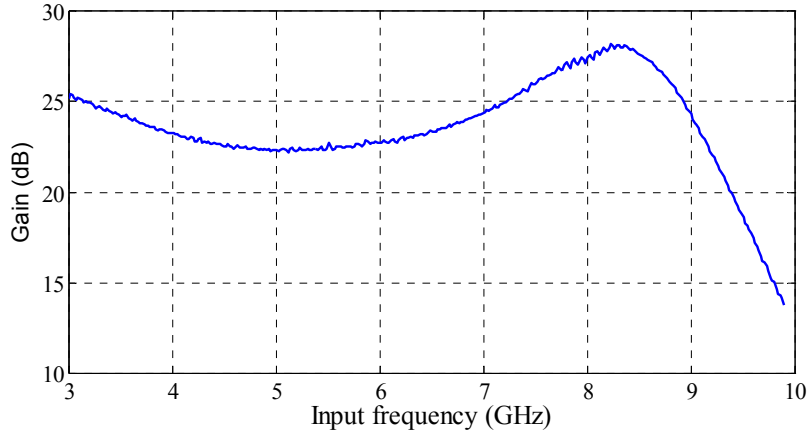



Fig. 4. 31 Measured voltage gain of the LNA

The LNA's IIP3 for the 5$^{th}$ frequency band with high LNA gain setting of 22 dB is measured to be -6 dBm. Fig. 4. 32 shows measured NF of the LNA.

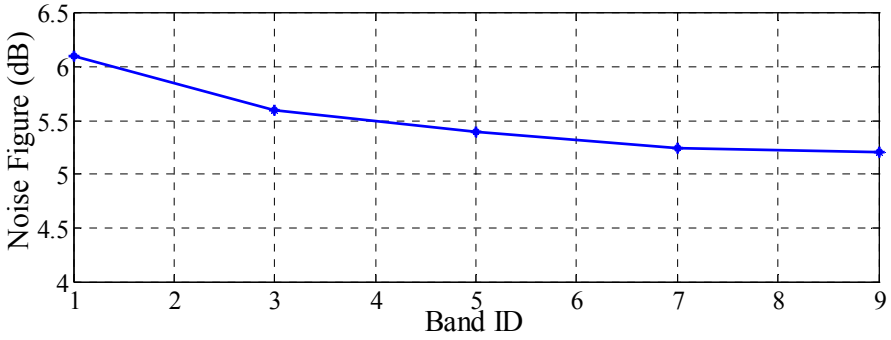



Fig. 4. 32 Measured NF of the LNA

TABLE 4. 7 compares the measured performance with other UWB LNAs.

TABLE 4. 7 COMPARISON WITH OTHER UWB LNAS

| | *[4. 5]* | *[4. 17]* | *This work* |
|---|---|---|---|
| *Input Freq.* | 2.3~9.2 GHz | 1.2~11.9 GHz | 3.1~8 GHz |
| *Noise Figure* | 4~7 dB | 4.5~.5.1 dB | 5~6.3 dB |
| *Gain* | 9.3 dB | 9.7 dB | >22 dB |
| *IIP3* | -6.7 dBm | -6.2 dBm | -6.0 dBm |
| *DC power* | 9 mW | 20 mW | 23 mW |
| *Process* | 0.18μm CMOS | | |
| *Topology* | LC ladder, single-ended | CG, single-ended | CG, differential |
| *Area* | 1.1 $mm^2$ | 0.59 $mm^2$ | 0.91 $mm^2$ |

The differential LNA consumes reasonable power consumption, achieves similar NF, with much higher gain than the others, which is very helpful in reducing the noise contribution from latter stages of the UWB receiver, and relaxes mixer noise consideration. Series peaking and T-coil are found to be very effective in bandwidth extension when parasitic capacitors are not lumped at one node, in addition, type II series peaking and T-coil achieves wideband noise rejection from the loading resistor. The RF choke inductors for the common-gate input stage degrade the NF of the LNA at low-frequency end, thus should be maximized.

## REFERENCES


[4. 1] C. Ling, R. Montemayor, A. Cicalini, K. Wang, L. Jansson, L. Muche, P. Trihka, and S. V. Kishore, "A low-power integrated tuner for cable telephony applications," *IEEE J. Solid-State Circuits*, Vol. 37, No. 12, pp. 1757-1767, Dec. 2002.

[4. 2] MultiBand OFDM Alliance SIG, "Multiband OFDM physical layer proposal for IEEE 802.15 task group 3a," Sep. 14, 2004.

[4. 3] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[4. 4] R. M. Fano, "Theoretical limitations on the broad band matching of arbitrary impedances," *Journal of the Franklin Institute*, Vol. 249, pp. 57-83, January 1950, and pp. 139-154, February 1950.

[4. 5] A. Bevilacqua, and A. M. Niknejad, "An Ultra-wideband CMOS LNA for 3.1 to 10.6 GHz wireless receivers," *IEEE International Solid-State Circuits Conference*, Feb. 2004.

[4. 6] W. Zhuo, X. Li, S. Shekhar, S. H. K. Embabi, J. Pineda de Gyvez, D. J. Allstot, and E. Sanchez-Sinencio, "A capacitor cross-coupled common-gate low-noise

amplifer," *IEEE Trans. Circuits and Systems II*, Vol. 52, No. 12, pp. 875-879, Dec. 2005.

[4. 7] F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Noise canceling in wideband CMOS LNAs," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Vol. 45, pp. 406-407, Feb. 2002.

[4. 8] Thomas T. Y. Wong, *Fundamentals of Distributed Amplification*, Artech House, 1993.

[4. 9] Dan Wang, "Design and integration of a single-chip low-power single-conversion CMOS cable TV tuner," Ph.D.dissertation, Hong Kong University of Science and Technology, 2005.

[4. 10] S. Andersson, C. Svensson, and O. Drugge, "Wideband LNA for a multistandard wireless receiver in 0.18 μm CMOS," *IEEE European Solid-State Circuits Conference*, 2003.

[4. 11] J. V. Sinderen, F. Seneschal, E. Stikvoort, F. Mounaim, M. Notten, H. Brekelmans, O. Crand, F. Singh, M. Bernard, V. Fillatre, and A. Tombeur, "A 48-860MHz digital cable tuner IC with integrated RF and IF selectivity," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 444-445, Feb. 2003.

[4. 12] L. Connell, N. Hollenbeck, M. Bushman, D. McCarthy, S. Bergstedt, R. Cieslak, and J. Caldwell, "A CMOS broadband tuner IC," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 400-476, Feb. 2002.

[4. 13] S. Stroh, "Ultra-wideabnd: multimedia unplugged," *IEEE Spectrum*, Vol. 40, pp. 23-27, Sep. 2003.

[4. 14] L. Selmi, D. B. Estreich, B. Ricco, "Small-signal MMIC amplifier with bridged T-coil matching networks," *IEEE J. Solid-State Circuits*, pp. 1093-1096, July 1992.

[4. 15] S. Galal, and B. Razavi, "Broadband ESD protection circuits in CMOS technology," *IEEE J. Solid-State Circuits*, pp. 2334-2340, Dec. 2003.

[4. 16] C. Lee, L. C. Cho, and S. I. Liu, "A 0.1-25.5-GHz differential cascaded-distributed amplifier in 0.18-μm CMOS technology," in Proc. *IEEE Asian Solid-State Circuits Conf. (ASSCC)*, pp. 129-132, Nov. 2005.

[4. 17] C.-F. Liao, and S.-I. Liu, "A broadband noise-canceling CMOS LNA for 3.1-10.6-GHz UWB receiver," *IEEE Custom Integrated Circuits Conference*, pp. 161-164, Sep. 2005.

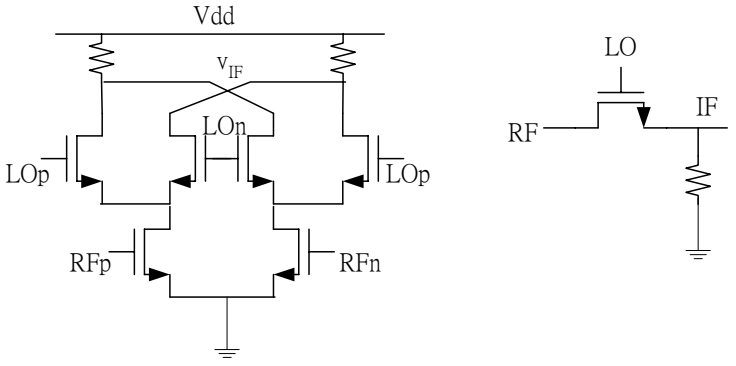
# Chapter 5 Mixer Design for UWB Transceiver

## 5. 1. Introduction

The RF signal amplified by the LNA is applied to the mixer. Mixer performs frequency translation to intermediate frequency (IF) (usually lower frequency), to facilitate easier signal amplification and processing.

Similar to LNA, good mixer has low NF, high linearity with moderate conversion gain. Due to the gain in front of the mixer, noise problem from mixer is relaxed, but its linearity is of great importance in the receiver chain.

Mixers perform frequency translation by multiplying two signals (and possibly their harmonics). Down-conversion mixers employed in the receive path have two distinctly different inputs, called the RF port and the local oscillator (LO) port, as shown in Fig. 5. 1. The RF port senses the signal to be down-converted and the LO port senses the periodic waveform generated by the local oscillator. There are two categories of mixers, active and passive mixers. Passive mixers don't provide any gain and typically achieve a higher linearity [5. 1]; by contrast, active mixers generally provide gain and reduce the noise contributed by subsequent stages and are widely used in RF systems.




(a) (b)

Fig. 5. 1 (a) Double-balanced Gilbert active mixer; (b) passive mixer

## 5. 2. UWB Transceiver System Architecture and Consideration

Many UWB transceivers focus on the direct-conversion transceiver architecture to eliminate the image problem [5. 2]-[5. 5]. However, such architecture would impose stringent requirement on wideband and high-frequency LO signals with accurate IQ outputs for acceptable sideband rejection (SBR). As the number of frequency bands increased up to 9 to cover from 3.1 GHz to 8.0 GHz, the IQ LO generation becomes even more challenging, in particular in low-cost CMOS technology. The feasibility to generate the desired IQ LO signals for UWB systems has been demonstrated. As illustrated in [5. 6], two single side-band (SSB) mixers can be used to generate the IQ outputs. However, the phase accuracy largely depends on the quadrature input signals of the SSB-mixers, which is not quite reliable especially at such high frequencies. The divided-by-2 solution can guarantee good phase accuracy [5. 2], but it would require VCO operation at double frequency and larger power.

To address these problems, a transceiver front-end with dual-conversion zero-IF architecture is designed for the first 9 frequency bands of MB-OFDM UWB systems from 3.1 GHz to 8.0 GHz.

As shown in Fig. 5. 2, in receiver (RX) chain, the RF signals are first amplified by a wideband LNA, then down-converted to a fixed IF1, and further converted to zero IF2 by an I/Q down-conversion mixer. In the transmitter (TX) path, the signal frequency conversion is the reverse of that in RX path. RF signal is finally transmitted out at buffer output. An on-chip synthesizer is included to provide all necessary LO signals. Fig. 5. 3 shows the proposed dual-conversion frequency scheme. The first variable LO1 signal down-converts RF signals to a fixed IF1 frequency at 2.904 GHz, and the second fixed-frequency IQ LO2 signals further down-converts the IF1 signals to zero IF2. The proposed scheme avoids a need for wide-band and high-frequency IQ

LO signals. In addition, it can help alleviate the problems with first LO leakage and with frequency pulling. However, image signals exist during the first-step conversion and need to be rejected. To achieve the required image suppression without any filtering, upper-sideband mixing with LO1 signals from 6.336 GHz to 10.56 GHz is proposed so that all the images are pushed completely outside of the desired frequency bands. With proper pre-filtering from the duplexer and the LNA, more than 30 dB image rejection can be readily achieved.

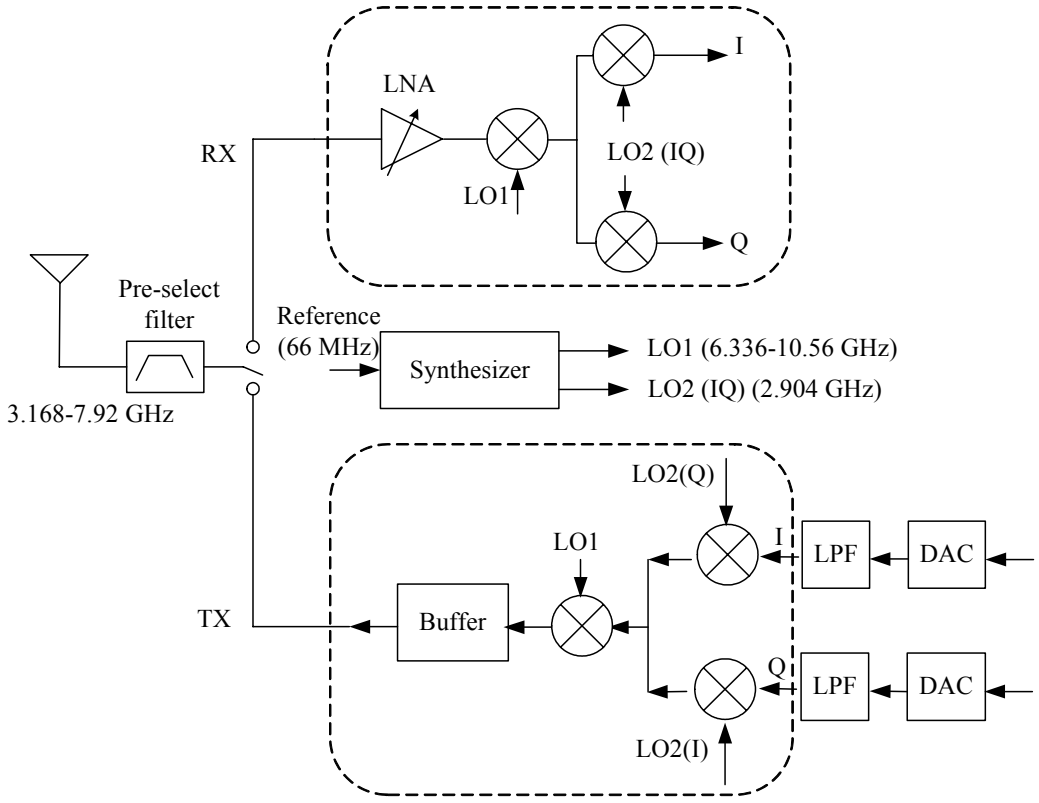



Fig. 5. 2 Proposed dual-conversion zero-IF2 front-end for UWB systems

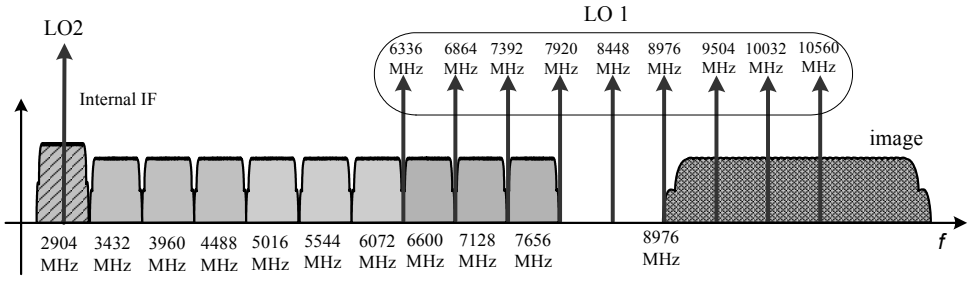



Fig. 5. 3 Proposed dual-conversion frequency plan

For the wideband and high-frequency operation, the interfaces between the synthesizer and the down-conversion mixers are quite critical, especially the interconnection between the high-frequency 1st LO output and the 1st stage mixer. However, due to the floorplan limitation, there exists long interconnection between mixer and synthesizer, which contributes large parasitic capacitors to synthesizer output, making synthesizer high-frequency and wideband operation difficult, as show in Fig. 5. 4 (a). Buffer can be inserted to drive the long interconnection, as shown in Fig. 5. 4 (b), but it consumes extra power, in addition, wideband loading occupies large chip area because inductors are usually needed in it. We can move the LO switches in mixer close to synthesizer, as shown in Fig. 5. 4 (c), however, long interconnections between RF to LO part and LO to IF part arise. To minimize the synthesizer's output loading and thus to improve the circuit performance in terms of frequency and power, a mixer with LO switching devices at the bottom of RF input pairs is proposed for the first stage down-conversion, as shown in Fig. 5. 4 (d). Potential problem with fast switching of the proposed mixer and the corresponding solution will be discussed in the following section.

