國立交通大學

電子工程學系 電子研究所碩士班

碩 士 論 文

應用於超寬頻系統之

低功率、高線性度射頻接收器

Low Power and High Linearity RF Receiver Circuit for Ultra Wideband System Application

研究生：賴俊豪

指導教授：郭建男 教授

中 華 民 國 九十八 年 八月

應用於超寬頻系統之低功率、高線性度射頻接收器

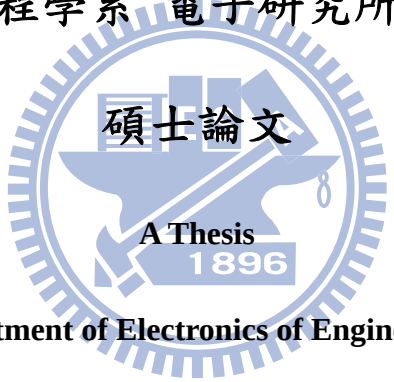
Low Power and High Linearity RF Receiver Circuit for Ultra Wideband System Application

研究生: 賴俊豪　　Student: Chuan-Hao Lai

指導教授: 郭建男　　Advisor: Chien-Nan Kuo

國立交通大學

電子工程學系 電子研究所碩士班



碩士論文

A Thesis

Submitted to Department of Electronics of Engineering & Institute of Electronics College of Electrical Engineering and Computer Engineering

National Chiao Tung University

In Partial Fulfillment of the Requirements

For the Degree of

Master

In

Electronic Engineering

August 2009

Hsinchu, Taiwan, Republic of China

中華民國 九十八年 八月

應用於超寬頻系統之
低功率、高線性度射頻接收器

學生 ： 賴俊豪　　指導教授 ： 郭建男　教授

國立交通大學

電子工程學系　電子研究所碩士班

摘要

本篇論文利用到複數轉導分析方法，對電晶體提出轉導等效小訊號電路模型，藉由此分析方法應用於多閘級電晶體架構之線性度分析，相較於之前所提出的分析方法，在此複數轉導分析下可以提供寬頻的線性度提升特性，以利於改善寬頻放大器電路之線性度。根據此種分析方法，超寬頻放大器與前端接收電路分別經由晶片製作來驗證。

第一顆晶片應用於超寬頻系統之-5dBm 之輸入第三階交會點超寬頻放大器利用到複數導數相消技術。量測結果顯示此一超寬頻放大器在線性度方面具有 5dBm 以上的提升，同時不用消耗多餘的功率。這顆超寬頻放大器在 8.26mW 功率損耗下，具有 8.1dB 之轉換增益，8dB 之輸入反迴損耗以及-5dBm 以上之輸入第三階交會點。

在第二顆晶片中，設計一個應用於超寬頻系統之低功率、高線性度前端接收端電路，此接收器電路包含了一個低雜訊放大器，一個主動相位分離器以及一個直接降頻混頻器。 模擬結果顯示此前端電路有 20.3dB 的轉換增益，10dB 之輸入反迴損耗以及-4.7dBm 之輸入第三階交會點，此外，此電路消耗功率為 10.4mW。

Low Power and High Linearity RF Receiver Circuit for Ultra Wideband System Application

Student : Chuao-Hao Lai Advisor : Chien-Nan Kuo

Department of Electronics Engineering & Institute of Electronics

National Chiao-Tung University

ABSTRACT

A compact equivalent circuit using a complex transconductance analysis is proposed for linearity design in the multiple gated transistors configuration. This complex transconductance analysis gives broadband linearity improvement in the common source amplifier as compared to previous published analysis. Following the complex transconductance analysis, an UWB LNA and an UWB RF front-end circuit were verified through two individual chips.

In the first chip, -5dBm IIP3 UWB low noise amplifier using complex derivative cancellation technique is analyzed and designed for ultra-wideband system. Measurement results show that the improvement of the linearity is more than 5dB without extra power consumption, and the designed LNA has conversion gain of 8.1dB, input return loss of 8dB, and input third-order intercept point (IIP3) of -5dBm with 8.26mW power dissipation.

In the second chip, a low power and high linearity UWB receiver intends to use in the receiver path of the ultra-wideband system. This front-end circuit is composed of a low noise amplifier, an active balun, and a direct down-conversion mixer. Simulation results show that the circuit has conversion gain of 20.3dB, input return loss of 10dB, IIP3 of -4.7dBm, while consuming only 10.4mW.

誌謝

能夠完成畢業論文，順利取得碩士學位，要感謝的人真的很多。首先是，我的父母親，能夠提供我一個沒有經濟負擔的環境一路往上念。再來，我要對我的指導教授郭建男教授至上誠摯的感謝。這兩年來，不僅提供了良好的研究環境與設備，並且讓我對射頻及類比電路的領域有深層的了解。適時的指點迷津，讓我學習到有邏輯的分析方法與嚴謹的研究態度。此外我要對昶綜, 明清和鴻源三位學長致上萬分的謝意，除了教導我許多軟體與硬體上的使用，也時常分享寶貴的研究與人生經驗，讓我一直有學習的好榜樣。

感謝實驗室的其他學長姐，鈞琳、俊興、燕霖、煥昇、易耕等的不吝指教與技術支援。感謝一起奮鬥一起研究的同學子超跟建忠，以及博一、根生、佑緯、勁夫、敬修、馳光、宇航等學弟，你們使實驗室不再是一個無聊的地方。

最後感謝女朋友惠中一路研究所的陪伴，謝謝你在這一年碩士生活中陪伴我走過許多人生的起起伏伏。在我研究上遇到瓶頸時，你的鼓勵與支持讓我走過種種一切的不如意。在我忙碌於研究時，你總是默默的陪伴在我身邊替我加油，讓我能無後顧之憂的向成功邁進。

賴俊豪

九十八年 八月

CONTENTS

FIGURE CAPTIONS

TABLE CAPTIONS

Chapter 1

Introduction

1.1 Ultra-Wideband Communication System

Ultra-wideband (UWB) is a wireless personal area network technology that transmits a low signal power over a very wide bandwidth and has the promising ability to provide high data rate at low cost with low power consumption. The Federal Communication Commission (FCC) has allocated 7.5GHz of spectrum for unlicensed use of UWB applications in the 3.1 to 10.6GHz frequency band. The multi-band UWB has greater flexibility in coexisting with other international wireless systems and future government regulators, and could avoid transmitting in already occupied bands. The multi-band OFDM based UWB, with fourteen 500MHz sub-bands and a frequency hopping scheme, is the possible approach to meet the requirements of IEEE 802.15.3a standard. Each of these bands must have a bandwidth greater than 500 MHz to obey the FCC definition of UWB. Fig. 1.1 shows the division of UWB frequency spectrum.

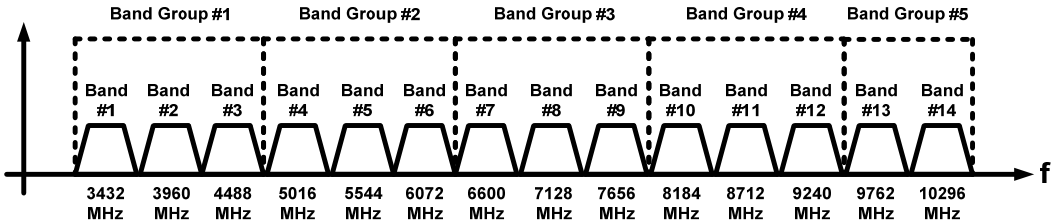


Fig. 1.1 Multiband spectrum allocation

1.2 Motivation

As introduced in Section 1.1, UWB is becoming more attractive for low cost consumer communication applications and that allows overlying existing narrowband systems to result in a much efficient use of the available spectrum. Thus, UWB is developed to provide a specification for a low complexity, low cost, low power consumption, and high data rate wireless connectivity among device within or entering the personal operating space and also addressed the quality of service capabilities required to support multimedia data types.

The direct down-conversion receiver becomes more attractive since it has the advantages of low complexity, low power, and less extra components. And it lets system-on-chip become possible. Typical, the first stage of the receiver is a low noise amplifier (LNA), which provides high gain and low noise to suppress the overall system's noise performance. Because LNA is dominating the linearity of the direct down-conversion receiver, we need high linearity to avoid the distortion of the signal. How to improve the linearity of the UWB LNA without extra power consumption is the main object of this thesis.

In the first one for UWB system application, the UWB LNA provides high gain, broadband input matching using dual reactive feedback topology, and broadband linearity improvement using multiple gated transistors (MGTR) technology. For MGTR analysis, we use a complex transconductance analysis for broadband linearity improvement design in this configuration. To design the UWB LNA with low power consumption using multiple gated transistors technology for broadband linearity improvement is the critical challenge for design.

In the second one for UWB system application, a low-power front-end circuit which includes a LNA, an active balun, and the Gilbert cell mixer is designed. The

LNA provides high gain and low noise to suppress the overall system's noise performance. The active balun provides broadband differential output signal for mixer. The double balance Gilbert cell mixer has an acceptable linearity for the receiver. To design an active balun for broadband differential output balance is critical challenge for design.

1.3 Thesis Organization

In the chapter 2 of the thesis, some basic concepts of RF design are introduced. These basic concepts which include the introduction of receiver architecture, noise and linearity provide the guidance for RF circuit design.

In the chapter 3 of the thesis, the design consideration of some circuit blocks which include LNA and mixer is introduced. Based on these circuits, a mixer and a front-end circuit are designed and verified in later chapter.

In the chapter 4 of the thesis, -5dBm IIP3 UWB LNA using complex derivative cancellation technique are designed. The multiple gated transistors configuration is introduced and a compact equivalent circuit using a complex transconductance is proposed for broadband linearity improvement design in this configuration. Following the above analysis, a low-power and high-linearity UWB LNA is designed. Finally, measurement result of the LNA chip fabricated by TSMC 0.18um CMOS technology is discussed.

In the chapter 5 of the thesis, a low-power, high linearity UWB front-end circuit is designed. The first stage is the LNA which is the same as that in chapter 4. The second stage is the active balun which provides the signal into differential form. The last stage is the double balance Gilbert cell mixer which has an acceptable linearity for the receiver. Overall front-end circuit is implemented.

In the chapter 6, the work is summarized and concluded.

Chapter 2

Basic Concepts in RF Circuit

2.1 Receiver Architecture

2.1.1 Heterodyne Receiver

A simple heterodyne architecture is shown in Fig. 2.1. This architecture is the most reliable reception technique today. But if the cost, complexity, integration and power dissipation are the primary criteria, the heterodyne receiver will become unsuitable due to its complexity and the need for a large number of external components.

In heterodyne architectures, the signal band is translated into much lower frequencies by a down-conversion mixer, and the filters are used to select the band and channel of interest. In general, a low noise amplifier is placed in front of the down-conversion mixer to suppress the noise of the down-conversion mixer.

Frequency planning is important in the heterodyne receiver. For high-side injection, an undesired signal (image) at the frequency of $\omega_{IM}=\omega_{LO}+(\omega_{LO}-\omega_{RF})$ is translated into the same intermediate frequency (IF) as the desired signal. Similarly, for low-side injection, the image frequency is at $\omega_{IM}=\omega_{LO}-(\omega_{LO}-\omega_{RF})$. Therefore the image would cause extra noise at the intermediate frequency. As shown in Fig. 2.2, some techniques are necessary to suppress the image, such as image reject filter. But here comes the question: how to choose the intermediate frequency? If $2\omega_{IF}$ is sufficiently large, the rejection of a image filter will have a relatively small loss in the signal band and a large attenuation in the image band. But a lower $2\omega_{IF}$ will relax the quality factor of the channel select filter to get great suppression of nearby interferers. Therefore a trade-off between image rejection and channel selection must be taken.

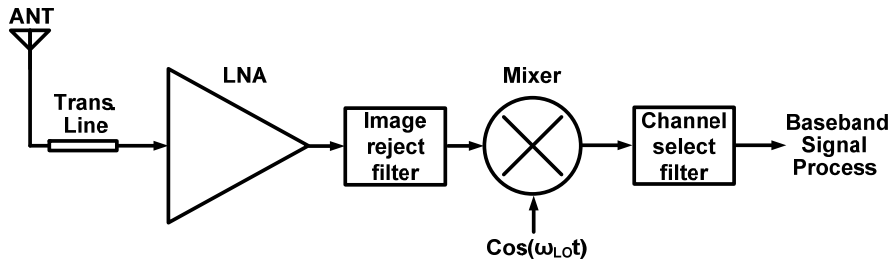


Fig. 2.1 Simple heterodyne architecture

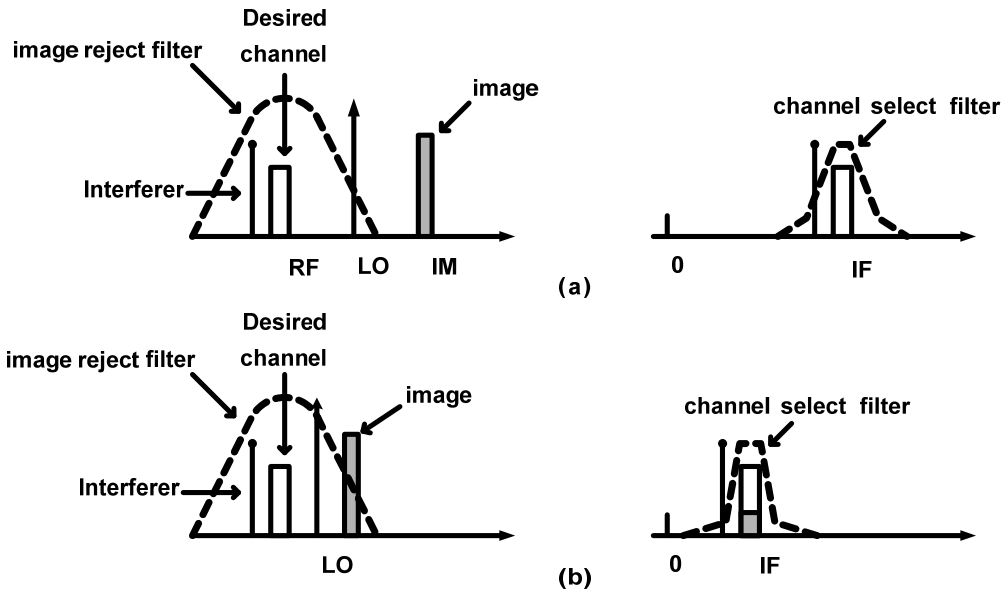


Fig. 2.2 Rejection of image versus suppression of interferers
(a) large ω_{IF} (b) small ω_{IF}

2.1.2 Homodyne Receiver

The homodyne receiver is also called "direct-conversion" or "zero-IF" architecture, since the RF signal is directly down-converted to the baseband in the first down conversion. In the homodyne receiver, the LO frequency is equal to the input carrier frequency, and channel selection requires only a low-pass filter with relatively sharp cutoff characteristics. The simple homodyne architecture is shown in Fig. 2.3. However, quadrature outputs are needed for frequency and phase-modulated signals, since the two sides of FM or QPSK spectra carry different information.

In recent years, this architecture becomes the topic of active research gradually and it may be due to the following reasons:

(1) The problem of image is removed due to $\omega_{IF} = 0$. Therefore no image filter is required, and the LNA need not drive a 50-Ω load.

(2) It is attractive for monolithic integration because this architecture needs less external components.

For the above reasons, this architecture is suitable for low-power and single-chip design. But some other issues that do not exist or are not as serious in a heterodyne receiver must be entailed, such as channel selection, DC offset, I/Q mismatch, even-order distortion, and flicker noise.

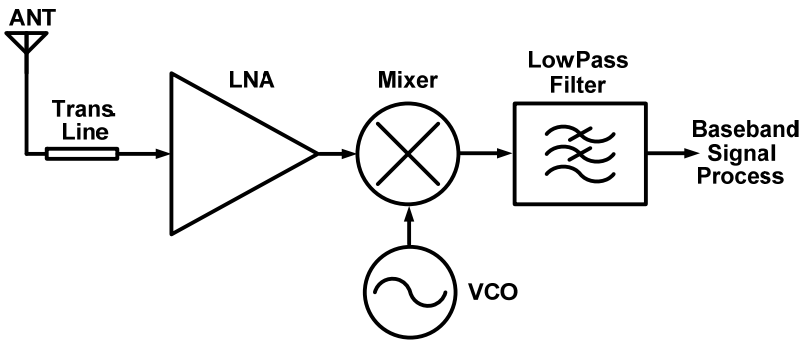


Fig. 2.3 Simple homodyne receiver architecture

2.2 Noise Basic

Noise can be generally defined as any random interference unrelating to the signal of interest, and it is characterized by a PDF and a PSD. In analog circuits, the signal-to-noise ratio (SNR), defined as the ratio of the signal power to the total noise power, is an important parameter. But in RF design, most of the front-end receiver blocks are characterized in terms of their noise figure, which is a measure of SNR degradation resulting from the added noise from the circuit/system but rather than the input-referred noise. Noise factor can be expressed as following:

$$\text{Noise Factor} = \frac{\text{Total output noise power}}{\text{Output noise power due to input source}} \qquad (2\text{-}1)$$

The noise figure (NF) is simply the noise factor expressed in decibels. If there is no noise in the system, then noise figure is 0 dB regardless of the gain. In reality, the finite noise of a system degrades the SNR, yielding noise figure > 0 dB. For those whose noise factor is close to unity, noise temperature (T_N) is an alternative way to express the effect of noise contribution. Noise temperature is the description of the noise performance in higher-resolution, and is defined as the required temperature increment for the source resistance. Noise temperature is calculated by all of the output noise at the reference temperature T_{ref} (which is 290 K). It relates to the noise factor as following:

$$\text{Noise Factor} = 1+\frac{T_N}{T_{ref}} \Rightarrow T_N = T_{ref}\cdot(\text{Noise Factor-1}) \quad (2\text{-}2)$$

2.2.1 Noise Source

Thermal noise:

Thermally agitated charge carriers in a conductor constitute a randomly varying current that gives rise to a random voltage resulting from the Brownian motion. Thermal noise is often called Johnson noise or Nyquist noise. The noise voltage has averaged value of zero, but a nonzero mean-square value.

In a resistor R, thermal noise can be represented by a series of noise voltage source $\overline{v_n^2} = 4kTR\Delta f$ or by a shunt noise current source $\overline{i_n^2} = \frac{4kT\Delta f}{R}$, where k is Boltzmann's constant (about 1.38×10^{-23} J/K), T is the absolute temperature in Kelvins, and Δf is the noise bandwidth. However, purely reactive elements generate no thermal noise.

Shot Noise:

Shot noise occurs in PN junctions, and there occurred two conditions for shot noise:

(1) There must be direct current flow.

(2) There must be energy barrier over which a charge carrier hops.

Charge comes in discrete bundles. The randomness of the arrival time gives rise to the whiteness of shot noise. Therefore the shot noise can be modeled by a shunt noise current source $\overline{i_n^2} = 2qI_{DC}\Delta f$, where q is the electronic charge, I_{DC} is the DC current in amperes, and Δf is the noise bandwidth in hertz.

Flicker Noise:

Flicker noise appears as 1/f character and is found in all active devices, as well as in some discrete passive element such as carbon resistors. In diodes, flicker noise is caused by traps associated with contamination and crystal defects in the depletion regions. The traps capture and release carriers in a random fashion and the time constants associated with the process giving rise to the 1/f nature of the noise power density. The flicker noise in diode can be represented as $\overline{i_j^2} = \frac{K}{f}\cdot\frac{I}{A_j}\cdot\Delta f$, where K is the process-dependent constant, A_j is the junction area, and I is the bias current. In MOSFET, charge trapping phenomena are invoked in surface, and its type of noise is much greater than that of the bipolar transistor. The flicker noise in MOSFET can be given by:

$$\overline{i_n^2} = \frac{K}{f}\cdot\frac{g_m^2}{WLC_{ox}^2}\cdot\Delta f \approx \frac{K}{f}\cdot\omega_T^2\cdot A\cdot\Delta f \qquad (2\text{-}3)$$

where K is the process-dependent constant, and A is the area of the gate.