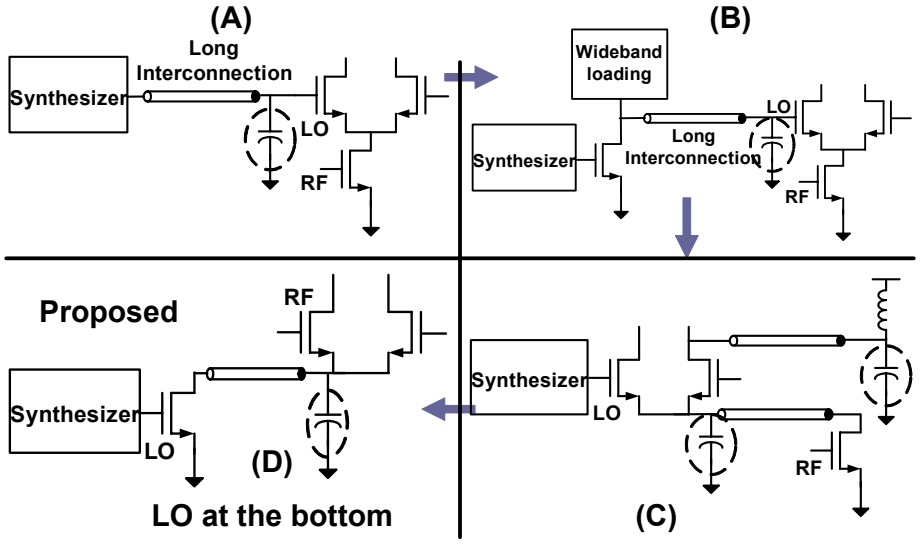



Fig. 5. 4 Mixer topology consideration

## 5. 3. UWB Mixer Design

### 5.3.1 Down-Conversion Mixer Design

*A. Proposed Mixer Topology*

As shown in 5. 2, two-stage mixers are required to perform dual conversion with the first IF1 at 2.9 GHz and the second IF2 at DC. Unlike the conventional mixers, the LO switch of the first mixer is proposed to be placed at the bottom below the RF input pair as shown in Fig. 5. 5.

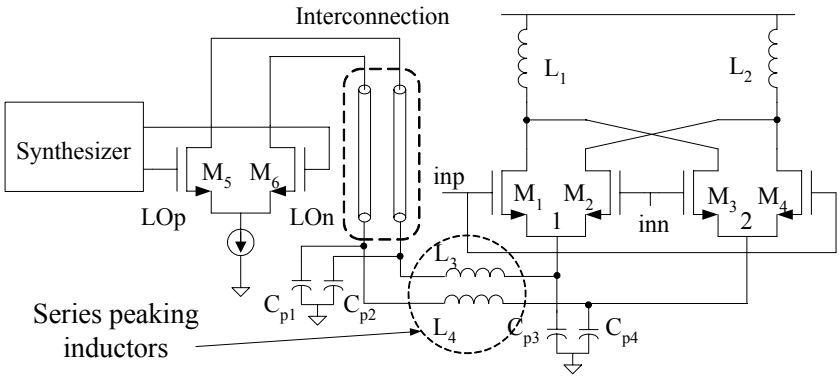



Fig. 5. 5 Schematic of the proposed 1st stage down-conversion mixer

With such modification, the LO transistors can be laid out very closely to the output of the synthesizer. As a result, both the interconnection between the synthesizer and the mixer's LO devices and the loading to the synthesizer can be significantly reduced, which is very critical for the synthesizer's wideband and high-frequency operation. A long interconnection is still required to connect the mixer's LO devices to its RF devices, which would also contribute large parasitic capacitance. For the proposed mixer topology, LO amplitude at source nodes of transconductors (nodes 1 and 2) needs to be large to fully turn on and off cross-coupled transconductors effectively one at a time.

However, at high frequencies, large parasitic capacitance at these nodes may prevent the LO voltages from rising quickly and from being sufficiently high to turn

off the RF transconductors, as shown in Fig. 5. 6. Consequently, both the RF transconductors may be simultaneously on in a large portion of one LO period, which would degrade the mixer's NF and gain.

To alleviate the problem, as discussed in the following, inductive series peaking technique is adopted to extend the bandwidth at the LO port, thus no buffer is needed to drive those parasitic capacitors.

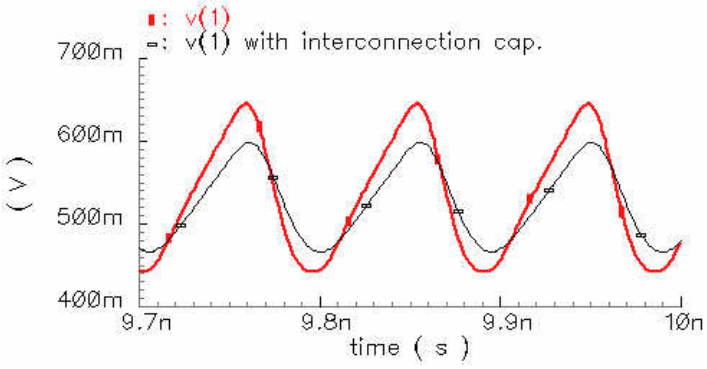



Fig. 5. 6 LO amplitude affected by the parasitic capacitors at transconductor source nodes

*B. Inductive Series Peaking in Mixer*

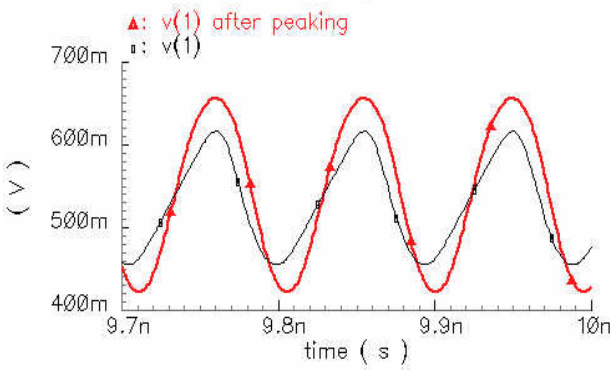



Fig. 5. 7 Simulated LO waveform at transconductor pair's source node without and with inductive series peaking

In the proposed mixer, series spiral inductors $L_3$ and $L_4$ are employed in place of long interconnections to form series peaking with parasitic capacitors to increase the bandwidth at the output node of the LO switches. Fig. 5. 7 shows the simulated LO

waveforms at node 1 without and with series peaking. Simulation shows that the series peaking inductors of 2.7 nH improve the gain from 0 dB to 2 dB and reduce the NF from 24 dB to 22 dB within the whole operation band.

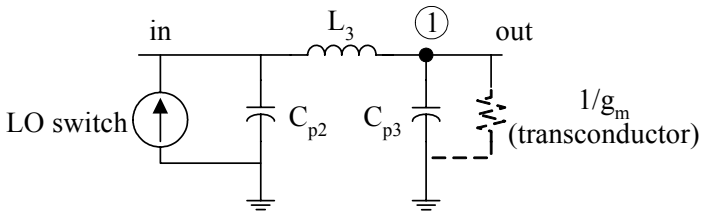



Fig. 5. 8 Model of series peaking

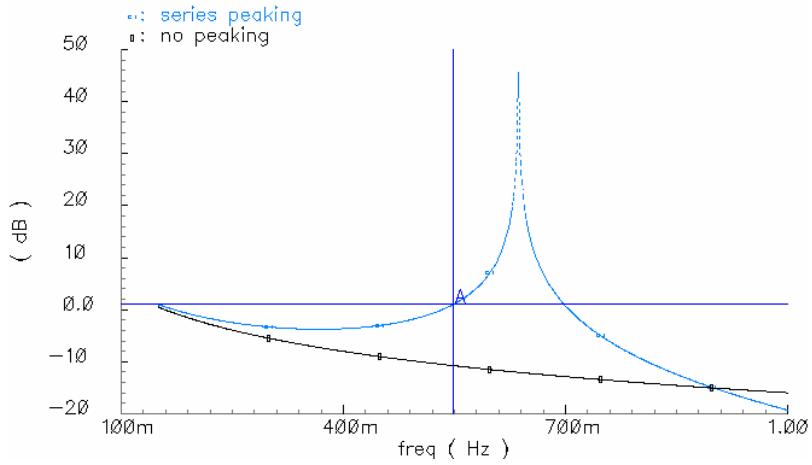



Fig. 5. 9 Transimpedance without and with series peaking

Nodes 1 and 2 need a high voltage to turn off the transconductor pair, when the transconductor exhibits high-impedance ($1/g_m \to \infty$). The series peaking network can then be modeled as shown in Fig. 5. 8, where the resistor $1/g_m$ is actually open. The transimpedance $v_{out}/i_{in}$ is plotted in Fig. 5. 9, in which $C_{p2}=C_{p3}$, $1/g_m=\infty$. There is a big peak in series peaking frequency response at $\omega = 1/\sqrt{L_3 C_{p3}}$, within this frequency, the response is much flatter than the case without peaking, showing the effectiveness of series peaking. It is also showed in Fig. 5. 9 that L3 should be chosen such that $L_3 < 1/\sqrt{\omega_0^2 C_{p3}}$, where $\omega_0$ is the maximum operating frequency. If we assume

$f_0$=10.56 GHz and $C_{p3}$=75 fF, L3 should be less than 3 nH. If L3 is too small, series peaking is not effective.

*C. Gain and Linearity Analysis*

With the LO voltage being applied periodically to the source nodes of the transconductors, the mixer's transconductor varies periodically. The output small-signal current for one of the transconductors ($M_1$, $M_2$, $M_3$, or $M_4$) can be expressed by [5. 7]:

$$i_o = g_1(t)\cdot v_{in} + g_2(t)\cdot v_{in}^2 + g_3(t)\cdot v_{in}^3 + \cdots \quad (5.\ 1)$$

where $v_{in}$ is the RF small signal applied at the gate of the transconductor, and $g_i$ represents the $i^{th}$-order transconductance coefficient of the transconductor. $g_i$ is a function of $V_{gs}$ and $V_{gs}$ now varies periodically with large applied LO signal, so $g_i$ also varies periodically with the applied LO voltage. Recall $g_i$-$V_{gs}$ plot in Fig. 3. 30, Fig. 5. 10 illustrates the periodically varying $g_i(t)$ for a 60-μm/0.18-μm nMOSFET.

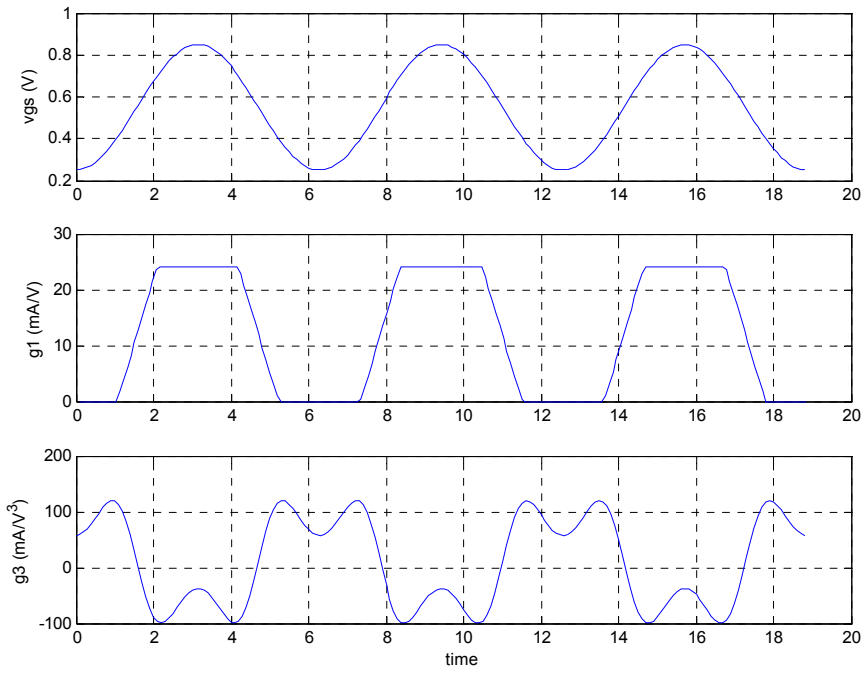



Fig. 5. 10 Nonlinearity coefficient $g_i$ as a function of time

Without loss of generality, the LO signal is assumed to be sinusoidal, $A_{LO}\cos(2\pi f_{LO}t)$, where $A_{LO}$ and $f_{LO}$ are LO amplitude at node 1 (or 2) and LO frequency. $g_1(t)$, $g_2(t)$, and $g_3(t)$ can thus be even functions of time and can be expanded in a series of sinusoids. In this case, (5. 1) provides:

$$i_o = g_{1,0}\cdot v_{in} + g_{2,0}\cdot v_{in}^2 + g_{3,0}\cdot v_{in}^3 + \sum_{k=1}^{\infty}[g_{1,k}\cdot v_{in} + g_{2,k}\cdot v_{in}^2 + g_{3,k}\cdot v_{in}^3 + \cdots]\cdot\cos(2\pi k f_{LO}t) \quad (5.\ 2)$$

where $g_{i,k}$ is the $k^{\text{th}}$ coefficient of the $g_i(t)$ in the series. As the mixer is used for down-converting RF signal by $f_{LO}$, the output current in the frequency band of interest can be expressed as

$$i_o = b_1\cdot v_{in} + b_2\cdot v_{in}^2 + b_3\cdot v_{in}^3 + \cdots \quad (5.\ 3)$$

where

$$b_i = \frac{g_{i,1}}{2} = \frac{1}{T_{LO}}\int_0^{T_{LO}} g_i(t)\cos(2\pi f_{LO}t)dt \quad (5.\ 4)$$

and $T_{LO}$ is the LO period.

If $g_1$ is approximated as a brick-wall shape:

$$g_1(V_{gs}) = \begin{cases} g_{1,\max}, when\ V_{gs} > V_{th} \\ 0, when\ V_{gs} \le V_{th} \end{cases} \quad (5.\ 5)$$

The $V_{gs}$ of the transconductor ($M_1$-$M_4$) is determined by the LO voltage:

$$V_{gs} = V_{th} - A_{LO}\cos(2\pi f_{LO}t) \quad (5.\ 6)$$

So

$$g_1(t) = \begin{cases} 0, 0 \le t \le T_{LO}/4\ and\ 3T_{LO}/4 \le t \le T_{LO} \\ g_{1,\max}, T_{LO}/4 < t < 3T_{LO}/4 \end{cases} \quad (5.\ 7)$$

and

$$b_1 = \frac{1}{T_{LO}}\int_0^{T_{LO}} g_1(t)\cos(2\pi f_{LO}t)dt = -\frac{1}{\pi}g_{1,\max} \quad (5.\ 8)$$

Because two transconductance devices are connected to the same RF input, the overall transconductance becomes $\frac{2}{\pi} g_{1,\max}$. In Fig. 5. 5, the transistor sizes are designed such that $W_{M1} = W_{M2} = W_{M3} = W_{M4} = 0.5*W_{M5} = 0.5*W_{M6}$. As comparison, in the conventional mixer arrangement with the RF port at the bottom, the effective transconductance would be $\frac{2}{\pi} g_5$. Since the bias current for the proposed arrangement will only pass through half of the transconductors due to LO switching, the current density is 2 times that of the conventional mixer, and $g_{1,\max} = \frac{\sqrt{2}}{2} g_5$. As a consequence, the proposed mixer's gain is 3 dB smaller.