2.2.2 Noise Model of MOSFET

The dominant noise source in CMOS devices is channel noise, which basically is thermal noise originating from the voltage-controlled resistor mechanism of a MOSFET. This source of noise can be modeled as a shunt current source in the output circuit of the device. The channel noise of MOSFET is given by

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \qquad (2\text{-}4)$$

where γ is bias-dependent factor, and g_{d0} is the zero-bias drain conductance of the device. Another source of drain noise is flicker noise and is given by Eqn. 2-3. Hence, the total drain noise source is given by

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f + \frac{K}{f}\cdot\frac{g_m^2}{WLC_{ox}^2}\cdot\Delta f \tag{2-5}$$

At RF frequencies, the thermal agitation of channel charge leads to a noisy gate current because the fluctuations in the channel charge induce a physical current in the gate terminal due to capacitive coupling. This source of noise can be modeled as a shunt current source between gate and source terminal with a shunt conductance g_g, and may be expressed as

$$\overline{i_{ng}^2} = 4kT\delta g_g\Delta f \tag{2-6}$$

where the parameter g_g is shown as

$$g_g = \frac{\omega^2 C_{gs}^2}{5g_{d0}} \tag{2-7}$$

and δ is the gate noise coefficient. This gate noise is partially correlated with the channel thermal noise, because both noise currents stem from thermal fluctuations in the channel and the magnitude of the correlation can be expressed as

$$c \equiv \frac{\overline{i_g\cdot i_d^*}}{\sqrt{\overline{i_g^2}\cdot\overline{i_d^2}}} \approx 0.395j \tag{2-8}$$

where the value of 0.395j is exact for long channel devices. Hence, the gate noise can be re-expressed as

$$\overline{i_{ng}^2} = \overline{(i_{ngc}+i_{ngu})^2} = 4kT\delta g_g\Delta f\,|c|^2 + 4kT\delta g_g\Delta f(1-|c|^2) \tag{2-9}$$

where the first term is correlated and the second term is uncorrelated to channel noise. From previous introduction of MOSFET noise source, a standard MOSFET noise model can be presented in Fig. 2.4, where $\overline{i_{nd}^2}$ is the drain noise source, $\overline{i_{ng}^2}$ is the

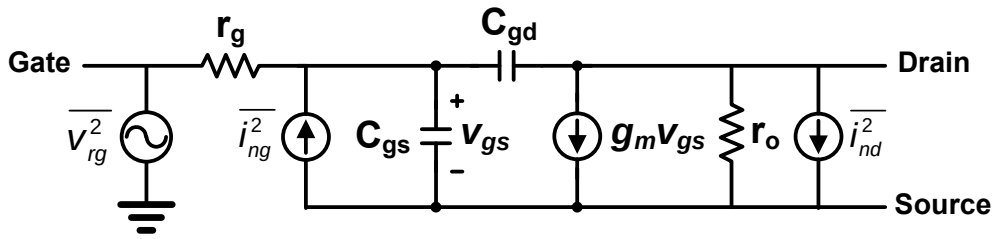


Fig. 2.4 A standard noise model of MOSFET

gate noise source, and $\overline{v_{rg}^2}$ is thermal noise source of gate parasitic resistor r_g.

2.2.3 Noise Figure of Cascaded Stages

For a cascade of m stages, the overall noise figure can be characterized by Friis formula

$$NF_{total} = 1 + (NF_1 - 1) + \frac{NF_2 - 1}{A_{p1}} + \ldots + \frac{NF_m - 1}{A_{p(m-1)}} \quad (2\text{-}10)$$

where NF_n is the noise figure of stage n, and A_{pn} denotes the power gain of stage n. This equation indicates that the noise results from the decrease in each stage as the gain preceding the increase in stage. Hence, the first few stages in a cascade are the most critical for noise figure. But if a stage exhibits attenuation, then the noise figure of the following circuit is amplified when referred to the input of that stage.

2.2.4 Noise Factor of a Two Port Network

Noise factor F is a useful measure of the noise performance of a system. It is defined as the ratio of the available noise power P_{no} at output divided by the product of the available noise power at input P_{ni} which times the networks's numeric gain G, or equivalently defined as the ratio of the signal to noise power at the input to the signal to noise power at the output .

$$F = \frac{P_{no}}{P_{ni} G} = \frac{S_i / N_i}{S_o / N_o} \quad (2\text{-}11)$$

The noise factor is a measure of the degradation in signal to noise ratio due to the noise from the system itself. Since the noise factor relates to the input noise power, a

standardized definition of noise source has been setup: a resistor at 290K. A more general expression of noise factor NF is called noise figure which is just noise factor expressed in decibels:

$$NF = 10\log F \tag{2-12}$$

When several networks are cascaded, each has its own gain G_i and noise factor F_i. The total output of the noise is composed of all the noise from each stage but with different amount of contribution to the noise performance. The noise factor of a cascade networks is given as

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_2} + \ldots \tag{2-13}$$

From (2-13), the noise factor of the first stage is most critical and must be keep as low as possible and its gain should be as large as possible to suppress the noise in the following stage. The result is intuitive since there is less interference of noise effect when the signal level is high.

2.2.5 Optimum Source Impedance for Noise Design

The noise factor of a two port network can be given as [5]

$$F = F_{\min} + \frac{4R_n}{Z_o} \frac{|\Gamma_s - \Gamma_{opt}|}{(1-|\Gamma_s|^2)|1+\Gamma_{opt}|^2} \tag{2-14}$$

where R_n is the correlation resistance which showed the relative sensitivity of the noise figure to departures from the optimum conditions, and Z_o is the characteristic impedance of the system. This equation expresses that there exists an optimum source reflection coefficient (Γ_{opt}) or equivalently an optimum source impedance (Z_{opt}) at the input of the network in order to deliver lowest noise factor ($F_{\min}$). The value of Γ_s provides a constant noise factor value forming non-overlapping circles on the Smith chart. It is usually the case that the optimum noise performance trades with the maximum power gain.

2.3 Linearity and Nonlinearity

All electronic circuits are nonlinear: a fundamental truth of electronic engineering. The linear assumption underlying most modern circuit theory is in practice only an approximation. Some circuits, such as small-signal amplifiers, are only very weak nonlinear; however, they are used in systems as if they were linear. In these circuits, nonlinearities are responsible for phenomena that degrade system performance and must be minimized. Other circuits, such as frequency multipliers, exploit the nonlinearities in their circuit elements; these circuits would not be possible if nonlinearities did not exist. Among these, it is often desirable to maximize the effect of the nonlinearities and even to maximize the effects of annoying linear phenomena. The problem of analyzing and designing such circuits is usually more complicated than for linear circuits, and it is the subject of much special concern.

Linear circuits are defined as those for which the superposition principle holds. Specifically, if excitations x_1 and x_2 are applied separately to a circuit having responses y_1 and y_2 respectively, the response to the excitation ax_1+bx_2 is ay_1+by_2, where a and b are arbitrary constants. This criterion can be applied to either circuits or systems.

This definition implies that the response of a linear and time-invariant circuit of system includes only those frequencies present in the excitation waveforms. Thus, linear and time-invariant circuits do not generate new frequencies. As nonlinear circuits usually generate a remarkably large number of new frequency components, this criterion provides an important dividing line between linear and nonlinear circuits.

Nonlinear circuits are often characterized as either strongly nonlinear or weakly nonlinear. Although these terms have no precise definitions, a good working

distinction is that a weakly nonlinear circuit can be described with adequate accuracy by a Taylor series expansion of its nonlinear current/voltage (I/V), charge/voltage (C/V), or flux/current (Φ/I) characteristic around some bias current or voltage. This definition implies that the characteristic is continuous and has continuous derivatives. And for most practical purposes, they do not require more than a few terms in its Taylor series. Virtually all transistors and passive components satisfy this definition if the excitation voltages are well within the component's normal operating ranges; that is, below saturation.

2.4 Nonlinear Phenomena

2.4.1 Harmonic Generation

Assumption of the current nonlinear element is given by the expression:

$$I = aV + bV^2 + cV^3 \tag{2-15}$$

where a, b, and c are constants, real coefficients. Assuming that V_s is a two-tone excitation of the term:

$$V_s = v_s(t) = V_1 \cos(\omega_1 t) + V_2 \cos(\omega_2 t) \tag{2-16}$$

Substituting (2.1) into (2.2) gives, for the first term,

$$i_a(t) = av_s(t) = aV_1 \cos(\omega_1 t) + aV_2 \cos(\omega_2 t) \tag{2-17}$$

After doing the same with the second term, the quadratic, and applying the well-known trigonometric identities for squares and products of cosines, we obtained:

$$\begin{aligned} i_b(t) = bv_s^2(t) = \frac{b}{2}\{V_1^2 + V_1^2 + V_1^2 \cos(2\omega_1 t) + V_2^2 \cos(2\omega_2 t) + \\ + 2V_1V_2[\cos((\omega_1 + \omega_2)t) + \cos((\omega_1 - \omega_2)t)] \end{aligned} \tag{2-18}$$

and the third term, the cubic, gives

$$
\begin{aligned}
i_c(t) = cv_s^3(t) = \frac{c}{4}\{&V_1^3\cos(3\omega_1 t) + V_2^3\cos(3\omega_2 t) \\
&+ 3V_1^2V_2[\cos((2\omega_1+\omega_2)t) + \cos((2\omega_1-\omega_2)t)] \\
&+ 3V_1V_2^2[\cos((\omega_1+2\omega_2)t) + \cos((\omega_1-2\omega_2)t)] \\
&+ 3(V_1^3 + 2V_1V_2^2)\cos(\omega_1 t) \\
&+ 3(V_2^3 + 2V_1^2V_2)\cos(\omega_2 t)\}
\end{aligned}
\tag{2-19}
$$

The total current in the nonlinear element is the sum of the current components in (2-17) through (2-19).

One obvious property of a nonlinear system is its generation of harmonics of the excitation frequency or frequencies. These are evident as the terms in (2-17) through (2-19) at $m\omega_1$ and $m\omega_2$. The mth harmonic of an excitation frequency is an mth-order mixing frequency. In narrow-band systems, harmonics are not a serious problem because they are far removed in frequency from the signals of interest and inevitably rejected by filters. In other systems, such as transmitters, harmonics may interfere with other communication systems and must be reduced by filters or other means.

2.4.2 Intermodulation Distortion

All the mixing frequencies in (2-17) through (2-19) that arise as linear combination of two or more tones, often called Intermodulation (IM) products. IM products were generated in an amplifier or communications receiver, often come with a serious problem. While the interfered spurious signals can be mistaken for desired signals. IM products are generally much weaker than the generating signals; however, a situation often arises wherein two or more very strong signals, which may be outside the receiver's passband, generate an IM product that is within the receiver's passband and obscures a weak and desired signal. Even-order IM products usually occur at frequencies well above or below the generating signals, and consequently are often of little concern because….. The IM products of greatest concern are usually the third-order ones that occur at $2\omega_1$-ω_2 and $2\omega_2$-ω_1, and because they are the

strongest of all odd-order products and close to the generating signals, they often cannot be rejected by filters. Thus, intermodulation is a major concern in microwave system.

2.5 Fundamental of the Volterra Series

The nonlinearity of the system often leads to interesting and important phenomena, such as harmonics, gain compression, desensitization, blocking, cross modulation, intermodulation, etc. These distortions will degrade the performance of the system, while Volterra series will be used for distortion computations. It can provide designers some information to derive which circuit parameters or circuit elements they have to modify to obtain the required specifications. Therefore, Volterra series will be introduced in the following section.

In fact, the Volterra series describe a nonlinear system in a way which is equivalent to the Taylor series approximating an analytic function. A nonlinear system excited by a signal with small amplitude can be described by the Volterra series, which can be broken down after the first few terms. The higher the input amplitude, the more terms of that series need to be taken into account to describe the system behavior properly. For very high amplitudes, the series diverges just as Taylor series. Hence, Volterra series are only suitable for the analysis of weak nonlinear circuits.

The Volterra series approach has been proven to be useful for hand calculations of small transistor networks. Since Volterra kernels retain phase information, they are especially useful for high-frequency analysis.