If the RF input signal $v_{in}$ consists of two tones of equal magnitude $V_{in}$ at two closely spaced frequencies $f_1$ and $f_2$:

$$v_{in} = V_{in}\cos(2\pi f_1 t) + V_{in}\cos(2\pi f_2 t) \tag{5. 9}$$

Then the generated third-order intermodulation becomes [5. 1]:

$$IM_3 = \frac{3}{4}\frac{b_3}{b_1} V_{in}^2 \tag{5. 10}$$

where

$$b_3 = \frac{1}{T_{LO}} \int_0^{T_{LO}} g_3(t)\cos(2\pi f_{LO} t)dt \tag{5. 11}$$

The waveform $g_3(t)$ is shown in Fig. 5. 10, which for simplicity can be approximate as:

$$g_3(t) = a \cdot \sin(\frac{\pi \cdot (V_{gs}(t) - V_0 + V_1)}{V_1}) = a \cdot \sin(\frac{\pi \cdot (V_{th} - A_{LO}\cos(2\pi f_{LO} t) - V_0 + V_1)}{V_1}) \tag{5. 12}$$

where $a$ is a coefficient determining the maximum $g_3$ over different $V_{gs}$, $V_0 - V_1$, $V_0$ and $V_0 + V_1$ are the $V_{gs}$ voltage that makes $g_3$ be 0. Putting (5. 12) into (5. 11) and

simplifying calculation by assuming $V_{th} = V_0$, $A_{LO} = V_1$ =160 mV, it can be calculated that:

$$b_3 = \frac{1}{T_{LO}} \int_0^{T_{LO}} a \cdot \sin\left(\frac{\pi \cdot (Vth - A_{LO}\cos(2\pi f_{LO} t) - V_0 + V_1)}{V_1}\right) \cos(2\pi f_{LO} t) dt = 0.285a = a - 10.9dB \quad (5.\ 13)$$

The time-averaging effect makes effective $g_3$ smaller. Even though $a$ is the peak value for $g_3$, for power-efficient mixer design, the mixer should not be biased at the point that has 10 dB better $g_3$ than its peak. The result in (5. 13) does depend on the assumption in (5. 12), but improvement of around 10 dB can still be estimated. Consequently, the transconductor's linearity is improved over a wide frequency band. Unlike in the conventional mixers for which the linearity is significantly degraded by the switching pairs [5. 7], in the proposed mixer, the linearity can be even better than that of the transconductor as numerically calculated in (5. 13). So the proposed mixer can achieve better linearity and wideband operation with the same power consumption as the conventional mixer. Fig. 5. 11 shows the simulated output spectrum of the mixer with only the LO and RF inputs swapped. In the simulation, the mixer is driven by a 500-mV peak-to-peak single-end LO signal at 8.45 GHz, and -48 dBV RF voltage at 5.54 and 5.55 GHz. The proposed mixer achieves 19 dB improvement of IM3 suppression with 3.8 dB gain reduction.

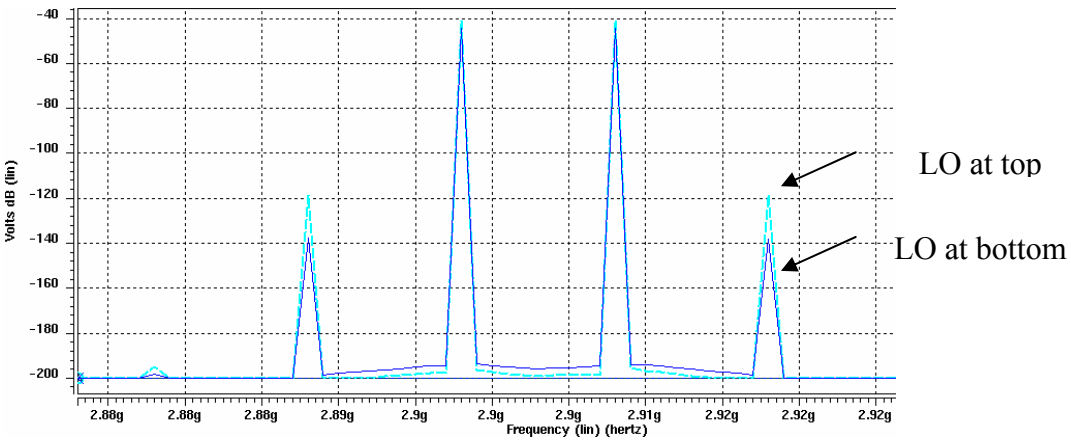



Fig. 5. 11 Simulated output spectrum of the mixer with RF and LO input swapped

The second-stage IQ mixers employ the conventional Gilbert cell for a fixed and lower RF frequency at 2.9 GHz. Due to its narrow-band requirement, inductive degeneration of 2 nH is adopted to improve its linearity.

### 5.3.2 Up-Conversion Mixer Design

In the MBOA UWB transmitter (TX) path, the average output power is limited to only -10 dBm despite the large peak-to-average ratio of OFDM signals, and the TX can operate with only a 4-dB backoff from the -1-dB compression point with little performance degradation [5. 3]. As a result, the required output 1-dB compression point of the TX is only around –6 dBm, which can be achieved by an up-conversion mixer without a need of a wideband power amplifier.

Because of wideband operation, the mixer needs to have a loading with low quality factor $Q$ and thus can only have limited gain. In addition, in order to drive an antenna directly without a power amplifier, the mixer also needs to have the output impedance matched to 50-Ω for maximum power transfer to the antenna. Finally, the bias current and the transistor sizes in the mixer cannot be too large due to the limited driving capability of the synthesizer.

Existing MBOA UWB transmitters [5. 3] [5. 8] mainly focus on lower 3 bands. With operation bands increasing, as discussed before, not only mixer design becomes more challenging, LO generation becomes more and more difficult. The proposed two-stage up-conversion IQ mixers are shown in the lower part of Fig. 5. 2. For the first IQ mixer stage, a fixed LO2 signal of 2.904 GHz is used to up-convert baseband IQ signals to a fixed IF1. For the second mixer stage, variable LO1 signals from 6.336 GHz to 10.560 GHz further up-converts the IF1 signals to the desired RF frequencies. This high-sideband LO2 signal pushes the unwanted upper sideband (8.976-13.728 GHz) generated in the second step conversion to outside of the desired frequency

bands. A buffer is added at the output to obtain output impedance matching to 50-Ω. The complete schematic of the whole mixer is shown in Fig. 5. 12.

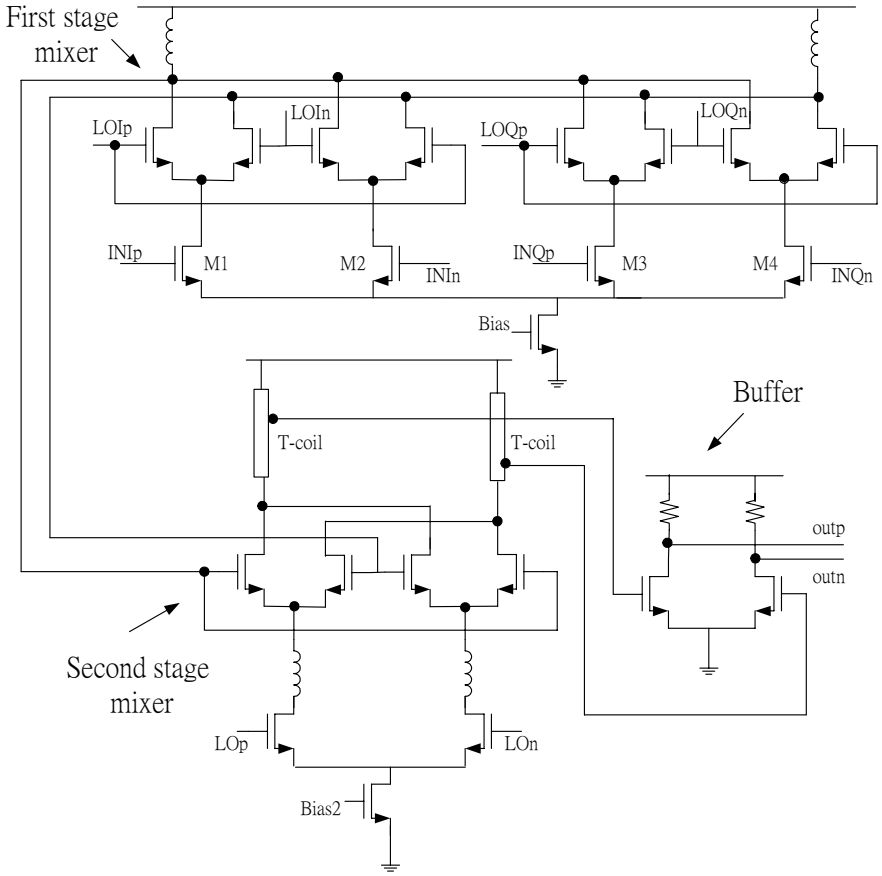



Fig. 5. 12 Schematic of the proposed up-conversion mixer

**Design Consideration**

In first stage mixer, the matching among input transistors M1-M4 is important, which determines the 1st LO leakage to the output. Because M1-M4 work at low-frequency so are not sensitive to the parasitic capacitors, their channel length is chosen to be large than minimum, L=0.25-μm, to improve matching characteristic among them.

Since the required output power of UWB transmitters is as low as -10 dBm, which corresponds to a voltage $V_{o,rms}$ of 71 mV for a 50-Ω impedance, the distortion from transistor drain node is not significant because of small output swing. Therefore,

having enough gain in each stage is critical to achieve the targeted output -1-dB compression point ($OP_{-1dB}$). Too small gain will require a large input for the desired output power, which leads to the possibility of reaching compression at the input of the transistors.

Adding a buffer is an efficient way of using power and greatly reduces the burden of the second mixer. The second mixer and the buffer need to be ultra wideband, the gain is limited and should be designed carefully. Moreover, the second mixer performs the signal mixing, its compression performance is worse than that of the buffer. From simulation, it is found that total gain around 0 dB for the second mixer and buffer is acceptable for the targeted $OP_{-1dB}$. In this design, the gain of the second mixer and buffer is -3 dB and 1 dB, respectively.

**Combination of Down / Up Conversion Mixers**

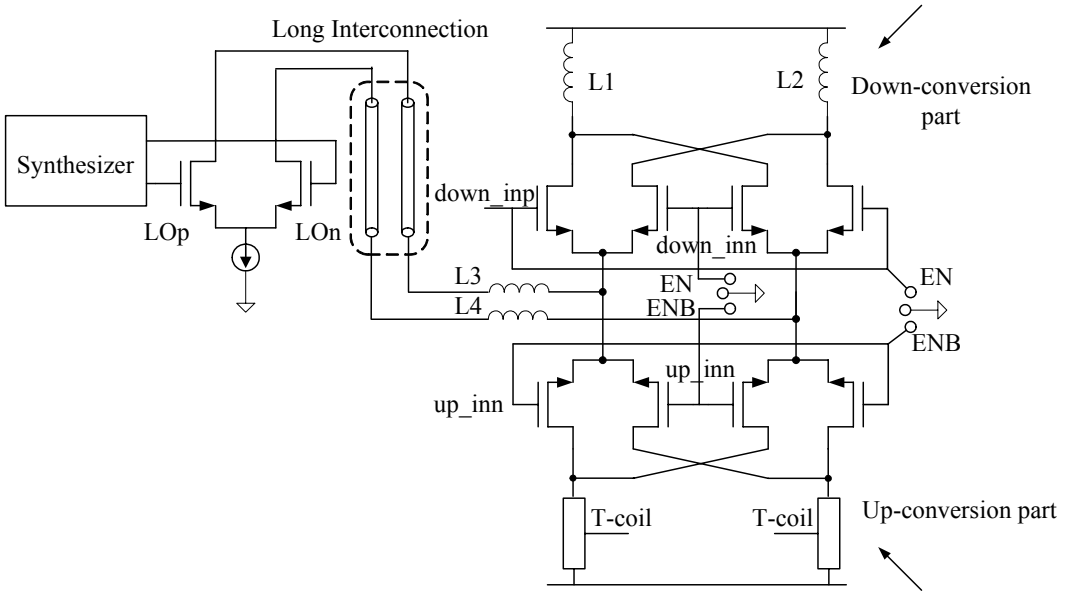



Fig. 5. 13 High-frequency LO switching pair shared by down- and up-conversion mixers

To minimize capacitive loading to synthesizer, the high-frequency LO switching pair is put at the bottom of the transconductor, the same as that in down-mixer, and

shared between down- and up- conversion mixers, as shown in Fig. 5. 13. By doing so, the LO switching pair can be laid out very close to the synthesizer, parasitic capacitor from interconnection is greatly reduced, and that from active device is reduced by half, which is critical for the wideband and high-frequency LO generation. Switches controlled by EN and ENB enable down- or up-conversion mixer one at a time.

## 5. 4. Measurement Results

The proposed dual-conversion zero-IF UWB front-end with a fully-integrated frequency synthesizer is designed and fabricated in TSMC 0.18-μm CMOS process ($V_{Tn}$ = 0.52 V, $V_{Tp}$ = - 0.54 V) with 6 metal layers. Fig. 5. 14 shows the photograph of the chip, which occupies the area of 2.9×2.5 $mm^2$.

The synthesizer can generate all the LO signals for the 9 UWB frequency bands with band switching time of less than 1 ns. Single-ended output amplitude of larger than 400 mVp-p is obtained over the 9 frequency bands [5. 9].

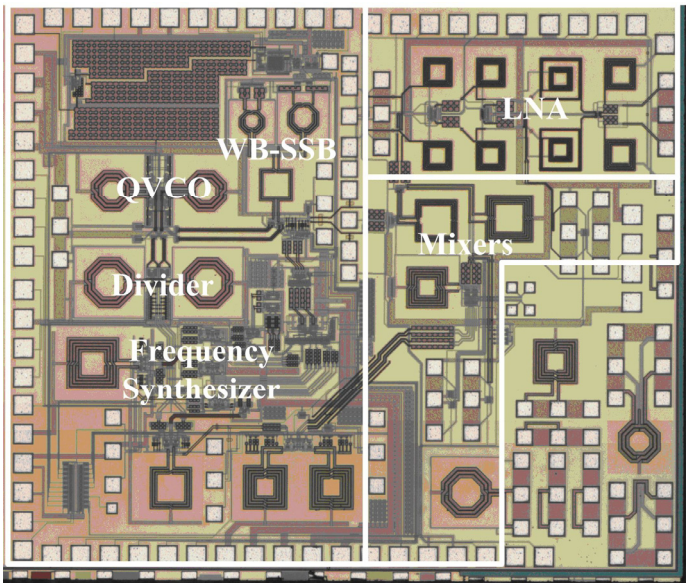



Fig. 5. 14 Photograph of the proposed UWB front-end

Measured performance of down-conversion mixer is summarized in TABLE 5. 1. It consumes 7 mA from 1.5 V supply and delivers 6.7 dB gain.

TABLE 5. 1 MEASURED DOWN-MIXER PERFORMANCE

| | 1st mixer | 2nd mixer |
|---|---|---|
| Gain / dB | 1.1 | 5.6 |
| IIP3 / dBm | 11 | 6 |
| NF / dB | 21.6 | |
| Power / mW | 4.5 | 6 |

The whole receiver front-end (RFE) measures gain around 30 dB across the whole operation band, as shown in Fig. 5. 15.