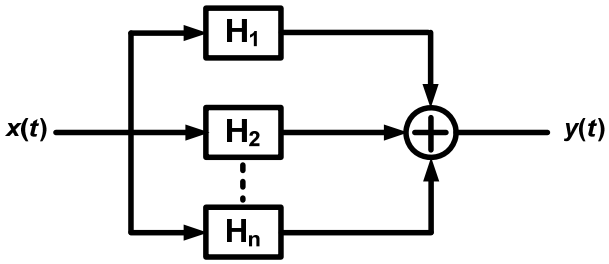


Fig. 2.5 Schematic representation of a system characterized by a Volterra series

The theory of Volterra series can be viewed as an extension of the theory of linear, first-order systems to weakly nonlinear systems. And a system is considered as the combination of different operators of different order in the Volterra series description, as shown in Fig. 2.5. Every block H_1, H_2, and H_n represents an operator of order 1, 2, …, respectively. The amount of operators must be used depend on the input amplitude. In general, the weakly nonlinear effects can be described accurately by taking into account third-order effects only.

In the time domain, the transformation on an input signal (x(t)) was performed by a nth-order Volterra operator that is given by:

$$H_n[x(t)] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h_n(\tau_1, \tau_2, \cdots, \tau_n) x(t-\tau_1) x(t-\tau_2) \cdots x(t-\tau_n) d\tau_1 d\tau_2 \cdots d\tau_n \quad (2\text{-}20)$$

The n-dimension integral can be seen as an nth-order convolution integral. The function $h_n(\tau_1, \tau_2, \cdots, \tau_n)$ is an nth-order Volterra kernel. The output of a nonlinear system can represent the sum of the output of a first-order Volterra operator with the output of a second-order one, a third-order one and so on, as shown in Fig. 2.5. The Volterra series of the nonlinear system can be expressed as

$$y(t) = H_1[x(t)] + H_2[x(t)] + H_3[x(t)] + \cdots + H_n[x(t)] \quad (2\text{-}21)$$

In the frequency domain, the nth-order Volterra kernel can be given by

$$H_n(s_1, \cdots, s_n) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h_n(\tau_1, \cdots, \tau_n) e^{-(s_1\tau_1 + \cdots + s_n\tau_n)} d\tau_1 d\tau_2 \cdots d\tau_n \quad (2\text{-}22)$$

and is called the nth-order nonlinear transfer function or the nth-order kernel transform.

2.6 Nonlinear Performance Parameters in Terms of Volterra Kernels

When a system that is described by a Volterra series up to order three, it is excited by the sum of two sinusoidal excitations $A_1 \cos \omega_1 t$ and $A_2 \cos \omega_2 t$. Then the output is given by the sum of the responses listed in Table 2.1. From Table 2.1, the expression for the second and third harmonic distortion in terms of general Volterra are given by

$$HD_2 = \frac{A_1}{2}\left|\frac{H_2(j\omega_1, j\omega_1)}{H_1(j\omega_1)}\right| \tag{2-23}$$

$$HD_3 = \frac{A_1^2}{4}\left|\frac{H_3(j\omega_1, j\omega_1, j\omega_1)}{H_1(j\omega_1)}\right| \tag{2-24}$$

Furthermore, among the intermodulation products, the third-order intermodulation products at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ is important. Since if the difference between ω_1 and ω_2 is small, the distortions at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ would appear in the vicinity of ω_1 and ω_2. Table 2.1 shows the third-order intermodulation distortion in terms of Volterra kernel transforms

$$IM_3 = \frac{3}{4}A_2^2\left|\frac{H_3(-j\omega_1, j\omega_2, j\omega_2)}{H_1(j\omega_1)}\right| \tag{2-25}$$

Table. 2.1 Different responses at the output of a nonlinear system described by Volterra kernels.

Order	Frequency of response	Amplitude of response	Type of response
1 1	ω_1 ω_2	$A_1\lvert H_1(j\omega_1)\rvert$ $A_2\lvert H_1(j\omega_2)\rvert$	Linear
2 2	$\omega_1+\omega_2$ $\lvert\omega_1-\omega_2\rvert$	$A_1A_2\lvert H_2(j\omega_1,j\omega_2)\rvert$ $A_1A_2\lvert H_2(j\omega_1,-j\omega_2)\rvert$	2nd-order intermodulation products
2 2	$2\omega_1$ $2\omega_2$	$\frac{1}{2}A_1^2\lvert H_2(j\omega_1,j\omega_1)\rvert$ $\frac{1}{2}A_2^2\lvert H_2(j\omega_2,j\omega_2)\rvert$	2nd harmonics
2 2	0 0	$\frac{1}{2}A_1^2\lvert H_2(j\omega_1,-j\omega_1)\rvert$ $\frac{1}{2}A_2^2\lvert H_2(j\omega_2,-j\omega_2)\rvert$	DC shift
3 3 3 3	$2\omega_1+\omega_2$ $\lvert 2\omega_1-\omega_2\rvert$ $\omega_1+2\omega_2$ $\lvert\omega_1-2\omega_2\rvert$	$\frac{3}{4}A_1^2A_2\lvert H_3(j\omega_1,j\omega_1,j\omega_2)\rvert$ $\frac{3}{4}A_1^2A_2\lvert H_3(j\omega_1,j\omega_1,-j\omega_2)\rvert$ $\frac{3}{4}A_1A_2^2\lvert H_3(j\omega_1,j\omega_2,j\omega_2)\rvert$ $\frac{3}{4}A_1A_2^2\lvert H_3(j\omega_1,-j\omega_2,-j\omega_2)\rvert$	Third-order intermodulation products
3 3	$\omega_1+\omega_2-\omega_2=\omega_1$ $\omega_1-\omega_1+\omega_2=\omega_2$	$\frac{3}{4}A_1A_2^2\lvert H_3(j\omega_1,j\omega_2,-j\omega_2)\rvert$ $\frac{3}{4}A_1^2A_2\lvert H_3(j\omega_1,-j\omega_1,j\omega_2)\rvert$	Third-order desensitization
3	$2\omega_1-\omega_1=\omega_1$	$\frac{3}{4}A_1^3\lvert H_3(j\omega_1,j\omega_1,-j\omega_1)\rvert$	Third-order

3	$2\omega_2 - \omega_2 = \omega_2$	$\frac{3}{4} A_2^3 \lvert H_3(j\omega_2, j\omega_2, -j\omega_2)\rvert$	compression or expansion
3 3	$3\omega_1$ $3\omega_2$	$\frac{1}{4} A_1^3 \lvert H_3(j\omega_1, j\omega_1, j\omega_1)\rvert$ $\frac{1}{4} A_2^3 \lvert H_3(j\omega_2, j\omega_2, j\omega_2)\rvert$	Third harmonics

This effect causes some distortion at our desired frequency and damages the desired signals. Therefore third intercept point (IP3) is used to characterize this behavior. This parameter is measured by supplying a two-tone signal to the system. This input signal must be chosen to be sufficiently small in order to remove higher-order nonlinear terms. In a typical test, A_1=A_2=A, hence the magnitude of third-order intermodulation products grows at three times the rate at which the fundamental signal on a logarithmic scale when input signal increases. The third-order intercept point is defined to be the point at which third-order intermodulation product equals to the fundamental signal, and the corresponding input signal is called input IP3 (IIP3) and the corresponding output signal is called output IP3 (OIP3). The A_{IP3}, therefore, can be obtained by setting $IM_3 = 1$ and expressing as

$$A_{IP3}^2 = \frac{4}{3}\left|\frac{H_1(j\omega_1)}{H_3(-j\omega_1, j\omega_2, j\omega_2)}\right| \tag{2-26}$$

Besides, a quick method of measuring IIP3 is as follows. As shown in Fig. 2.6, If the power of the two-tone signal, P_{in}, is small enough to ignore higher order nonlinear terms, then IIP3 can be expressed as

$$IIP_3\,|_{dBm} = \frac{\Delta P\,|_{dB}}{2} + P_{in}\,|_{dBm} \tag{2-27}$$

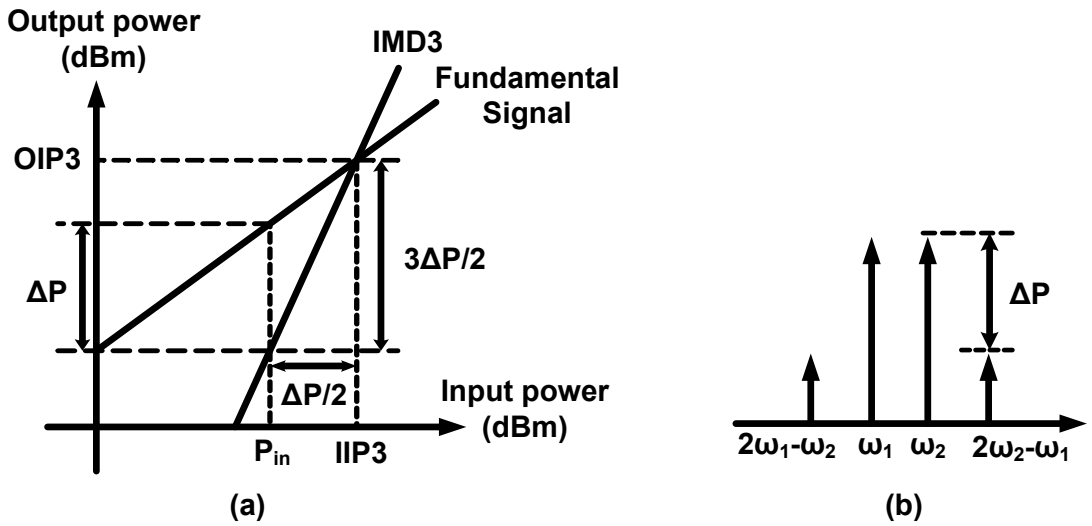


Fig. 2.6 (a) Growth of output components in an intermodulation test

(b) Intermodulation distortion

Chapter 3

General Consideration in RF Circuit Design

3.1 Low Noise Amplifier Basic

Low noise amplifier is the first gain stage in the receive path so its noise figure directly adds to that of the system. There, therefore, are several common goals in the design of LNA. These include minimizing noise figure of the amplifier, providing enough gain with sufficient linearity and providing a stable 50 Ω input impedance to terminate an unknown length of transmission line which delivers signal from antenna to the amplifier. Among LNA architectures, inductive source degeneration is the most popular method since it can achieve noise and power matching simultaneously, as shown in Fig. 3.1. The following analysis is based on this architecture.

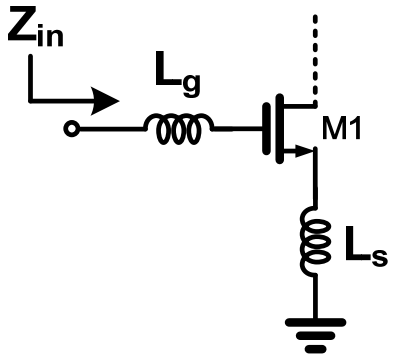


Fig. 3.1 Common source input stage with inductive source degeneration

3.1.1 Low Noise Amplifier Architecture Analysis

In Fig. 3.1, the input impedance can be expressed as

$$Z_{in} = s(L_g + L_s) + \frac{1}{sC_{gs}} + \left(\frac{g_m}{C_{gs}}\right)L_s$$
$$= \left(\frac{g_m}{C_{gs}}\right)L_s \approx \omega_T L_s \quad at \quad \omega = \omega_o = \frac{1}{\sqrt{(L_g + L_s)C_{gs}}} \tag{3.1}$$

as shown in (3.1), the input impedance is equal to the multiplication of cutoff frequency of the device and source inductance at resonant frequency. Therefore it can be set to 50 Ω for input matching while resonant frequency is designed to be equal to the operating frequency.

According to prior introduction, the equivalent noise model of common-source LNA with inductive source degeneration can be expressed as Fig. 3.2, where R_l is the parasitic resistance of the inductor, R_g is the gate resistance of the device. Note that the overlap capacitance C_{gd} has also been neglected in the interest of simplicity. Then the noise figure can be obtained by computing the total output noise power and output noise power due to input source. To find the output noise, we first evaluate the trans-conductance of the input stage. With the output current proportional to the voltage no C_{gs} and noting that the input circuit takes the form of series-resonant network, the transconductance at the resonant frequency can be expressed as

$$G_m = g_m Q_{in} = \frac{g_m}{\omega_o C_{gs}(R_s + \omega_T L_s)} = \frac{\omega_T}{2\omega_o R_s} \tag{3.2}$$

where Q_{in} is the effective Q of the amplifier input circuit. So the output noise power density due to the source can be expressed as

$$S_{a,Rs}(\omega_o) = S_{Rs} G_{m.eff}^2 = \frac{4kT\omega_T^2}{\omega_o^2 R_s (1 + \frac{\omega_T L_s}{R_s})^2} \tag{3.3}$$

In the similar way, the output noise power density due to R_g and R_l is

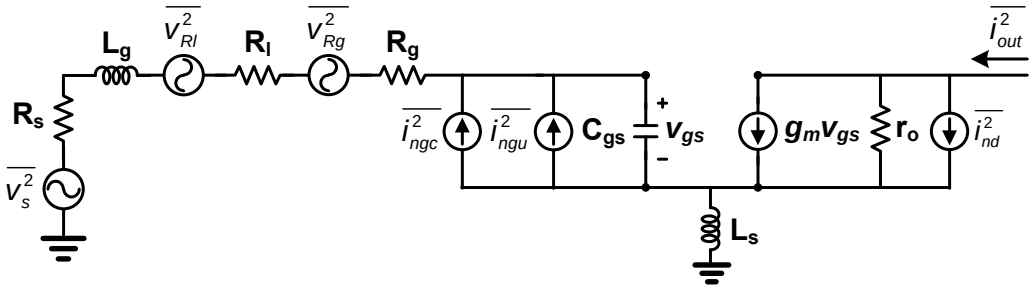


Fig. 3.2 Equivalent noise model of Figure 3.1

$$S_{a,R_g,R_l}(\omega_o) = \frac{4kT(R_g + R_l)\omega_T^2}{\omega_o^2 R_s^2 (1 + \frac{\omega_T L_s}{R_s})^2} \tag{3.4}$$

Furthermore, channel current noise of the device is the dominant noise contributor, and its noise power density associated with the correlated portion of the gate noise can be expressed as

$$S_{a,i_{nd},i_{ngc}}(\omega_o) = \frac{4kT\gamma\kappa g_{do}}{(1 + \frac{\omega_T L_s}{R_s})^2} \tag{3.5}$$

where γ is the coefficient of channel thermal noise, $\alpha = g_m / g_{d0}$ and

$$\kappa = \frac{\delta\alpha^2}{5\gamma}|c|^2 + \left[1 + |c|Q_L\sqrt{\frac{\delta\alpha 2}{5\gamma}}\right]^2 \tag{3.6}$$

$$Q_L = \frac{1}{\omega_o R_s C_{gs}} \tag{3.7}$$

The last noise term is the contribution of the uncorrelated portion of the gate noise, and its output noise power density can be expressed as

$$S_{a,i_{ngu}}(\omega_o) = \frac{4kT\gamma\xi g_{do}}{(1 + \frac{\omega_T L_s}{R_s})^2} \tag{3.8}$$

where

$$\xi = \frac{\delta\alpha^2}{5\gamma}(1 - |c|^2)(1 + Q_L^2) \tag{3.9}$$

According to (3.3), (3.4), (3.5) and (3.8), the noise figure at the resonant frequency can be expressed as

$$F = 1 + \frac{R_l}{R_s} + \frac{R_g}{R_s} + \frac{\gamma\chi}{\alpha Q_L}\left(\frac{\omega_o}{\omega_T}\right) \tag{3.10}$$

where

$$\chi = 1 + 2|c|Q_L\sqrt{\frac{\delta\alpha^2}{5\gamma}} + \frac{\delta\alpha^2}{5\gamma}(1+Q_L^2) \tag{3.11}$$

From (3.11), we observe that χ includes the terms which are constant, proportional to Q_L, and proportional to Q_L^2. It follows that (3.11) will contain terms which are proportional to Q_L as well as inversely proportional to Q_L. A minimum noise figure, therefore, exits for a particular Q_L.

3.1.2 Optimizations of Low Noise Amplifier Design Flow

The analysis of the previous section can now be drawn upon in designing the LNA. In order to pick the appropriate device size and bias point to optimize noise performance given specific objectives for gain and power dissipation, a simple second-order model of the MOSFET transconductance can be employed which accounts for high-field effects in short channel devices. Assume that the drain current, I_d, has the form

$$I_{DS} = WC_{ox}v_{sat}\frac{(V_{gs}-V_T)}{1+\frac{LE_{sat}}{V_{gs}-V_T}} = WC_{ox}v_{sat}LE_{sat}\frac{\rho^2}{1+\rho} \tag{3.12}$$

where $\rho \equiv \frac{V_{gs}-V_T}{LE_{sat}}$. And the (3-7) can be replace as

$$Q_L = \frac{3}{2\omega_o WLC_{ox}R_s} \Rightarrow C_{ox} = \frac{3}{2R_sQ_L\omega_o WL} \tag{3.13}$$

The power consumption of the LNA, therefore, can be expressed as

$$P_D = V_{DD} I_{DS} = \frac{3}{2} V_{DD} \frac{1}{Q_L R_s \omega_o} v_{sat} E_{sat} \frac{\rho^2}{1+\rho} \tag{3.14}$$

The noise figure can be expressed in terms of P_D and V_{gs}. Two parameters linked to power dissipation need to be accounted for.

$$\omega_T \approx \frac{g_m}{C_{gs}} = f_1(V_{gs}) \tag{3.15}$$

$$Q_L = \frac{3V_{DD} v_{sat} E_{sat}}{2P_D \omega_o R_s} \frac{\rho^2}{1+\rho} = \frac{P_o}{P_D} \frac{\rho^2}{1+\rho} = f_2(V_{gs}, P_D) \tag{3.16}$$

where $P_o = \frac{3V_{DD} v_{sat} E_{sat}}{2\omega_o R_s}$.

The noise figure of the LNA, therefore, can be expressed as

$$F = 1 + \frac{R_l}{R_s} + \frac{R_g}{R_s} + \frac{\gamma}{\alpha Q_L} \left(\frac{\omega_o}{\omega_T} \right) \left(1 + 2|c| Q_L \sqrt{\frac{\delta\alpha^2}{5\gamma}} + \frac{\delta\alpha^2}{5\gamma} (1 + Q_L^2) \right) = f(V_{gs}, P_D) \tag{3.17}$$

In general, there are two approaches to optimize noise figure. The first approach assumes a fixed transconductance, G_m. The second approach assumes fixed power consumption.

(1) Fixed G_m optimization: To fix the value of the transconductance, G_m, we need only assign a constant value to ρ. Once ρ is determined, the optimization of the noise figure can be obtained by (3.17):

$$\left. \frac{\partial f(V_{gs}, P_D)}{\partial P_D} \right|_{fixed\ Vgs} = 0 \Rightarrow P_{D.opt} \Rightarrow Q_{L.opt} \Rightarrow F = f(V_{gs}, P_{D.opt}) \tag{3.18}$$

From (3.18), we can obtain the optimal width to get the minimal noise figure for a given G_m under the assumption of matched input impedance. In this approach, the designer can achieve high gain and low noise performance by selecting the desired transconductance, but its disadvantage is that we must sacrifice the power consumption to achieve minimum noise figure.

(2) Fixed P_D optimization: An alternative method of optimization fixes the power dissipation and adjusts device size and bias point to minimize the noise figure. Once P_D is determined, the optimization of the noise figure can be obtained by (3.19):

$$\left.\frac{\partial f(V_{gs}, P_D)}{\partial V_{gs}}\right|_{fixed\ P_D} = 0 \Rightarrow V_{gs.opt} \Rightarrow Q_{L.opt} \Rightarrow F = f(V_{gs.opt}, P_D) \tag{3.19}$$

Then the optimum device size can be obtained to get the best noise performance for fixed power dissipation. In this approach, the designer can specify the power dissipation and find the optimal noise performance, but its disadvantage is that the transconductance is held up by the optimal noise condition.

3.1.3 Amplifier stability

The stability of an amplifier, or its resistance to oscillate, is a very important consideration in a design and can be determined from the S parameters, the matching networks, and the terminations. The non-zero S_{12} parameter of a two port networks as shown in Fig. 3.3 provides a feedback path by which the power transferred to the output can be feedback to the input and combined together. Oscillation may occur when the magnitude of reflection coefficient Γ_{IN} or Γ_{OUT}, defined as the ratio of the reflected to the incident wave, exceeds unity. It is expected that a properly designed amplifier will not oscillate no matter what passive source and load impedances are connected to it, which is said to be unconditionally stable and the reflection coefficient is given as

$$\begin{aligned} |\Gamma_{IN}| &= |S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}| < 1 \\ |\Gamma_{OUT}| &= |S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s}| < 1 \end{aligned} \tag{3.20}$$

A network that has $\Gamma_{IN} > 1$ or $\Gamma_{OUT} > 1$ for certain load impedance is said to be

conditionally stable. In such a case, input and load stability circles, the contour of $\Gamma_{IN}=1$ and $\Gamma_{OUT}=1$ for certain frequencies on the Smith chart, are useful to fine the boundary line for load and source impedances that cause stable and unstable condition. The stability circles can be calculated directly from the S parameters of the two port network, so another convenient parameter, stability factor K, is defined and given as

$$K=\frac{1-|S_{11}|^2-|S_{22}|^2+|\Delta|^2}{2|S_{12}S_{21}|} \tag{3.21}$$

where

$$|\Delta|^2=|S_{11}S_{22}-S_{12}S_{21}|^2$$

The amplifier is unconditionally stable provided that

$$K>1 \quad \text{and} \quad |\Delta|^2<1 \tag{3.22}$$

or equivalently

$$K>1 \quad \text{and} \quad B_1<1 \tag{3.23}$$

where $B1=1+|S_{11}|^2-|S_{22}|^2-|\Delta|^2$.