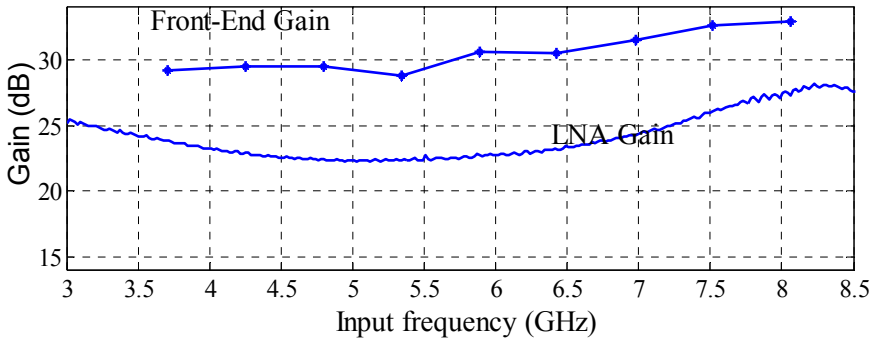



Fig. 5. 15 Measured voltage gains of the LNA and of the RFE

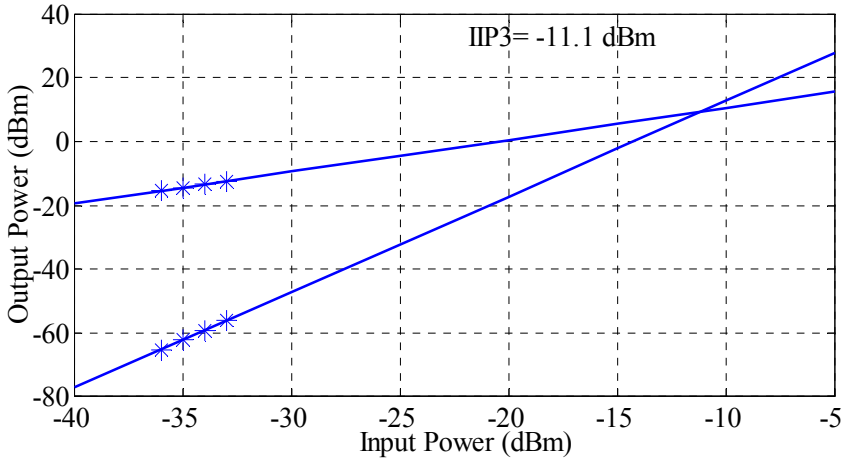



Fig. 5. 16 Measured IIP3 at 5th-band of the RFE with low LNA gain setting

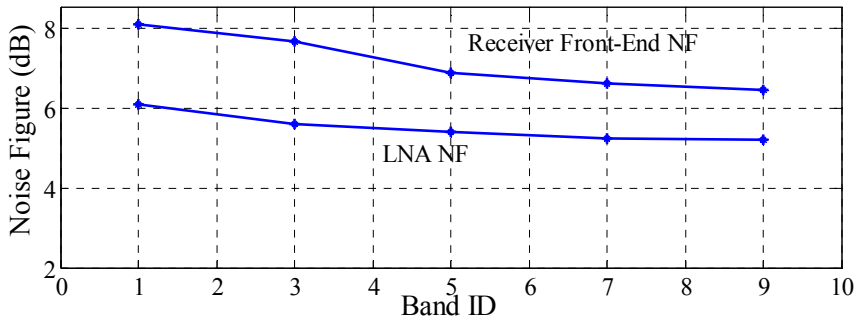



Fig. 5. 17 Measured NF of the proposed LNA and UWB RFE

The RFE's IIP3 for the 5$^{th}$ frequency band with a low LNA gain setting of 14 dB is measured to be -11.1 dBm as shown in Fig. 5. 16. The measured IIP2 is 21.6 dBm with the same low LNA gain setting.

Fig. 5. 17 shows the RFE and LNA's measured NF for the 9 different bands with a maximum NF of 8.1 dB. Since the communication resides in each band for 1/3 of the times within each band group (BG), the input-referred NF of the UWB RFE are averaged within each BG. The minimum averaged NF is measured to be around 6.5 dB for BG 3 while the maximum averaged NF is 7.9 dB for BG 1. The NF is larger at lower frequency, it is due to relatively small impedance of the RF choke inductors at low frequency bands in first LNA stage, which introduces signal loss as discussed before, in addition, at the output of the LNA, the filtering is not sharp enough to filter out the noise at image frequency for low frequency bands before down-conversion, so that noise adds to the input of mixer.

Operated under a 1.5-V supply, the synthesizer consumes 59 mA while the LNA and mixers dissipate a total current of 22.5 mA. TABLE 5. 2 summarizes the measured performance of the proposed UWB RFE. TABLE 5. 3 compares the measured performance with the other state of the art UWB RFEs.

TABLE 5. 2 PERFORMANCE SUMMARY OF THE PROPOSED UWB RFE

| | BG 1 | BG 2 | BG 3 |
|---|---|---|---|
| Voltage Gain (dB) | 29.4 | 29.97 | 32.3 |
| NF (dB) | 7.87 | 6.86 | 6.51 |
| S11 | < -13 | < -18 | <-20 |
| In-Band IIP3 (dBm) | -9.95 | -11.1 | -12.6 |
| Input P-1dB (dBm) | -21.8 | -23.8 | -24.5 |
| PN @ 10MHz (dBc/Hz) | < -126.3 | < -124.5 | < -123.1 |
| Sideband Rejection (dBc) | < -32 | < -22 | < -27 |
| Supply Voltage | 1.5 V | | |
| Current Consumption (mA) | 59 mA (Synthesizer)<br>22.5 mA (LNA+Mixers) | | |
| Process | TSMC 0.18-μm CMOS | | |
| Chip Area | 2.9×2.5 $mm^2$ | | |

TABLE 5. 3 COMPARISON WITH OTHER UWB RFE

| | [5. 3] | [5. 5] | This work |
|---|---|---|---|
| Frequency (GHz) | 3.1~4.8 | 3.1~9.5 | 3.1~8 |
| Voltage Gain (dB) | 21~24 | 27 (LNA) | 29~32 |
| NF (dB) | 5.4~8.4 | 6.3~7.8 | 6.5~8.1 |
| S11 (dB) | < -11 | NA | < -13 |
| In-Band IIP3 (dBm) | NA | >-17 | > -12.6 |
| Input $P_{-1dB}$ (dBm) | -9.5~-12<br>@ 7dB Gain | NA | > -24.5<br>@ 22dB Gain |
| PN @ 1MHz (dBc/Hz) | < -104 | -108~-98 | < -123.1<br>@ 10MHz |
| DC Power: LNA+Mixer | 7.5@1.5V | NA | 33.8@1.5V |
| Synthesizer (mW) | 67.5@1.5V | 47@1.1V | 88.5@1.5V |
| Process (CMOS) | 0.13-μm | 0.09-μm | 0.18-μm |
| Architecture | Direct-Conversion | Direct-Conversion | Dual-Conversion |

The up-mixer consumes 20 mA (4 mA, 4 mA, 12 mA for 1st, 2nd and buffer, respectively) from 1.8 V supply. The measured $OP_{-1dB}$ of up-mixer across the UWB bands is shown in Fig. 5. 18. Fig. 5. 19 shows the up-mixer output spectrum at the highest frequency band. Sideband rejection is better than 40 dBc for all the frequency bands, and LO leakage power is less than -58 dBm.

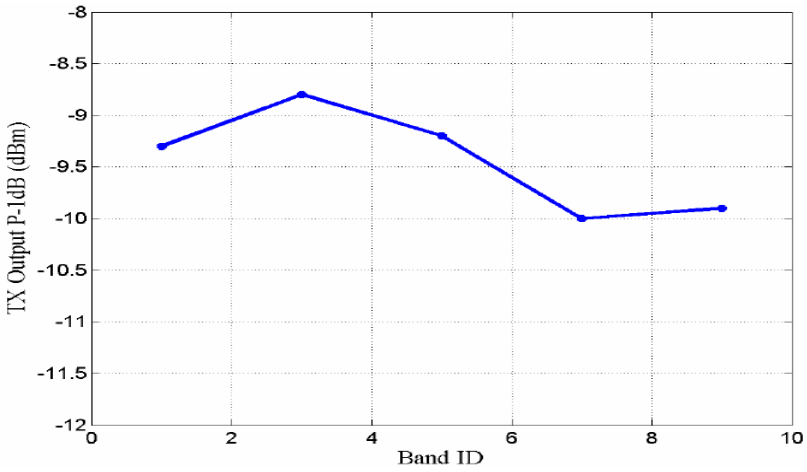



Fig. 5. 18 Measured $OP_{-1dB}$ of up-mixer across the UWB bands

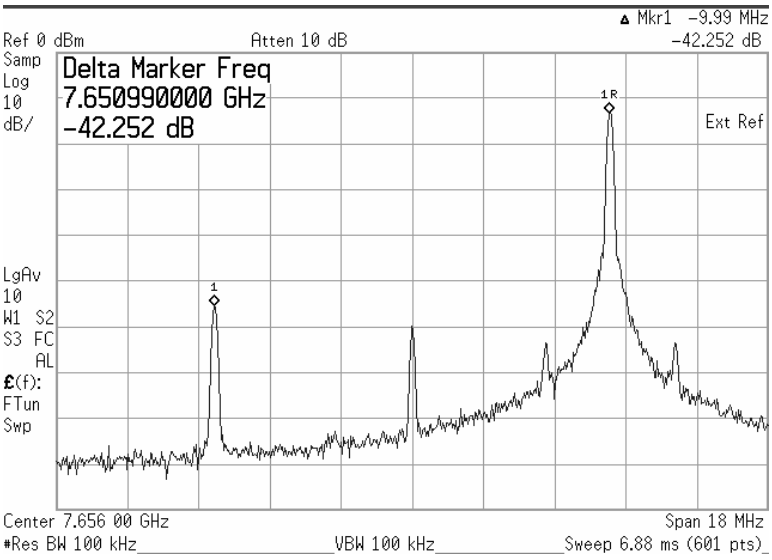



Fig. 5. 19 Measured output spectrum of up-mixer

## REFERENCES

[5. 1] B. Razavi, *RF Microelectronics*, Prentice Hall, 1998.

[5. 2] A. Ismail and A. Abidi, "A 3.1 to 8.2 GHz direct conversion receiver for MB-OFDM UWB communications," *IEEE International Solid-State Circuits Conference*, pp. 208-209, Feb. 2005.

[5. 3] B. Razavi, T. Aytur, C. Lam, F. R. Yang, K. Y. Li, R. H. Yan, H. C. Kang, C. C. Hsu, and C. C. Lee, "A UWB CMOS transceiver," *IEEE J. Solid-State Circuits*, Vol. 40, No. 12, pp. 2555-2562, Dec. 2005.

[5. 4] G. Cusmai, M. Brandolini, P. Rossi, and F. Svelto, "A 0.18-µm CMOS selective receiver front-end for UWB application," *IEEE J. Solid-State Circuits*, Vol. 41, No. 8, pp. 1764-1771, Aug. 2006.

[5. 5] A. Tanaka, H. Okada, H. Kodama, H. Ishikawa, “A 1.1V 3.1-to-9.5GHz MB-OFDM UWB transceiver in 90nm CMOS,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 120-121, Feb. 2006.

[5. 6] D. Leenaerts, “A SiGe BiCMOS 1ns fast hopping frequency synthesizer for UWB radio,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 202-203, Feb. 2005

[5. 7] M. T. Terrovitis, and R. G. Meyer, “Intermodulation distortion in current-commutating CMOS mixers,” *IEEE J. Solid-State Circuits*, pp. 1461-1473, Oct. 2000.

[5. 8] S. Aggarwal, D. Leenaerts, R. van de Beek, G. van der Weide, K. S. Harish, J. Bergervoet, A. Landesman, Y. Zhang, C. Razzell, H. Waite, and R. Roovers, “A low power implementation for the transmit path of a UWB transceiver,” *IEEE Custom Integrated Circuits Conference*, pp. 149-152, Sep. 2005.

[5. 9] H. Zheng, and H. Luong, “A 1.5 V 3.1 GHz-8 GHz CMOS synthesizer for 9-band MB-OFDM UWB transceivers,” *IEEE J. Solid-State Circuits*, Vol. 42, No. 6, pp. 1250-1260, Jun. 2007.

# Chapter 6 Receiver Front-End Design for SDR

## 6. 1. Introduction

Wireless bands and services are now proliferating across the world. Every six months it seems a new use for wireless appears, often leading to a new standard, such as DECT (Digital Enhanced Cordless Telecommunications), Bluetooth, Wi-Fi, UWB, and recently Bluetooth 2.0. Manufacturers of mobile handsets have a hard time keeping up, because the end user wants to access an increasing number of services from a single handset, and have it adapt to global roaming. The concept of software-defined radio (SDR) [6. 1] allows the utilization of a single flexible receiver hardware in multiple standards. Achieving this goal requires a receiver that can work across standards instead of multiple receivers customized to work in narrow frequency bands.

LNA and mixer are still indispensable blocks for SDR receiver. After designing a few LNAs for different standards, the ultimate goal is to design an LNA covering every standard. In this chapter, a wideband LNA is proposed to work in the frequency band of 0.9-10.6 GHz for SDR application. The challenges for mixer design and requirements for the whole receiver front-end will be discussed.

## 6. 2. SDR LNA Design

### 6.2.1 Existing Solutions

Wideband LNA is one candidate for SDR receiver [6. 2] [6. 3], reconfigurable narrowband LNA is another candidate [6. 4] [6. 5] [6. 6]. Wideband LNA delivers relative constant gain across wide bandwidth, and avoids the need of frequency tuning which may lead to performance degradation in gain and NF, on the other hand, reconfigurable narrowband LNA provides filtering to the received signal, which helps with the linearity of the whole receiver by improving the selectivity, and relaxes IIP2,

IIP3 requirements of the mixer. However, the tuning range is usually limited [6. 4], or the tuning degrades circuits performance such as gain, NF, and power consumption [6. 4] [6. 5] [6. 6].

In [6. 4], frequency tuning is realized by loading inductor switching. However, even the tuning range is as small as 300 MHz (from 2.1 GHz to 2.4 GHz), and the switches are in good process of 0.13-μm CMOS, the gain reduces 6 dB due to the degradation of tank impedance by switching.

In [6. 5], the frequency tuning range is large (from 0.75 to 3 GHz) by tuning delay cells in recursive positive feedback loop, but the NF is as large as 4.8 dB and IIP3 is low with large power consumption of 42 mW.

In [6. 6], the input impedance is tuned by changing the Miller capacitor of input transistor through bias tuning. However, the tuning is quite sensitive, with only 19% bias current change, frequency response varies from 2.4 GHz to 5.2 GHz. Also, NF and IIP3 depend on the input stage bias current, too, so this tuning scheme has to make tradeoff among several performances.

In the proposed LNA, wideband topology is adopted.

### 6.2.2 Proposed SDR LNA

The proposed wideband LNA is shown in Fig. 6. 1. It adopts two-stage topology to deliver 20 dB gain with large gain-bandwidth-product (GBW). M1 and M2 form the capacitive cross-coupling common-gate stage [6. 7] for input impedance matching, reduced power consumption, and reduced NF compared to simple common-gate stage. Cross-coupled biasing transistor M3 and M4 are adopted to further reduce the NF by partial noise cancellation in M1-M4 [6. 8]. Wideband and high linearity techniques discussed earlier are also implemented to improve the circuit performance. AC coupling between 1$^{st}$ and 2$^{nd}$ stage, and between 1$^{st}$ stage and squaring circuit are

added, for low voltage operation and flexibility of circuit tuning, but not shown in Fig. 6. 1 for simplicity.

*A. Input Impedance Matching*

If we ignore the input parasitic capacitance $C_p$, the input impedance can be calculated to be

$$Z_{in} = \frac{1}{2g_{m1} - g_{m3}} \tag{6. 1}$$

In real case, parasitic capacitors $C_p$ from active devices and wiring, and $C_{pad}$ from probing pad significantly affect $Z_{in}$, the maximum parasitic capacitance tolerable is 200 fF for a targeted -10 dB S11 at 10.6 GHz. However, due to cross-coupling among several transistors, $C_p$ is large. Inductors $L_1$ and $L_2$ are inserted to alleviate the S11 degradation at high-frequency.