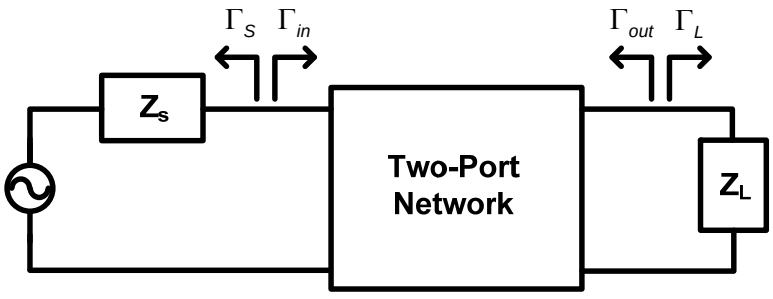


Fig. 3.3 Stability of two-port networks embedded between source and load.

3.2 Down Conversion Mixer Basic

The purpose of the mixer is to convert a signal from one frequency to another. In a receiver, this conversion is from radio frequency to intermediate frequency or zero-IF. Mixing requires a circuit with a nonlinear transfer function, since nonlinearity is fundamentally necessary to generate new frequencies. Fig. 3.4 shows a simplified CMOS Gilbert Cell mixer, which is composed of transcondcutance stage and switching stage.

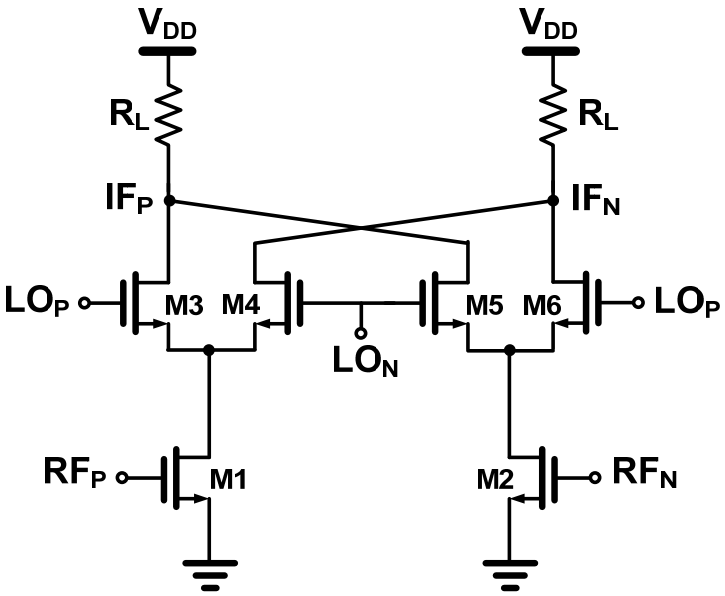


Fig. 3.4 The simplified CMOS Gilbert Cell mixer

The RF input must be linear, or adjacent channels could intermodulate and interfere with the desired channel. And the third-order intermodulation term from the two other signals will be directly on top of the desired signal. The LO input need not be linear, since the LO is clean and of known amplitude. In fact, the LO input is usually designed to switch the upper quad so that for half the cycle M3 and M6 are on and taking all current to output loading. For the other half of the LO cycle, M3 and M6 are off and M4 and M5 are on. This stage will be, therefore, like switch to mixing RF signal to IF signal.

3.2.1 Conversion Gain

The gain of mixers must be carefully defined to avoid confusion. The voltage conversion gain of a mixer is defined as the ratio of the rms voltage of the IF signal and rms voltage of the RF signal. Note that the frequencies of these two signals are different. The power conversion gain of a mixer is defined as the IF power delivered to the load divided by the available RF power from the source. If the impedances are both matched to 50 Ω, then the voltage conversion gain and power conversion gain of the mixer are equal when they are expressed in decibels.

Now, we assume that M3-M6 work like an ideal switch, and the conversion transconductance of the mixer can be expressed as

$$G_c = \frac{2}{\pi} g_m \tag{3.24}$$

where gm is the transcondcutanc of M1 and M2, and 2/π is produced by switching stage.

3.2.2 Switching Stage

For small LO amplitude, the amplitude of the output depends on the amplitude of the LO signal. Thus, gain is larger for larger LO amplitude. For Large LO signals, the upper quad switches and no further increase occur. Thus, at this point, there is no longer any sensitivity to LO amplitude. Besides, if upper quad transistors are alternately switched between completely off and fully on, the noise will be minimized. Since upper transistor contributes on noise when it is fully off, and when fully on, the upper transistor behaves like a cascode transistor which does not contribute significantly to noise.

The large LO signal is required to let upper quad transistors achieve complete switching. But if the LO voltage is made too large, a lot of current has to be moved into and out of the transistors during transitions. This can lead to spikes in the signals

and can actually reduce the switching speed and cause an increase in LO feed-through. Thus, too large a signal can be just as bad as too small a signal.

3.2.3 Mixer Noise

Noise figure for a mixer is defined as

$$\text{Noise Factor} = \frac{\text{Total output noise power at the IF}}{\text{Output noise power at IF due to input source}} \tag{3.25}$$

In general, the noise figure of the mixer is divided to two categories, single-sideband (SSB) noise figure and double-sideband (DSB) noise figure. The difference between the two definitions is the value of the denominator in (3.25). In the case of SSB noise figure, only the noise at the output frequency due to the source that originated at the RF frequency is considered, and it is usually used in heterodyne systems. In the case of DSD noise figure, all the noise at the output frequency due to the source is considered (noise of the source at the input and image frequencies), and it is usually used in homodyne systems.

Because of the added complexity and the presence of noise that is frequency translated, mixers tend to be much noisier than LNAs. In generally, mixers have three frequency bands where noise is important:

(1) Noise already present at the IF: The transistors and resistors in the circuit will generate noise at the IF. Some of this noise will make it to the output and corrupt the signal.

(2) Noise at the RF and image frequency: The noise presents at the RF and image frequency will be mixed down to the IF.

(3) Noise at multiples of the LO frequencies: Any noise that is near a multiple of the LO frequency can also be mixed down to the IF, just like the noise at the RF.

Beside, the flicker noise will become more important in the homodyne receiver. In the design of the direct down-conversion mixer, how to reduce the flicker noise of upper

quad transistors is the important thing. This noise can be reduced by increasing the device size for a given g_m.

3.2.4 Port-to-Port Isolation

The isolation between each two ports of a mixer is critical. The LO-RF feed-through results in LO leakage to the LNA and eventually the antenna, whereas the RF-LO feed-through allows strong interferers in the RF path to interact with the local oscillator driving the mixer. The LO-IF feed-through is important because if substantial LO signal exists at the IF output even after low-pass filtering, then the following stage may be desensitized. Fortunately, this feed-through can be reduced largely by used the double-balanced architecture. Finally, the RF-IF isolation determines what fraction of the signal in the RF path directly appears in the IF, a critical issue with respect to the even-order distortion problem in homodyne receivers.

The required isolation levels greatly depend on the environment in which the mixer is employed. If the isolation provided by the mixer is inadequate, the preceding or following circuits may be modified to remedy the problem.

3.2.5 Linearity

As far as cascaded stages are concerned, linearity is a very important performance in mixer stage. In general, it will dominate the distortion of the entire receiver. The detail analysis will be introduced in Chapter 4.

Chapter 4

-5dBm IIP3 UWB Low Noise Amplifier Using Complex Derivative Cancellation Technique

4.1 Introduction

Linearity plays an important role in broadband RF systems because nonlinearity degrades system performance with the consequence effects such as harmonic generation, cross-modulation and intermodulation. Of these distortions, the third-order intermodulation (IMD3) is one of the critical terms to be solved. The issue is especially serious to the low noise amplifier (LNA) in broadband system as an LNA needs high gain to suppress noise, while facing a broadband spectrum involving external interferers without much filtering. As such, it has been great attention on linearity improvement to RF circuit designers.

In recent years, several techniques have been proposed to improve the linearity of RF amplifier. Most of them are based on negative feedback circuit. One of the most famous ones is using source degeneration by resistor or inductor. Another scheme is the superposition of auxiliary transistors operated in different bias conditions to cancel the derivative of device transconductance. Such method, referred as derivative superposition or multiple gated transistors (MGTR), offers a good opportunity to extend linearity without increasing power consumption. However the conventional derivative cancellation through DC transconductance analysis is inaccurate at RF frequency. A complex transconductance analysis technique was proposed to search for the optimal design parameters. It is also found adequate to improving IIP3 in a broadband LNA design.

In this work, a 3.1~10.6GHz CMOS LNA employing the complex transconductance analysis is reported, with linearity improvement of 5 dB. In the next section the IIP3 analysis of a cascode amplifier is provided, followed by the LNA circuit design, fabrication and measurement results in 0.18 μm CMOS technology.

4.2 UWB LNA Design Consideration

A low noise amplifier (LNA) is an important component at receiver path for wireless communication. It provides a high gain with low noise figure for overall wireless communication system. For various LNA circuit topologies, the common source amplifier is generally popular as it provides a better noise performance with low power consumption. It is especially popular for extreme applications in which ultra low power. A widely used one for narrowband LNA design is a CS amplifier with inductive source degeneration, which has been well analyzed. Because the frequency dependency of the derived Z_{opt} is different from that of Z_{in}, broadband LNA is not feasible using that technique. This is observed in the broadband amplifier realized by employing a multi-order LC matching network. The noise performance is still band-limited. We designed the UWB LNA by employing dual reactive feedback topology, and the theory will be detailed in this paper.

In this chapter discussions are given for broadband noise and input matching realization in a CMOS LNA. Starting from the next section, we first analyze the dual reactive feedback circuit is proposed to achieve broadband noise and input matching, we use transformer simplifying the LNA circuit to save chip area. For better linearity, we try to use multiple gate transistors to improve the linearity of UWB LNA. The following section shows a design example of an UWB LNA implemented in TSMC 0.18μm CMOS process, along with simulation and measurement results.

4.3 Dual Reactive Feedback Technique For Broadband Input Impedance Matching

For this UWB LNA, the proposed solution is a dual reactive feedback topology composing of a capacitive shunt feedback and an inductive series feedback, which individually attain noise and input matching in two different frequency regions to constitute the broadband noise and input matching. These two feedbacks are seamlessly combined by employing an inductor at transistor drain, which conducts different loading conditions for each feedback structure. Then for some reason, three inductors in this circuitry are merged into a transformer to reduce the chip area.

4.3.1 The proposed dual reactive feedback circuit

The proposed dual reactive feedback structure with input matching network is shown in Fig. 4.1. The Z_{opt}^{*} has an equivalent circuit representation shown in Fig. 4.2, as Z_{opt}^{*} is not affected by lossless feedbacks. By employing the ladder filter structure

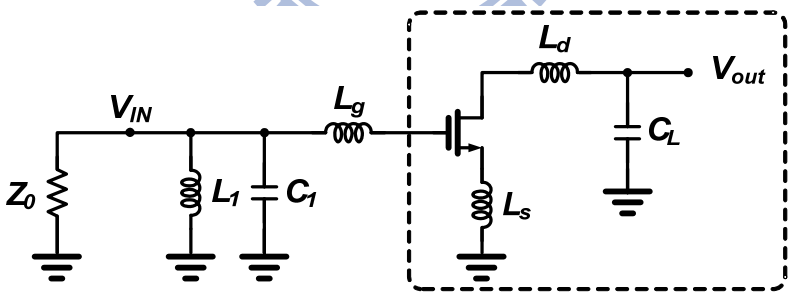


Fig. 4.1 Dual reactive feedback structure

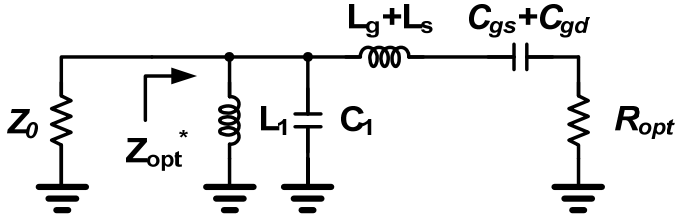


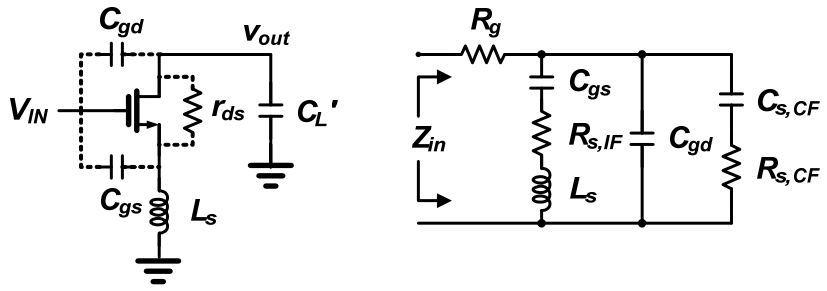
Fig. 4.2 Equivalent circuit of Z_{opt}^{*}.

with L_1 and C_1, the Z_{opt}^{*} to Z_0 matching bandwidth is extended. Following the general filter design guideline,

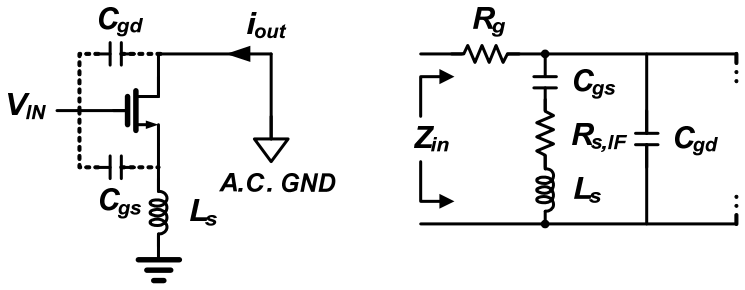
$$L_1 \cdot C_1 = (L_g + L_s) \cdot (C_{gs} + C_{gd}) = 1/(2\pi f_c)^2 , \tag{4.1}$$

in which f_c is the center frequency of pass band. To the optimal wideband matching result the R_{opt} can be designed slightly less than Z_0, not strictly following the standard filter design algorithm. The design of L_1 and C_1 also takes into account the gain response as described the in next subsection.

After the broadband noise matching is preliminarily achieved, the input impedance is then matched by the proposed dual reactive feedback circuitry, as in the dashed-line box of Fig. 4.1. This circuitry provides different reactive feedbacks in different frequency regions. In the frequency region lower than the L_d-C_L series resonance frequency, when looked from the transistor drain, the L_d-C_L tank behaves



(a)



(b)

Fig. 4.3 Input impedance changed among the two feedbacks with frequency

(a) capacitive shunt feedback in lower frequency region;

(b) inductive series feedback in higher frequency region.

like a capacitor C_L' as shown in the left of Fig. 4.3(a). Hence the input impedance can be represented as an equivalent circuit in the right of Fig. 4.3(a), in which

$$R_{s,IF} = \frac{g_m \cdot L_s}{C_{gs}}, \tag{4.2}$$

$$R_{s,CF} = \frac{C_L'+C_{gd}}{g_m C_{gd}} \approx \frac{C_L + C_{gd}}{g_m C_{gd}}, \tag{4.3}$$

and

$$C_{s,CF} = g_m r_{ds} \cdot C_{gd}. \tag{4.4}$$

Two noiseless resistances can be found in this circuit: the $R_{s,IF}$ by the series inductive feedback of L_s, and the $R_{s,CF}$ by the shunt capacitive feedback of C_{gd} with C_L'. Because the $C_{s,CF}$, as in (4.4), is much larger than C_{gs}, the branch of $C_{s,CF}$ dominates the input impedance in the lower frequency region. As such the $R_{s,CF}$ is the noiseless resistance added on R_g to match to R_{opt} in this region.

In the higher frequency region where the effect of C_{gs} is significant, the L_d is designed to resonate with C_L to provide a short circuit at the transistor drain, as shown in Fig. 4.3(b). Hence the branch of $C_{s,CF}$ in the right of Fig. 4.3(b) vanishes and the $R_{s,IF}$ takes over the role of $R_{s,CF}$. Such handover of reactive feedbacks can well minimize the R_g-to-R_{opt} difference in wide frequency range.

The determination of these relative frequency parameters includes the L_1, C_1, and L_g in Fig. 4.1. For the capacitive feedback, the L_g in Fig. 4.1 and the $C_{s,CF}$ and R_g+$R_{s,CF}$ in Fig. 4.3(a) constitute a low-Q resonance tank with resonance frequency

$$f_{0,CF} = 1/2\pi\sqrt{L_g C_{s,CF}} = 1/2\pi\sqrt{L_g \cdot g_m r_{ds} \cdot C_{gd}}. \tag{4.5}$$

With its low-Q character, the slight frequency shift of $f_{0,CF}$ by L_1 and C_1 is ignorable. For the inductive feedback, the resonance frequency is located at

$$f_{0,IF} = 1/2\pi\sqrt{(L_g + L_s)\cdot(C_{gs} \,||\, C_1)}. \tag{4.6}$$

As mentioned before, at this frequency the L_d is expected to tune out C_L. Therefore

$$L_d \cdot C_L = 1/(2\pi f_{0,IF})^2 \ . \tag{4.7}$$

The frequency relationship of $f_{0,CF} < f_c < f_{0,IF}$ is obtained from (4.1), (4.5)–(4.6).

4.3.2 Gain response

The dual reactive feedback amplifier is expected to have a broadband response to suppress noise by the succeeding stages. The gain response is mainly shaped by the drain network in Fig. 4.4. as the input network is a broadband structure. Based on the low-pass response by C_L (the gray curve in Fig. 4.4.), the L_d conducts a series peaking providing gain expansion at frequency

$$f_{0,Peak} = 1/2\pi\sqrt{L_d \cdot (C_{gd} \parallel C_L)} \ , \tag{4.8}$$

which is higher than $f_{0,IF}$. The voltage gain at this frequency can be derived with the circuit approximation, in which

$$R_d' \approx \frac{C_{gs} + C_{gd}}{g_m C_{gd}} \ . \tag{4.9}$$

As to the gain in the lower out-of-band frequency, it is suppressed by the input shunt inductor L_1. As a result it produces a gain peak at the lower band edge. The magnitude of this peak is designed close to that of gain peak at the higher band edge such that the expected gain response is as the solid curve in Fig. 4.4. Following this design guideline the voltage gain is mainly determined by that at $f_{0,Peak}$.

If a very wide bandwidth needs to be covered, an additional gain expansion by the next stage is necessary to compensate the mid-band depression. Such a case applies to a 3–11GHz UWB LNA. A flat gain response is obtainable by this stage itself as the two gain peaks can be designed fairly close to each other.

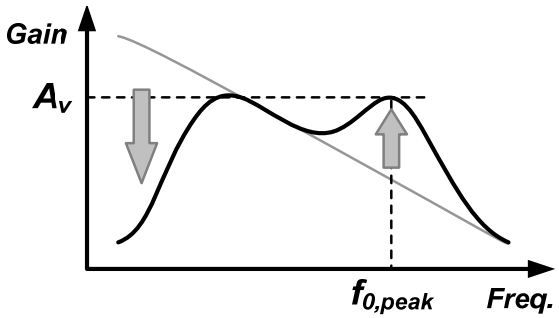
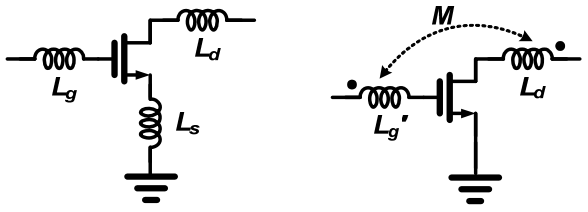


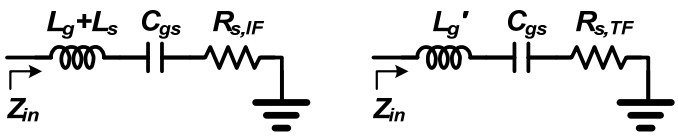
Fig. 4.4 Gain response design of the dual reactive feedback stage.

4.3.3 Dual reactive feedback with Transformer feedback

The proposed dual reactive feedback amplifier as shown in Fig. 4.1 demands four inductors to implement, which occupy huge die area. To reduce the inductor number, the transformer feedback topology is proposed to replace the three inductors L_g, L_d, and L_s, as shown in Fig. 4.5(a). In transformer feedback the L_g' and L_d are overlap sharing the same die area and having a mutual inductance M to constitute a transformer. The mutual inductance M senses the drain current and contributes a series voltage feedback at input, which provides the same function as L_s in the inductive source degeneration amplifier. The input impedance can be expressed in equivalent circuit as shown in Fig. 4.5(b), in which $L_g' = L_g + L_s$ and



(a)



(b)

Fig. 4.5 (a) The inductive source degeneration feedback can be substituted by the transformer feedback. (b) Their equivalent circuit for input impedance.