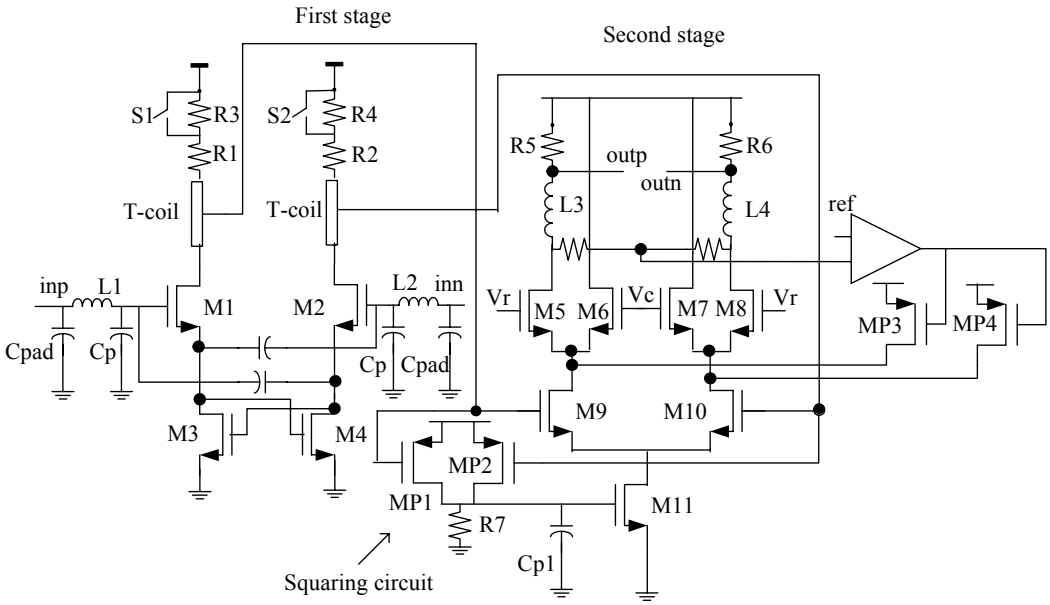



Fig. 6. 1 Schematic of the wideband LNA

Taking into consideration $L_1$ and $C_p$ ($C_{pad}$ (~45 fF) is much smaller than $C_p$ (~200 fF)), as shown in Fig. 6. 2, the input impedance can be written as:

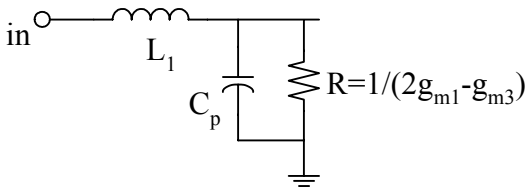



Fig. 6. 2 Input network including reactive components

$$Z_{in} = \frac{R}{1 + j\omega C_p R} + j\omega L_1 = \frac{R + j\omega L_1 - \omega^2 L_1 C_p R}{1 + j\omega C_p R} \tag{6. 2}$$

Reflection coefficient Γ is calculated as:

$$\begin{aligned} \Gamma^2 &= \left|\frac{Z_{in} - R}{Z_{in} + R}\right|^2 = \left|\frac{j\omega L_1 - \omega^2 L_1 C_p R - j\omega C_p R^2}{2R + j\omega L_1 - \omega^2 L_1 C_p R + j\omega C_p R^2}\right|^2 \\ &= \frac{(\omega^2 L_1 C_p R)^2 + (\omega L_1 - \omega C_p R^2)^2}{R^2(2 - \omega^2 L_1 C_p)^2 + (\omega L_1 + \omega C_p R^2)^2} \end{aligned} \tag{6. 3}$$

The optimum $L_1$ for minimizing Γ needs to be found by solving a 3-order equation based on (6. 3), which is quite complicated. The benefit of $L_1$ can be readily seen by an example. Assume that $C_p$=250 fF, ω=10.56 GHz*2π, let $L_1$=$C_p$*$R^2$=625 pH, this is already an improved solution compared to $L_1$=0, with Γ improved by 1.4 dB. The actual optimum $L_1$ is less than 625 pH. In this design, $L_1$ and $L_2$ are chosen to be 400 pH, they are implemented with transmission line and are area-efficient.

*B. Noise Figure*

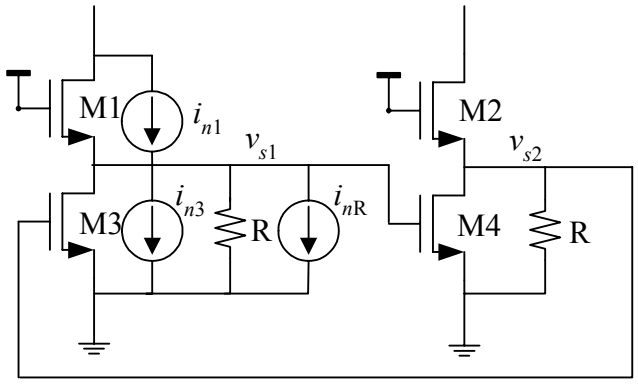



Fig. 6. 3 Simplified schematic for noise calculation

The noise canceling technique usually has the drawback of canceling noise from one certain device, but adding extra noise from the cancellation path, which makes the technique less attractive. Recently proposed noise cancellation technique in [6. 8] takes care of the noise from the noise cancellation path as well and optimizes noise of the circuit as a whole. In this design, the common-gate input stage in [6. 8] is changed to capacitive cross-coupling common-gate stage, and the corresponding optimum noise conditions are derived.

We consider the noise from devices M1-M4, as shown in Fig. 6. 3, where M1 and M2 are input devices, M3 and M4 are biasing transistors, which also form the noise canceling path. For noise current of M1, $i_{n1}$, the following steps calculate its noise transfer function to the output.

$$\begin{cases} i_{n1} + (0 - v_{s1}) g_{m1} = \dfrac{v_{s1}}{R} + v_{s2} g_{m3} \\ -v_{s2} g_{m1} = \dfrac{v_{s2}}{R} + v_{s1} g_{m3} \end{cases} \tag{6. 4}$$

where conditions $g_{m1}=g_{m2}$, $g_{m3}=g_{m4}$ are used. We get

$$\begin{cases} v_{s1} = -\dfrac{i_{n1}(1/R + g_{m1})}{g_{m3}^2 - (1/R + g_{m1})^2} \\ v_{s2} = \dfrac{i_{n1} g_{m3}}{g_{m3}^2 - (1/R + g_{m1})^2} \end{cases} \tag{6. 5}$$

Therefore, the output noise currents at the positive and negative output are

$$\begin{cases} i_{n1,op} = \dfrac{g_{m3}^2 - (1/R + g_{m1})^2 + g_{m1}(1/R + g_{m1})}{g_{m3}^2 - (1/R + g_{m1})^2} i_{n1} \\ i_{n1,on} = \dfrac{-g_{m1} g_{m3}}{g_{m3}^2 - (1/R + g_{m1})^2} i_{n1} \end{cases} \tag{6. 6}$$

The differential output noise is given by

$$i_{n1,diffout} = \frac{g_{m3}^2 - 1/R^2 - g_{m1}/R + g_{m1} g_{m3}}{g_{m3}^2 - (1/R + g_{m1})^2} i_{n1} \tag{6. 7}$$

For the noise current from M3, $i_{n3}$, and noise current from source impedance $R$, $i_{nR}$, the output referred noise can be similarly calculated to be:

$$i_{n3,diffout} = \frac{-1/R - g_{m1} - g_{m3}}{g_{m3}^2 - (1/R + g_{m1})^2} g_{m1} i_{n3} \tag{6. 8}$$

$$i_{nR,diffout} = \frac{-1/R - g_{m1} - g_{m3}}{g_{m3}^2 - (1/R + g_{m1})^2} g_{m1} i_{nR} \tag{6. 9}$$

Using the impedance matching condition (6. 1), and changing $g_{m1}$ to $2g_{m1}$ in (6. 7)-(6. 9) due to capacitive cross-coupling, we can rewrite (6. 7)-(6.9) to be

$$i_{n1,diffout} = (g_{m1}R - 1) i_{n1} \tag{6. 10}$$

$$i_{n3,diffout} = \frac{1}{2} R g_{m1} i_{n3} \tag{6. 11}$$

$$i_{nR,diffout} = \frac{1}{2} R g_{m1} i_{nR} \tag{6. 12}$$

The NF can then be written as

$$\begin{aligned} F &= 1 + \frac{i_{n1,diffout}^2 + i_{n3,diffout}^2}{i_{nR,diffout}^2} = 1 + \frac{(g_{m1}R - 1)^2 g_{m1}\gamma/\alpha + (g_{m1}R)^2 g_{m3}\gamma/\alpha}{g_{m1}^2 R} \\ &= 1 + \frac{\gamma}{\alpha}\left(\frac{(g_{m1}R - 1)^2}{g_{m1}R} + 2g_{m1}R - 1\right) \end{aligned} \tag{6. 13}$$

The minimum NF is achieved when

$$g_{m1} = 1/(\sqrt{3}R) \tag{6. 14}$$

and the minimum NF is $F_{\min} = 1 + (2\sqrt{3} - 3)\gamma/\alpha = 1 + 0.464\gamma/\alpha$ (6. 15)

It is seen that the condition to cancel noise from M1 and M2 ($g_{m1}R$-1=0) is not the optimum condition for the overall NF, indicating the importance of global optimization considering the circuit as a whole. The technique avoids the need of large RF choke inductor to bias the common-gate stage at low-frequency, and even improves NF a bit ($1 + 0.5\gamma/\alpha$ from capacitive cross-coupling common-gate stage) despite the added current sources M3 and M4, so is suitable for our application. $g_{m3}$

can be calculated to be $g_{m3} = (2/\sqrt{3} - 1)/R$ based on (6. 1), which is much smaller than $g_{m1}$, the stability of the amplifier is therefore guaranteed.

*C. Design Considerations on Bandwidth and Linearity*

As discussed in Chapter 4, T-coil and inductive series peaking is very useful to extend output bandwidth, in addition, they can greatly reject the noise of loading resistors from arriving at the output. After noise canceling in active devices, loading resistors contribute significant noise, thus T-coil and inductive series peaking is helpful in reducing NF at high-frequency band.

IM2 injection technique for linearization is implemented in the second stage. A squaring circuit, composed of MP1, MP2 and R7, is used to generate the desired low-frequency IM2 tone of the input signal. The IM2 tone, with proper magnitude and phase, will be converted to current in M11, and mix with the fundamental tone through 2$^{nd}$-order nonlinearity in M9 and M10, generating IM3 that is anti-phase and equal-magnitude with intrinsic IM3 of M9 and M10 for IM3 cancellation. The injected low-frequency IM2 at the gate of M11 should be in-phase with the envelope of input two tones for best IM3 cancellation. The generated IM2 by the squaring circuit inherently satisfies the phase requirement as long as IM2 frequency is low enough so that its phase is not affected by R7 and Cp1. The low-pass cut-off corner is set to be 300 MHz for R7 and Cp1. The linearization will be used in the narrowband standards when the signal occupies 100-200 MHz bandwidth.

The gain can be varied by changing the switches S1 and S2 in the first stages, and current steering pair M5-M8 in the second stage. In the second stage, the voltage headroom is limited due to large current used to reduce the noise contribution and due to 3 devices stacking. To facilitate low voltage operation, MP3 and MP4 serve as current-steering to share part of the current from loading resistor. Common-mode

feedback (CMFB), which is composed of an opamp, MP3, and MP4, sets the output DC voltage.

### 6.2.3 Simulation Results

Fabricated in a 0.13-μm CMOS process, and operated at 1.2-V supply, the LNA consumes 10 mW. The gain is around 20 dB, with 3 dB ripple in 3.1-10.6 GHz band, as shown in Fig. 6. 4. NF is less than 3.4 dB, with a minimum of 2.8 dB at 8 GHz, as shown in Fig. 6. 5. IIP3 is -0.6 dBm at 5 GHz, with 6.7 dB improved by the linearization. The linearization doesn't affect both gain and NF, and consumes only 0.1 mA. S11 is better than -10.5 dB within the whole band, as shown in Fig. 6. 6. The chip area is 0.8*1.0 $mm^2$.

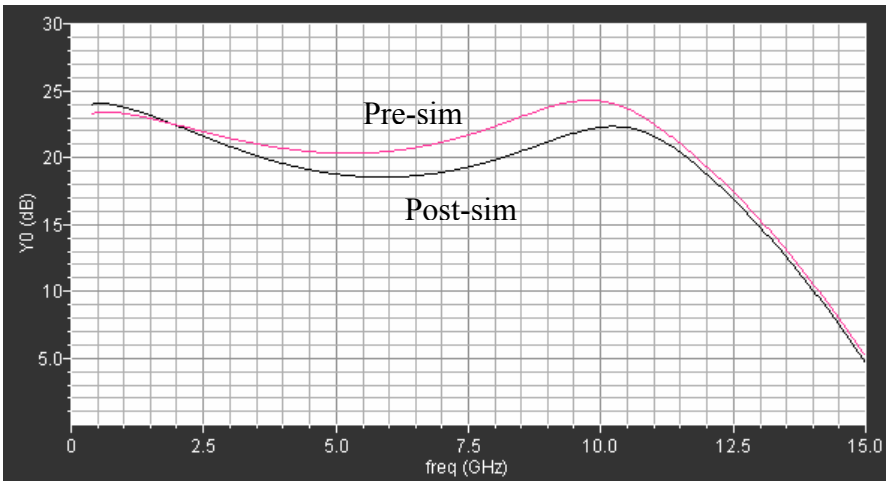



Fig. 6. 4 Simulated gain of the LNA

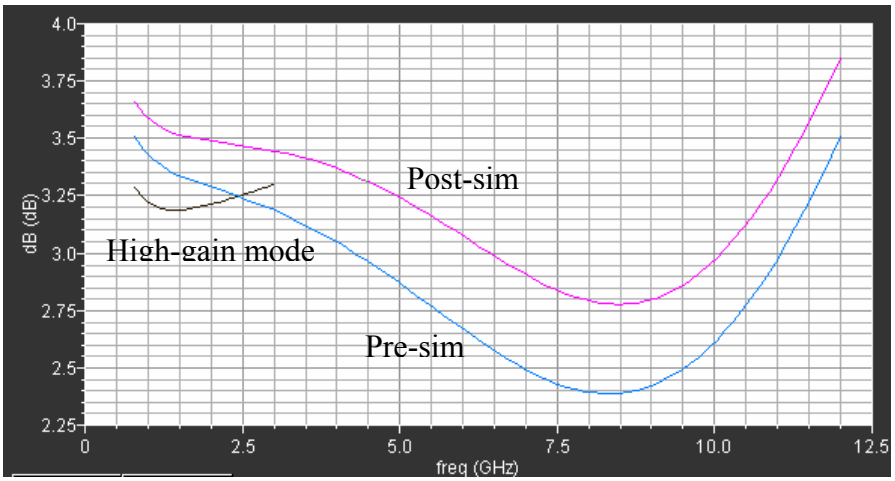



Fig. 6. 5 Simulated NF of the LNA

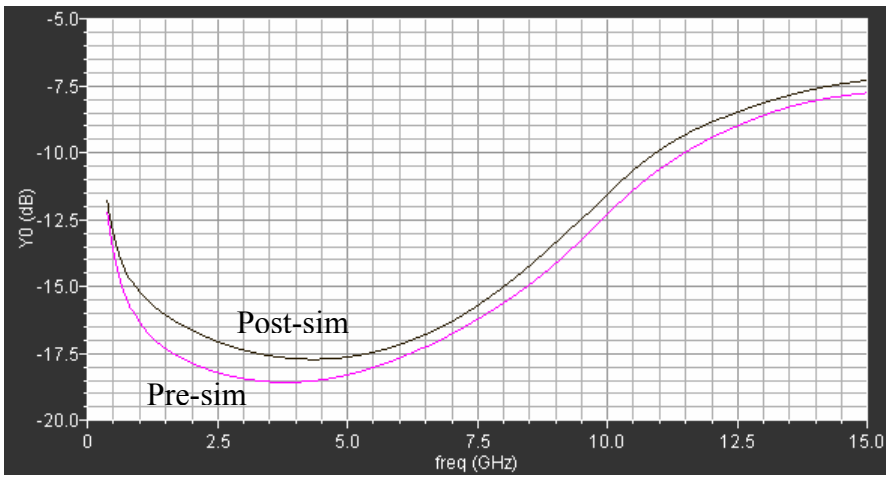



Fig. 6. 6 Simulated S11 of the LNA

### 6.2.4 Measurement Results

The measurement results of gain, NF, and S11 are shown in Fig. 6. 7, Fig. 6. 8, and Fig. 6. 9, respectively.

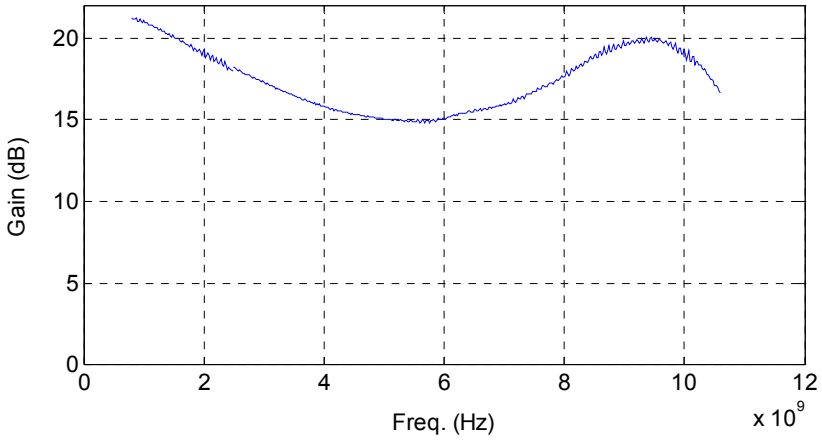



Fig. 6. 7 Measured gain of the LNA

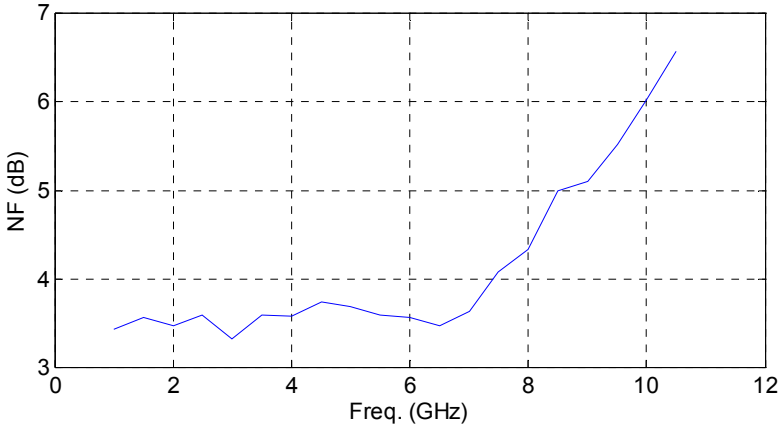



Fig. 6. 8 Measured NF of the LNA

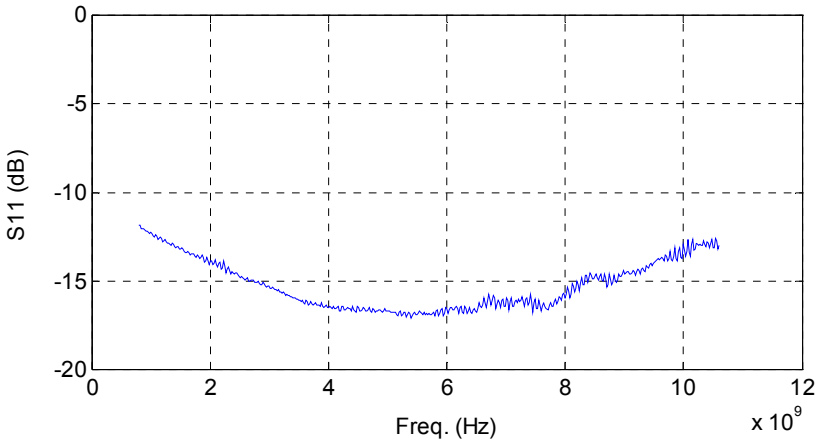



Fig. 6. 9 Measured S11 of the LNA

Compared with simulation results, measured bandwidth is narrower, resulting in rapid NF increase at high frequency. T-coil and series peaking inductor in schematic of Fig. 6. 1 are measured separately, their parasitic capacitors are much larger than those predicted by Momentum EM simulation, as compared in TABLE 6. 1 (their models are shown in Fig. 6. 10).

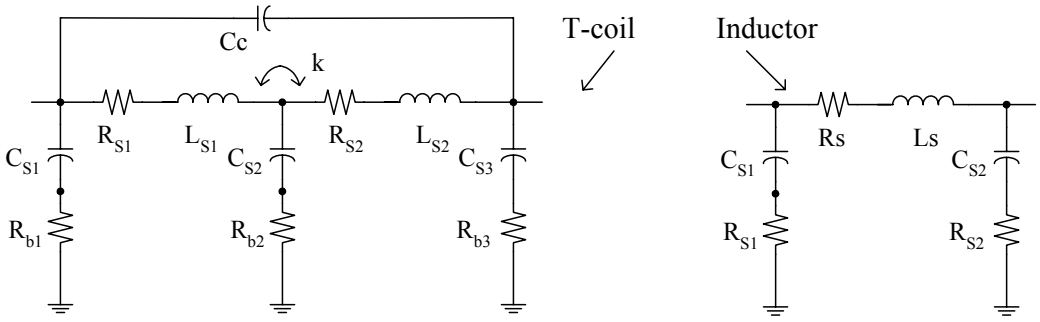



Fig. 6. 10 T-coil and inductor model

TABLE 6. 1 INDUCTOR AND T-COIL MODEL PARAMETERS

| **Inductor** | Ls/nH | Rs/Ω | $C_{S1}$/fF | $R_{S1}$/Ω | $C_{S2}$/fF | $R_{S2}$/Ω |
|---|---|---|---|---|---|---|
| Simulation | 4.2 | 8.6 | 8.8 | 493 | 9.2 | 419 |
| Measurement | 4.04 | 10.1 | 28.5 | 306 | 28.5 | 306 |

| **T-coil** | $L_{S1}$/nH | $R_{S1}$/Ω | $L_{S2}$/nH | $R_{S2}$/Ω | k | Cc/fF |
|---|---|---|---|---|---|---|
| Simulation | 3.69 | 8.24 | 3.28 | 8.10 | 0.189 | 2.0 |
| Measurement | 3.7 | 7.47 | 3.2 | 10.9 | 0.198 | 11.5 |
| | | | | | | |
| **T-coil** | $C_{S1}$/fF | $R_{b1}$/Ω | $C_{S2}$/fF | $R_{b2}$/Ω | $C_{S3}$/fF | $R_{b3}$/Ω |
| Simualtion | 14.5 | 2.92 | 23.6 | 941 | 10.1 | 1.04 |
| Measurement | 41.8 | 182 | 65.6 | 106 | 41.8 | 182 |

Possible reason for the increase of parasitic capacitors: there is deep N-well under inductor and T-coil structure in the layout, but it is not modeled in the Momentum EM simulation. Possible solution is to remove deep N-well in the future run.

With the measured models of inductor and T-coil, NF increases from 3.3 dB to 5.2 dB at 10.6 GHz in simulation. The simulated 1st stage gain and overall NF with simulated and measured T-coil models are shown in Fig. 6. 11. Due to the reduction

of 1st stage gain at high-frequency, 2nd stage has more noise contribution and makes NF increase rapidly at high frequency.

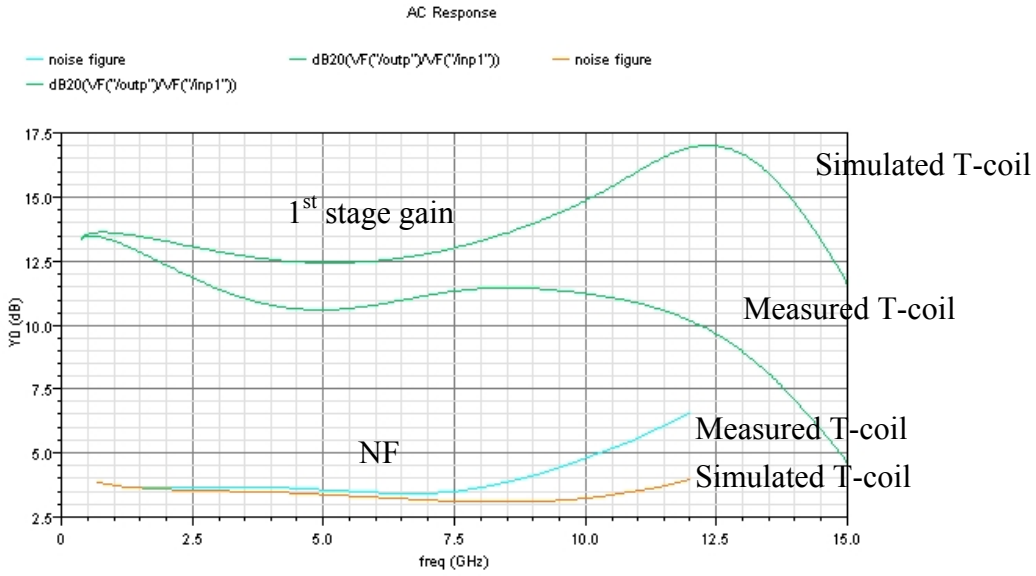



Fig. 6. 11 Simulated 1st stage gain and overall NFwith different T-coil models

TABLE 6. 2 summarizes the performance comparison with other wideband LNAs. The performance is comparable with others except a large NF beyond 8 GHz.

TABLE 6. 2 PERFORMANCE COMPARISON OF THE WIDEBAND LNA

| | [6. 2] | [6. 3] | [6. 9] | This work |
|---|---|---|---|---|
| Frequency/GHz | 0.1-6.5 | 1-7 | 3.1-10.6 | 0.9-10.6 |
| S11/dB | <-10 | <-10 | <-9.9 | <-12 |
| Gain/dB | ~18 | >16 | ~15.1 | >15 |
| NF/dB | 2.9-4.2 | 2.7-3 (1~3 GHz) | 2.1-3 | 3.5-6.5 |
| IIP3/dBm | 1 | -4.1 | -5.7 | -0.6 |
| Supply/V | 1.8 | 1.4 | 1.2 | 1.2 |
| Power/mW | 11.7 | 25 | 9 | 10 |
| Process | 0.13-μm CMOS | | | |

## 6. 3. Issues in SDR Mixer Design

Mixer is a more challenging block in SDR receiver, and needs more careful derivation on its specifications based on the targeted standards than just designing for one specific standard. The specifications include 1/f noise corner, IIP3, IIP2, and gain.

*A. Low 1/f Noise*

Most of the receivers use direct-conversion frequency plan to avoid image problem and achieve high level of integration. However, the down-converted signal located at DC and low-frequency at mixer output suffers from 1/f noise. In some standards, such as MBOA UWB, there is 2 MHz bandwidth around DC not occupied by the signal, 1/f noise does not significantly affect signal quality. However, in narrowband applications such as paging or GSM, the mixer 1/f noise is very likely to overwhelm the down-converted signal, and as a consequence, limits the overall receiver NF.

There exist a few mixer topologies to achieve low 1/f noise corner. In [6. 10], a current buffer is proposed to work as a low impedance load at the switches output, while not sacrificing the gain by using large loading resistor at buffer output. The switches are biased with no DC current to further improve the 1/f noise corner.

In [6. 11], a 1/f noise reduction technique based on physical understanding of noise mechanisms in active mixers is proposed. Since the 1/f noise pulses are only present at the switching instant of the LO differential pairs, by injecting a dynamic current equal to the bias current of each pair at only the switching event, it can eliminate the output flicker noise component. The proposed technique achieves a 1/f noise corner of almost an order of magnitude lower than that of a standard implementation, without penalty in the linearity, gain, or power consumption.

*B. IIP2 and IIP3*

If we use a wideband LNA for SDR receiver, upon entering the receiver, far away channels are filtered for the first time at the mixer load. Second- and third-order

nonlinearity in the LNA and mixer can cause unwanted channels to corrupt the desired channel through AM detection and cross-modulation, respectively.

To receive a -99 dBm GSM signal in the presence of a -15 dBm unwanted 802.11g channel, the receiver IIP2 must be at least +61 dBm. On the other hand, if the interferer is changed to a -15 dBm WCDMA unwanted signal, IIP2 of +48 dBm is required to receive a wanted GSM channel at -99 dBm with a margin of 6 dB [6. 10]. So the IIP2 requirement on the front-end is quite complicated and needs careful planning. The mixer should be designed to meet the most stringent requirement of all the situations.

Similarly, due to the cross-modulation from third-order nonlinearity, for a -99 dBm GSM channel to be received in the presence of an unwanted 802.11g channel of -15 dBm, the front-end IIP3 must be -14dBm; while the same GSM signal can be detected in the presence of a -15 dBm WCMA unwanted signal in a receiver with IIP3 of -18 dBm [6. 10].

So both IIP2 and IIP3 of the mixer should be as good as possible because we face various conditions when the front-end is in multi-standard use.

To achieve good IIP2, it needs a well matched layout, well matched devices. The tail current source should be added for the differential pair despite the reduced voltage headroom, in addition, common-mode feedback is found useful to improve IIP2 [6. 12].

For IIP3 linearization, IM2 injection technique is effective with little extra power consumption, but it may lead to worse IIP2. A way to alleviate IIP2 degradation problem in IM2 injection linearization needs to be found.

## REFERENCES

[6. 1] J. Mitola, “The software radio architecture,” *IEEE Commum. Mag*., Vol. 33, No. 5, pp. 26-38, May 1995.

[6. 2] S. Chehrazi, A. Mirzaei, R. Bagheri, and A. Abidi, "A 6.5 GHz wideband CMOS low noise amplifier for multi-band use," in *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 801-804, Sep. 2005.

[6. 3] R. Ramzan, S. Andersson, J. Dabrowski, and C. Svensson, "A 1.4V 25mW inductorless wideband LNA in 0.13μm CMOS," *IEEE International Solid-State Circuits Conference*, 2007.

[6. 4] A. Liscidini, M. Brandolini, D. Sanzogni, and R. Castello, "A 0.13 μm CMOS front-end, for DCS 1800/UMTS/802.11b-g with multiband positive feedback low-noise amplifier," *IEEE J. Solid-State Circuits*, Vol. 41, No. 4, pp. 981-989, Apr. 2006.

[6. 5] S. Andersson, and C. Svensson, "A 750 MHz to 3 GHz tunable narrowband low-noise amplifier," *NORCHIP Conference*, pp. 8-11, Nov. 2005.

[6. 6] P. W. Lee, H. W. Chiu, T. L. Hsieh, C. H. Shen, G. W. Huang, and S. S. Lu, "A SiGe low noise amplifier for 2.4/5.2/5.7GHz WLAN applications," *IEEE International Solid-State Circuits Conference*, 2003.

[6. 7] W. Zhuo, S. Embabi, J. P. de Gyvez, and E. Sanchez-Sinencio, "Using capacitive cross-coupling technique in RF low noise amplifiers and down-conversion mixer design," *European Solid-State Circuits Conference*, pp. 77-80, Sep. 2000.

[6. 8] A. Amer, E. Hegazi, and H. F. Ragaie, "A 90-nm wideband merged CMOS LNA and mixer exploiting noise cancellation," *IEEE J. Solid-State Circuits*, Vol. 42, No. 2, pp. 323-328, Feb. 2007.

[6. 9] 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.

[6. 10] R. Bagheri, A. Mirzaei, S. Chehrazi, M. E. Heidari, M. Lee, M. Mikhemar, W. Tang, and A. A. Abidi, "An 800-MHz-6-GHz software-defined wireless receiver in 90-nm CMOS," *IEEE J. Solid-State Circuits*, Vol. 41, No. 12, pp. 2860-2876, Dec. 2006.

[6. 11] H. Darabi, and J. Chiu, "A noise cancellation technique in active RF-CMOS mixers," *IEEE J. Solid-State Circuits*, Vol. 40, No. 12, pp. 2628-2632, Dec. 2005.

[6. 12] M. Brandolini, M. Sosio, F. Svelto, "A 750mV 15kHz 1/f noise corner 51dBm IIP2 direct-conversion front-end for GSM in 90nm CMOS," *IEEE International Solid-State Circuits Conference*, 2006.

# Chapter 7 Conclusion and Future Work

## 7. 1. Problems Encountered and Proposed Solutions

### 7.1.1 Accuracy of Passive and Active Devices Model

In the WLAN LNA design, the inductance of gate inductor $L_g$ is measured to be only 1/3 of the expected value, the layout structure is shown in Fig. 7. 1. Even though the layout follows the design rules, the problem is due to the unexpected short between turns at the transition part of M5 and M6 to the sidewall M6. Due to the transition, M6 spreads and covers more area than the case without M5 under it, so the spacing between transition part to the nearby metal should be kept larger than the minimum allowed spacing to guarantee the yield of the device.

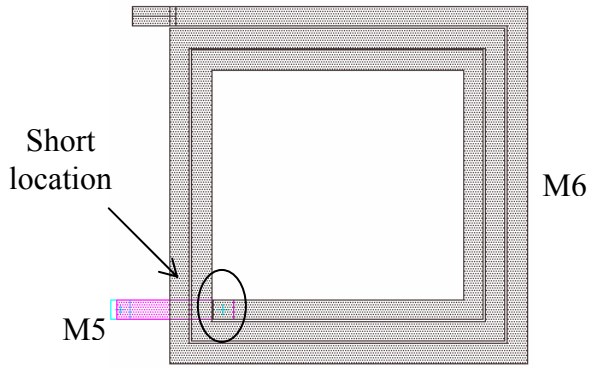



Fig. 7. 1 Layout of gate inductor in WLAN LNA

Active devices also need accurate models. In RF application, transistors are usually built in deep N-well, connecting source to bulk to get low threshold voltage, and avoid noise coupling. In addition to the parasitic capacitance which degrades the high frequency gain, the substrate effect can not be ignored at high frequencies. The substrate network for an NMOS transistor is shown in Fig. 7. 2.

In UWB LNA design, the NMOS transistors were first built in deep N-well, however, the measurement results deviated from simulation in both gain and

bandwidth. It is because mixed-signal MOSFET model does not model the substrate effect well.

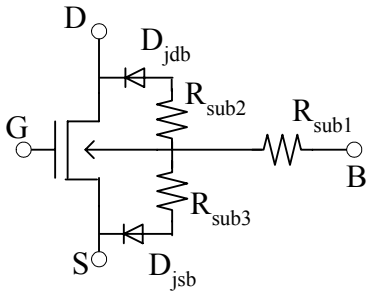



Fig. 7. 2 Substrate network for an NMOS transistor

Substrate effect degrades the gain by lowering the output impedance of the transistor. It can not be compensated by peaking techniques, and needs to be carefully modeled during design stage to predict the performance accurately. In measurement, amplifier implemented in deep N-well showed narrower bandwidth and smaller gain at high-frequency. In case deep N-well NMOS transistors are used, RF MOSFET model, instead of mixed-signal model, should be chosen in simulating circuit performance. It is helpful to place substrate contacts around the transistors as close as possible to reduce the substrate effect.

### 7.1.2 Cross-Coupling between Metal Connections

Cross-coupling between metal lines usually leads to signal corruption and ultimately entails in degradation of the overall performance of the whole circuit. For example, in the up-mixer design for UWB transceiver, low-frequency LO signal leaks to the output of the 1$^{st}$ mixer, which is close to the IF output spectrum and cannot be filtered.

The problem is caused by the cross-coupling between LO and IF lines of 1$^{st}$ up-mixer. Due to the floorplan limit, LO and IF lines go in parallel for long distance thus form the coupling path. To solve the problem, a shielding line connecting to ground is

inserted between LO and IF lines, as shown in Fig. 7. 3. The LO leakage power reduces from -30 dBm to -60 dBm with shielding metal line added.

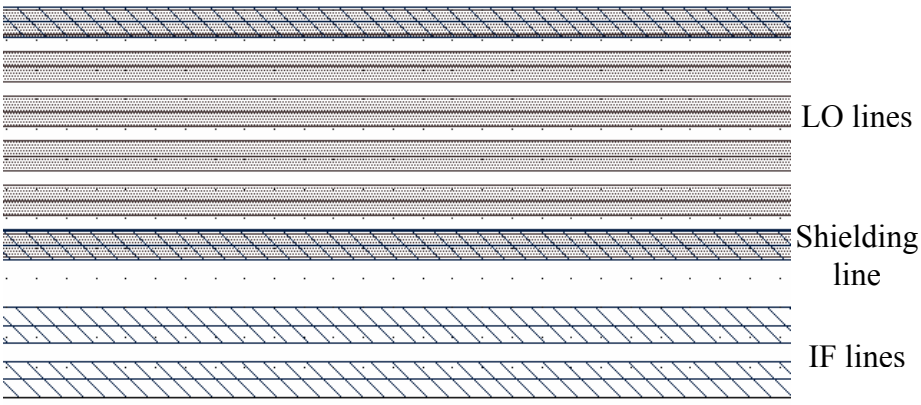



Fig. 7. 3 Shielding line inserted to reduce cross-coupling

## 7. 2. Contributions of the Dissertation

### 7.2.1 Novel Design of the UWB LNA

A low-power wideband LNA is designed for 9-band MB-OFDM receiver. The inductive series peaking is proposed and analyzed to extend the output bandwidth of the amplifier, it is also found useful in rejecting the noise of loading resistor. 3-stage topology is employed to enlarge gain-bandwidth product (GBW) of the amplifier. The differential LNA delivers larger than 22 dB gain over 3~8 GHz range with ~5.5 dB NF, and consumes only 15.5 mA from 1.5-V power supply.

### 7.2.2 Novel Mixer Design and System Integration for UWB Transceiver

The UWB transceiver operates in high-frequency wide-bandwidth, interface (floorplan and interconnection) between building blocks needs to be carefully considered and modeled, e.g., the interface between LNA and mixer, and the interface between mixer and synthesizer.

To relax the burden of the synthesizer, which operates at frequency as high as 10.56 GHz, mixer employs the novel topology with LO switches at the bottom of the

transconductors. LO switches are put as close to the synthesizer as possible to reduce the interconnections and thus the loading to the synthesizer. The effect of long interconnection between LO switches and transconductors in mixer are relaxed by proposed inductive series peaking. In addition, with the proposed mixer topology, it is convenient to share the LO switches between down- and up-conversion mixers, a combined down/up mixer is design to further reduce the parasitic capacitors and improve high-frequency performance. All the interconnections are modeled using ASITIC and included in simulation to make sure the performance of the whole transceiver.

The gain and linearity of the proposed mixer are analyzed. The proposed mixer is helpful in linearity.

### 7.2.3 Novel Design of Linearization Using 2nd Intermodulation Injection

The linearization technique is proposed using low-frequency 2nd intermodulation (IM2) injection to cancel 3rd intermodulation (IM3) of the transconductor. With proper phase and amplitude, the injected IM2 tone mixes with the fundamental tone though 2nd-order nonlinearity of the transconductor, and generates a IM3 that is anti-phase and equal-magnitude with intrinsic IM3 for cancellation.

The IM2 injection linearization technique doesn't affect both gain and NF of the amplifier, and consumes little power with simple extra circuitry. It is applicable to different applications because the injected signal is located at low-frequency but not RF frequency. It measures 17 dB IM3 suppression in the RFID LNA with only 1% extra power of the main amplifier.

### 7.2.4 Design and Implementation of WLAN and Cable TV LNA

Fully integrated low-voltage WLAN LNA at 5.25 GHz is designed and measured. It adopts conventional inductive degeneration topology. The noise analysis includes

the noise contribution from lossy components of on-chip inductors and cascode transistor. Based on the new noise analysis, the optimal input transistor width for minimum NF is calculated, and procedures for power constrained noise optimization are proposed.

Fully integrated wideband cable TV LNA is also designed and measured. Active shunt feedback is adopted to achieve wideband input matching with reasonable NF. Gain tuning is realized by current steering and negative $g_m$-cell impedance boosting. For good linearity at low-gain setting, current steering pair is carefully designed to make sure its current density is suitable for the linearity performance. The LNA measures less than 3 dB NF at high-gain setting over the frequency range of 50-864 MHz, and consumes 23 mA current from 1.8 supply.

## 7. 3. Potential Future Work

### 7.3.1 Narrowband Tunable LNA for SDR Receiver Front-End

Wideband LNA covering frequency band of 0.9-10.6 GHz is proposed in this thesis. However, due to the lack of narrowband filtering, strong interferer may block the input signal and desensitize the receiver, in addition, intermodulation IM3 and IM2 both need special attention to avoid signal corruption.

If LNA has narrowband filtering, the problems mentioned above can be relaxed by the selectivity of the filtering. However, tunable narrowband filtering is usually accompanied by sacrificing performance in gain or NF, e.g., using switchable capacitor / inductor array to change resonant frequency of the LC tank. The tuning range is also limited by the variable range of capacitor / inductor. More practical tuning mechanism and scheme are still to be found, and more ideal switches used in switchable capacitor / inductor array need to be designed.

### 7.3.2 Automatic Frequency Tuning

For the narrowband applications, inductors are usually used to resonate with capacitors. In the presence of process variation, the resonant frequency will shift, degrading circuit performance. Therefore, automatic frequency tuning is desired to compensate for the frequency shift.

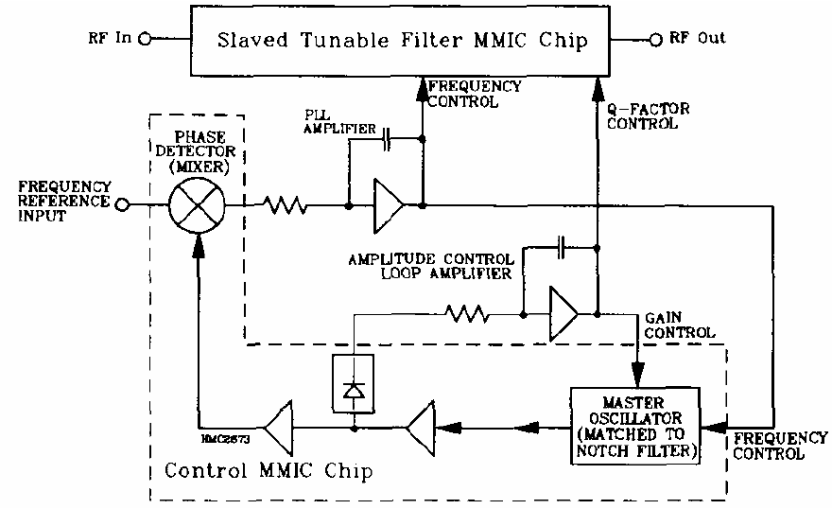



Fig. 7. 4 Master-slave scheme of automatic frequency tuning

Most existing designs on automatic frequency tuning use master-slave scheme. The main LC tank is duplicated and implemented in a master VCO, a PLL locks the oscillation frequency of the master oscillator to the desired circuit operation frequency. The frequency-control voltage generated by the PLL is also applied to the slave LC tank in the main circuit, so the LC tank center frequency is equal to the locked oscillation frequency of the master VCO.

Master-slave tuning scheme requires good matching between master and slave tanks, which poses limitations on this scheme. Direct frequency tuning need to detect and process phase information, and is still an open topic for research.

# Appendix A

## Differentail Impedance Measurement using Power Splitter

Most of the circuits in this thesis are in differential topology (differential-input differential-output). Due to the lack of 4-port network analyzer, 2 inputs and 2 outputs need to be combined first to do the measurement, as shown in Fig. 3. 19.

To verify the setup for impedance measurement with power splitter, comparison is made between two measurements (single-port measurement to measure the impedance): the first one is normal measurement, with calibration done at node 1, a loading with SubMiniature version A (SMA) connector is used as device-under-test (DUT) and is connected to node 1 for measuring its impedance. The second measurement is to include the power splitter in the calibration, which is the case in our real measurement. Calibration is done at power splitter output (using two sets of calibration kits with the same model No.), two identical loadings as used in the first measurement are connected at power splitter output. Two measurement results are shown in Fig. A. 2, as can be seen, two curves align well, showing the possibility of measuring differential circuit's input impedance using power splitter.

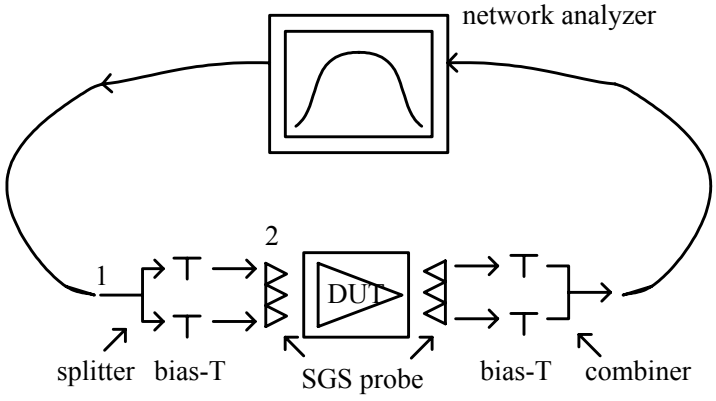



Fig. A. 1 Impedance and gain measurement

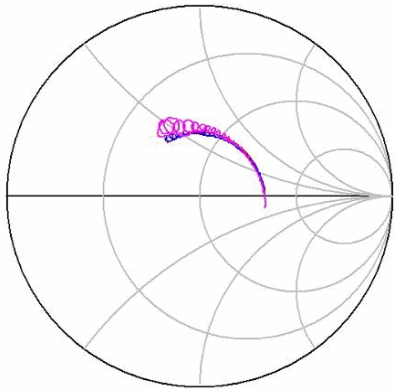


Fig. A. 2 Measured S11 plot with and without power splitter

# Appendix B

## NF Measurement of 75-Ω System with 50-Ω Testing Equipment

In the cable TV receiver system, the input impedance is matched to 75-Ω, however, the measurement equipment is matched to 50-Ω as a standard. The impedance mismatch between equipment and DUT will affect the measurement accuracy.

Wideband impedance transformer is need between measurement equipment and DUT. Resistive network as shown in Fig. A. 3 is one candidate. If we look right from source impedance $R_s$, we see 50-Ω; if we look left from load impedance $R_L$, we see 75-Ω.

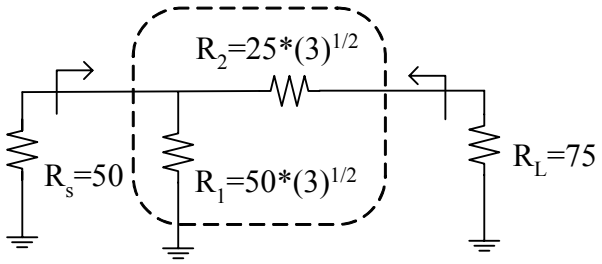



Fig. A. 3 Resistive impedance transformation network

The measured NF, $NF_{50}$, after adding the matching network can be de-embedded to find the DUT NF, $NF_{75}$, under 75-Ω environment. The relationship can be calculated as follows:

$$F_{50} = 1 + \frac{R_s^2 / R_1}{R_s} + \frac{R_2}{R_s}\left(\frac{R_L + R_2}{R_L}\right)^2 + (F_{75} - 1)\frac{R_L}{R_s}\left(\frac{R_L + R_2}{R_L}\right)^2 \quad \text{(A. 1)}$$

Simplifying (A. 1), we can get

$$F_{50} = (2 + \sqrt{3})F_{75} \Rightarrow NF_{50} = 5.72dB + NF_{75} \quad \text{(A. 2)}$$

So using the measured NF, $NF_{50}$, to subtract 5.72 dB, we can get the $NF_{75}$ we want to know.

# Appendix C

## Inductor Simulation and Measurement

On-chip inductor is one of the important passive components for RF circuit design. Inductor modeling and measurement are critical to predict the circuit performance and find the possible reasons for the failure of the circuit.

The steps of designing an inductor in this thesis are as follows:

1. **According to the desired inductor value, find the geometry parameters in ASITIC.**

   ASITIC is a fast simulation tool to predict the inductor model, it is useful in the initial inductor design. Launch ASITIC by typing

   ```
   > asitic_sun -t tsmc018.tek -8
   ```

   Following `-t`, `tsmc018.tek` is the technology file.

   Create a new file, e.g., named "try1", write the inductor geometry information in the file, e.g.,

   ```
   sq name=a: len=241: w=33: s=3: n=2: metal=m6: exit=m5:
   xorg=150: yorg=150
   ```

   where `sq` means square inductor, `len` is the out dimension, `w` and `s` are the metal width and spacing between turns, `xorg` and `yorg` specify the inductor location in the 2-dimensional plane. Load the file "try1" in ASITIC by `input` command:

   ```
   ASITIC> input try1
   ```

   then the following result shows, and we create an inductor called "a".

   ```
   ASITIC> sq name=a: len=241: w=33: s=3: n=2: metal=m6:
   exit=m5: xorg=150: yorg=150
   ```

   Simulate inductor "a" at 2.4 GHz using the command:

```
ASITIC> pix a 2.4
```

then the Pi model of the inductor is shown

```
Pi Model at f=2.40 GHz:  Q = 4.23, 4.26, 4.38
L  =     1 nH    R  =  3.35
Cs1=   120 fF    Rs1=   187
Cs2=  92.8 fF    Rs2=   372       f_res = 14.48GHz
```

Adjust the dimension `len`, metal width `w`, and number of turns `n`, to get the desired inductance value, and optimize the `Q` value by locating the peak `Q` of the inductor at the operating frequency.

More ASITIC commands can be found in the web:

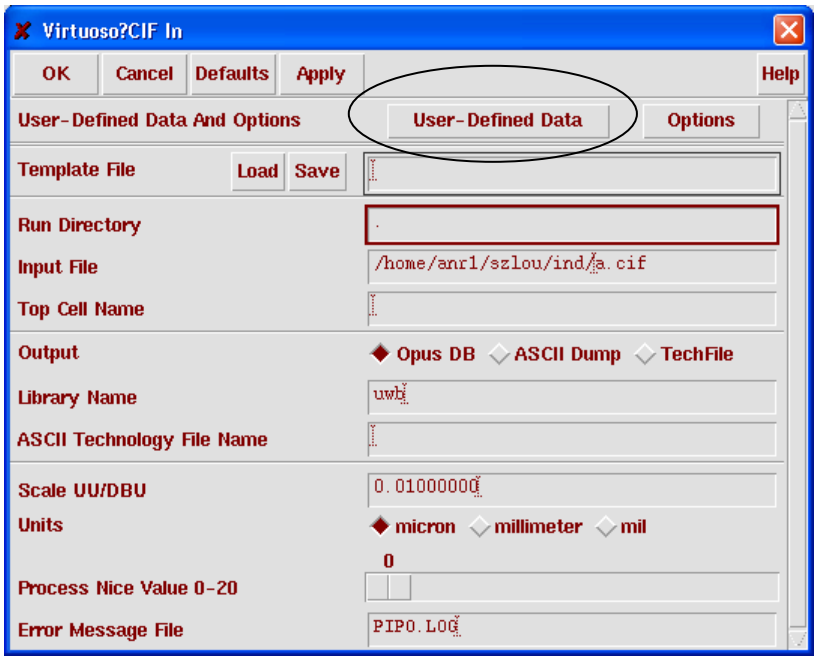
http://rfic.eecs.berkeley.edu/~niknejad/asitic.html

**2. Import the designed inductor to cadence.**

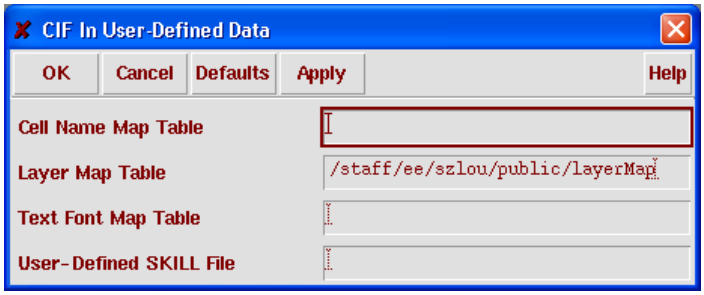
Use the command `cifsave` to save the inductor design, then import it to cadence to generate the corresponding layout, e.g.,

```
ASITIC> cifsave a a.cif
```

In cadence, choose `File` -> `Import` -> `CIF`, and fill in the form as follows:



Click `User-Defined Data`, and fill in layer map table between cadence and ASITIC.




Then click `OK`, the layout will be generated in cadence.

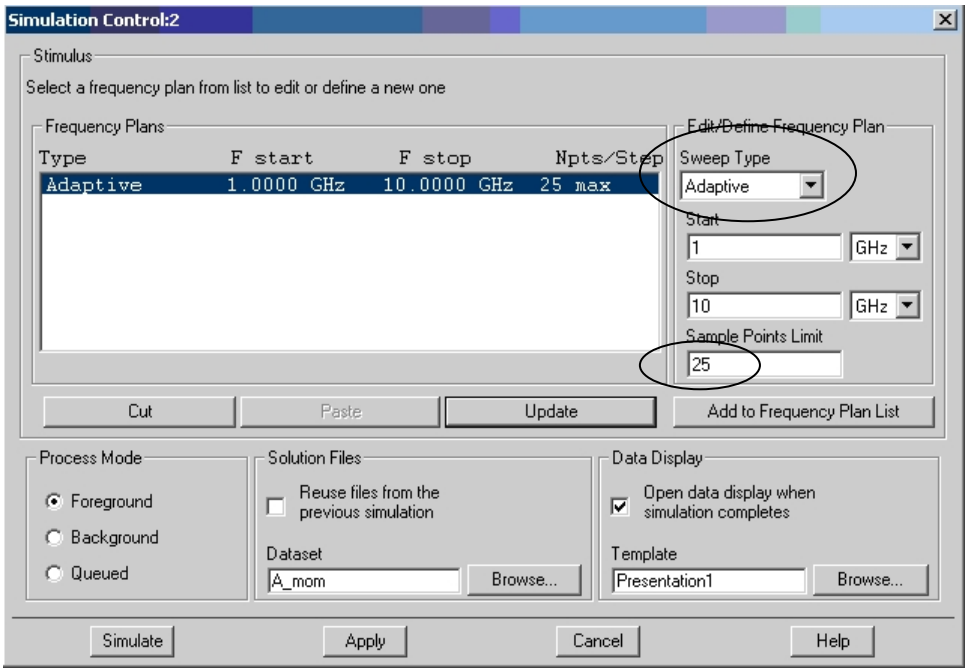
**3. Verify the design using Momentum.**

The inductor design is finished now. We will run a more accurate simulation in Momentum to get its π-model, which will be used in circuit simulation.

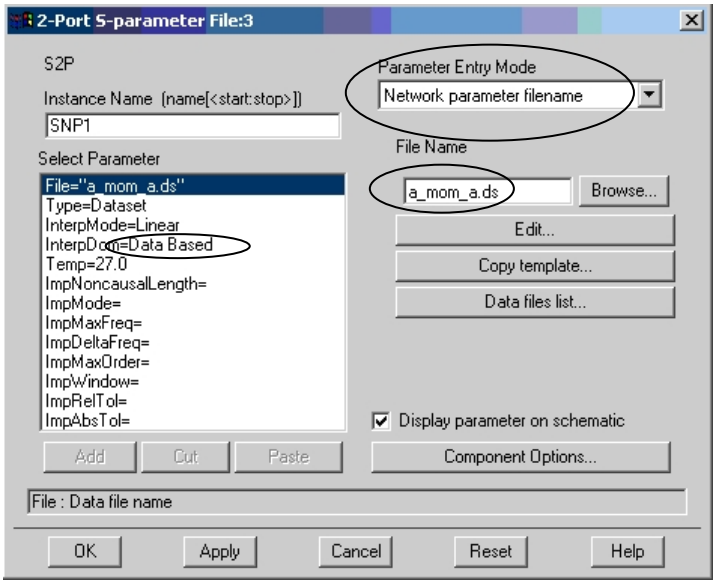
Export the inductor layout structure to .gds file, e.g., a.gds, by clicking `File` -> `Export` -> `Stream` in cadence, the forms are similar to the forms in `CIF In`. Open a new layout window in ADS, then import "a.gds" to ADS, manually changing the metal layers may be needed.

In ADS layout window, press `Options` -> `Layers` -> `Read`, to specify the layer table, e.g., tsmc018.lay. Then add contacts for m5 and m6 at the desired connection locations (the contacts imported from cadence don't work in ADS), by inserting rectangular "bond" and "hole", and a wire "resi". Enable `Midpoint Snap`, `Edge Snap`, `Grid Snap` in `Option` menu of layout window may help align the objects you are drawing.

To run Momentum simulation, add two ports at the inductor input and output ends, make sure the port layer is the same as the layer it is connected to. Then click `Momentum` -> `Substrate` -> `Open` to choose the substrate file, e.g. m23456.slm. Click `Momentum` -> `Simulation` -> `S-parameters`, the form can be filled in as follows. Then the simulation can be started.




After the simulation is done, the results are in S-parameters format, we need to fit the results to a lumped network. Open a new schematic window, insert a two-port `data item` in the schematic, relate it to the previous simulation results by specifying its property as follows:




Inductor model fitting schematic is shown in Fig. A. 4. Optimization tool is used, to fit the S-parameters (S11, S12, S21, S22, in both phase and magnitude) of the lumped component model to the EM Momentum simulation results. The lumped components values are variables, after the optimization, the lumped

network will have similar S-parameters response to the EM Momentum simulation in certain frequency range, which can be used in circuit simulation.

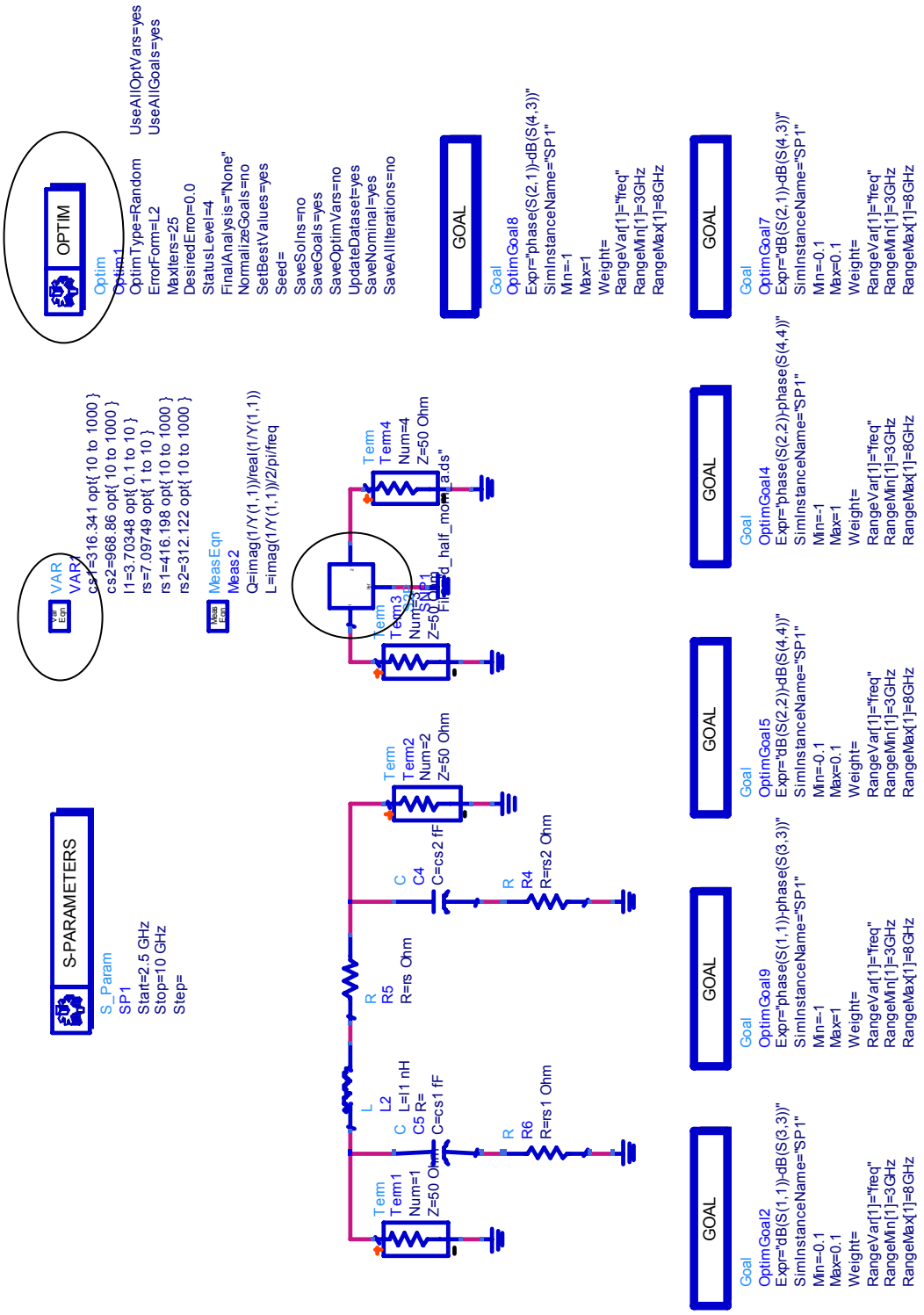



Fig. A. 4 Inductor model fitting schematic

To prepare the inductor testing structure, inductor ends are connected to the GSG pad. Most inductors use one port measurement, with one inductor end connected to S, another end connected to G. If two-port measurement is required, make sure to connect G terminal of two GSG pads with metal 1, as shown in Fig. A. 5.

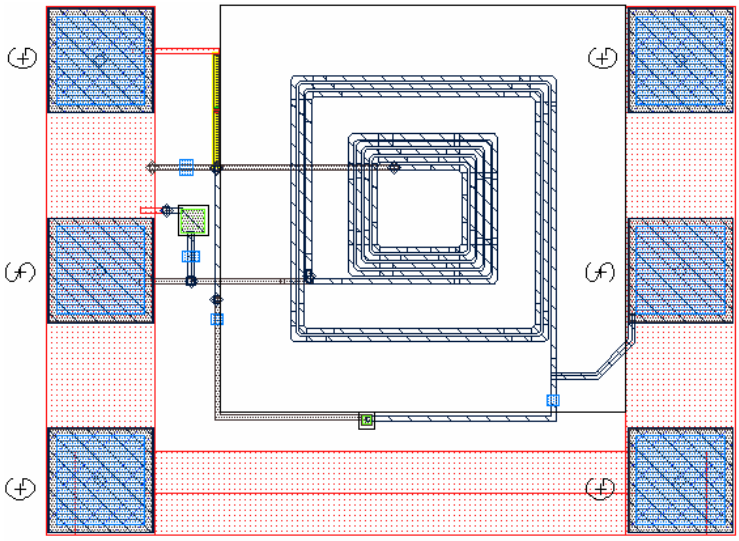
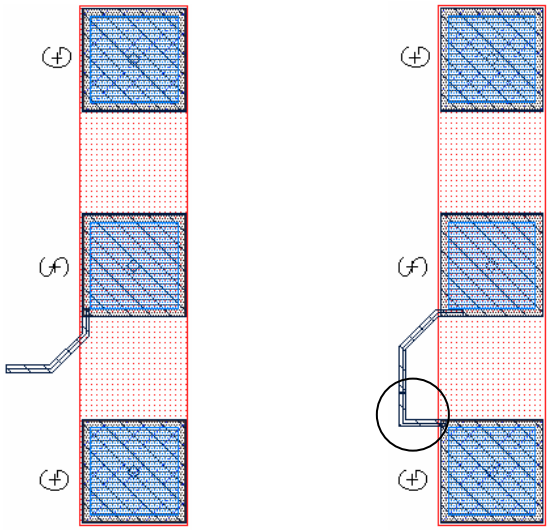


Fig. A. 5 2-port testing structure of a T-coil

Open and short structures are needed for de-embedding. Taking the right pad in Fig. A. 5 as an example, open and short structures are shown in Fig. A. 6.



(a) (b)

Fig. A. 6 (a) Open structure, and (b) short structure

Open and short structures can be used to measure the parasitic capacitance and inductance of the pad and wiring, respectively. In the short structure, the added wire to form the short path, as indicated in Fig. A. 6 (b), should be as short as possible in order to make measurement accurate.

The inductor measurement procedures are as follows:

**1. Network analyzer calibration.**

Set the network analyzer to the desired frequency range, input power, number of measurement points, and IF bandwidth first. If we use GGB GSG probe for measurement, GGB CS-5 substrate should be used for calibration. There are patterns `open`, `short`, `load`, and `through` on the substrate. Probe on each pattern one by one to do the corresponding calibration, after the calibration, do not raise the probe immediately. Check the calibration results by looking at S11 from the network analyzer, if it is less than -55 dB, we have confidence that the calibration is well done.

**2. Measure the testing structures.**

Measure both DUT and open / short structures. After the measurement is done, save the measured data in the floopy, in the `Define Disk-Save` menu, make sure to enable `data only`, and save the data to `ASII` format.

**3. Model fitting in ADS.**

The model fitting procedure is similar to that mentioned before, the only difference is to use the measured data instead of Momentum simulation data in `data item`.

De-embedding can be done in this way (assume 1-port case):

$Y_{de,open,DUT}=Y_{11,DUT}-Y_{open}$ */ de-embed DUT from open

$Z_{de,open,DUT}=ytoz\ (Y_{de,open,DUT})$ */ ytoz is a built-in function for Y- to Z-parameters transformation

$Y_{de,short}=Y_{11,short}-Y_{open}$ */ de-embed short from open

$Z_{de,short}$=ytoz ($Y_{de,short}$)
$Z_{de,DUT}$=$Z_{de,open,DUT}$-$Z_{de,short}$ */ de-embed DUT from short
$S_{de,DUT}$=ztos ($Z_{de,DUT}$)

$S_{de,DUT}$ is the de-embedded measurement result to use in model fitting.