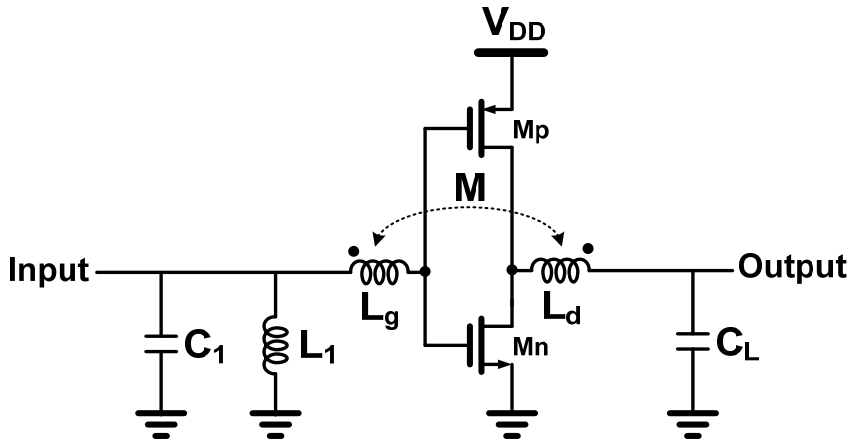


Fig. 4.6 The dual reactive feedback amplifier with transformer feedback.

$$R_{s,TF} = \frac{g_m \cdot M}{C_{gs}}, \tag{4.10}$$

which provides the wanted noiseless resistance.

The transformer feedback topology is also advantageous that the transistor source is connected to ground directly. This allows this amplifier be implemented in CMOS inverter structure, which employs both NMOS and PMOS to reuse drain current for a larger transconductance. The final shape of the proposed dual reactive feedback amplifier is therefore shown in Fig. 4.6. Here the mutual inductance is represented in coupling factor k with the relation $M = k\sqrt{L_g L_d}$, and the L_g in Fig. 4.6 represents the L_g' in Fig. 4.5.

Because the M in general is much smaller than L_g and L_d, the coupling factor k is much smaller than 1. Design experience shows the k value between 0.1 and 0.2. Possible layout schemes of such weak coupling transformer include common-centroid coil and overlapping coil, Generally the common-centroid coil has a better quality factor but consumes more die area. In the design example, the common-centroid coil was adopted for a better result.

4.4 A 3-11GHz Ultra-Wideband Low Noise Amplifier

A design example of a 3–11GHz UWB LNA employing the dual reactive feedback is demonstrated, as shown in Fig. 4.7. The inverter amplifier is self biased with a 8kΩ feedback resistor. The C_{gs} of M_T plays the role of C_L. The L_3 between M_T and M_1 further increases the gain expansion at the higher band edge. The body of M_1 is biased to its source with another 8kΩ resistor. The L_2 and R_1 provide a voltage gain with a low-Q peak at the center frequency to compensate the mid-band gain depression by the first stage. The M_2 and M_3 constitute the output buffer and the 0.1nH output inductor improves the output matching to 50Ω. For broadband input matching, the L_1 and C_1 increase the order of matching network hence both the S_{11} and S_{opt} around the center of Smith Chart and achieve good matching.

The second stage of this LNA plays the role of gain response trimming to have a flat in-band response. The L_3 further expands the gain response at the higher band edge with series peaking, and the L_2 and R_1 provide a low-Q shunt peaking at mid-band, as shown in Fig. 4.8.

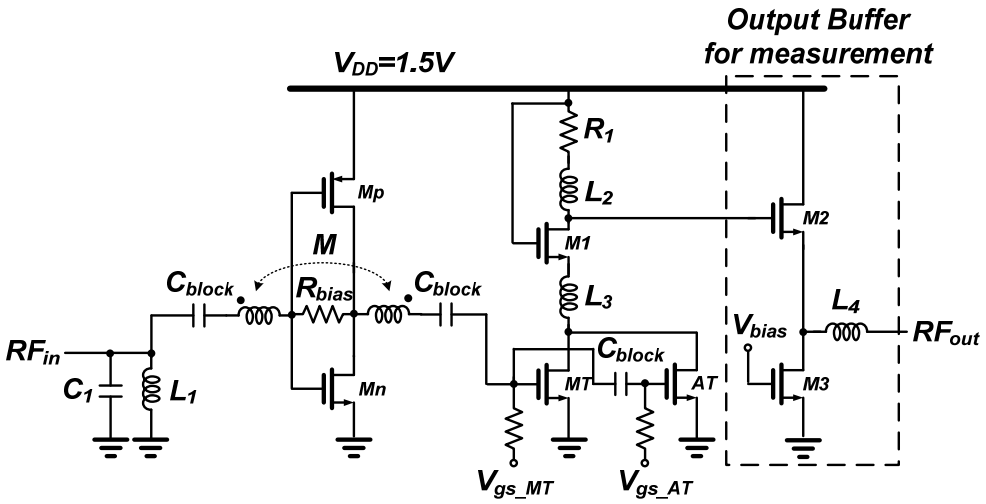


Fig. 4.7 A 3–11GHz UWB LNA as a design example of dual reactive feedback

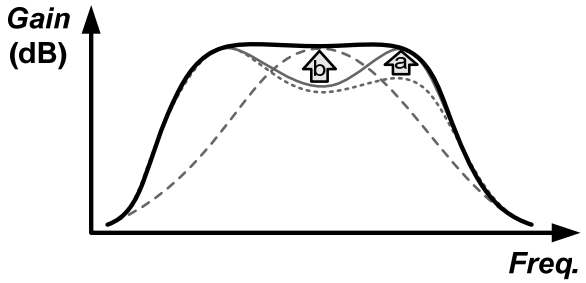


Fig. 4.8 Gain response

This LNA was designed in TSMC 0.18um CMOS process with aluminum as metal material. In simulation it was designed to target >10dB return loss, < 5dB maximum in-band noise figure, >15dB power gain with less than 1dB in-band variation, while draws <10mW DC power from a 1.5V supply.

4.5 Analysis of Linearity

In UWB RF transceiver design, linearity requirement becomes more and more challenging. Circuit nonlinearity results in various system distortions associated with the even and odd order nonlinearities. Of these distortions, the third-order intermodulation is one of the most critical terms responsible for linearity degradation in general RF systems. Due to the fact, how to improve the linearity of RF circuits without

extra power consumption becomes an important topic to be studied.

As far as cascade amplifier stages are concerned, system design calls for high linearity in the LNA to alleviate the distortion issues. Linearity is most limited by the transconductance amplifier, hence we will discuss the nonlinear effect of the common source amplifier in section 4.5.1.

In recent years, several techniques have been proposed to improve the linearity of RF circuits by linearization of the nonlinear transconductance, such as degeneration feedback. Another scheme is the superposition of auxiliary transistors operated in different bias conditions to null the derivative of device transconductance. Combined with the technique of out-of-band impedance termination, circuit linearity can be further enhanced, as indicated by the Volterra series analysis. The scheme, named as derivative superposition or multiple gated transistors (MGTR), offers a good opportunity to extend linearity without increasing power consumption. Further, we try to improve the linearity of broadband LNA using MGTR.

It is proposed that a complex transconductance shall be employed to search for the optimal design parameters for MGTR design consideration. Therefore we propose a compact equivalent circuit for the design of the multiple gated transistors technique in section 4.5.2. In section 4.5.3, we improve UWB LNA using multiple gated transistor technique by complex transconductance analysis.

4.5.1 Nonlinear Effects of Common Source Amplifier

For the common-source amplifier, the nonlinear effect is dominated by the transconductance and the output conductance of the device. The nonlinear effects of the capacitances and substrate can be neglected, and they can be considered as linear elements. Therefore we only consider the transconductance and the output

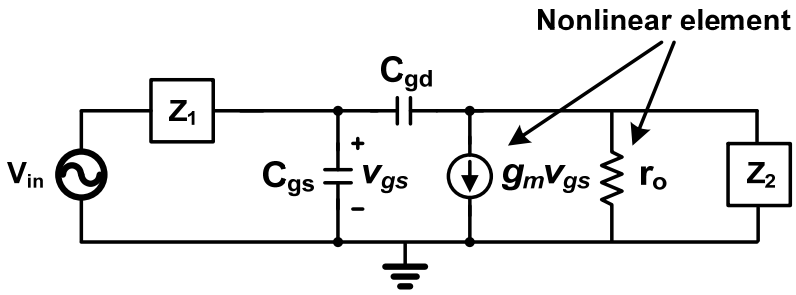


Fig. 4.9 The equivalent circuit of the common-source amplifier

conductance as the nonlinear source for the following analysis.

Fig. 4.9 shows a common-source amplifier, where Z_1 is the input impedance, and Z_2 is the output impedance. From the above-mentioned introduction and assuming that the common-source amplifier works in the weakly nonlinear region, the equivalent circuit of the common-source amplifier can be shown as Fig. 4.9, where the transconductance (g_m) and the output conductance (r_o) are the nonlinear elements. Therefore the I-V curve of the device can be expressed as

$$\begin{aligned} i_{DS} &= I_{dc} + g_m.v_{gs} + \frac{g_m^{'}}{2!}.v_{gs}^2 + \frac{g_m^{''}}{3!}.v_{gs}^3 + ... \\ &+ g_o.v_{ds} + \frac{g_o^{'}}{2!}.v_{ds}^2 + \frac{g_o^{''}}{3!}.v_{ds}^3 + ... \end{aligned} \tag{4-11}$$

where $g_m^{(n)}$ and $g_o^{(n)}$ represents the n[th]-order derivatives of the transconductance and output conductance, and $g_o = 1/r_o$.

The nonlinear distortion can be obtained by calculating the Volterra kernels of order one, two and three of voltages.

First-order kernels

In order to obtain the first-order Volterra kernels, the nonlinear elements must be replaced with its linearized equivalent, as shown in Fig. 4.10. Applying Kirchoff's current law in Fig. 4.10 yields:

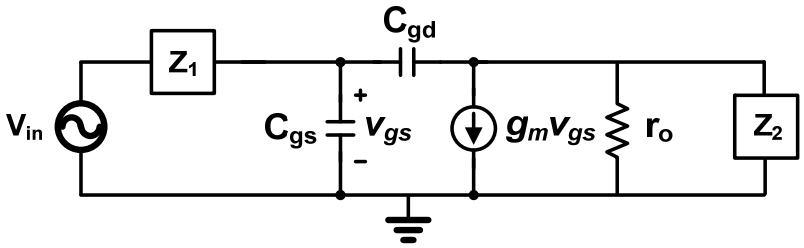


Fig. 4.10 Linearized equivalent of the circuit

$$\begin{bmatrix} (\frac{1}{Z_1(s)} + sC_{gs} + sC_{gd}) & -sC_{gd} \\ (g_m - sC_{gd}) & sC_{gd} + \frac{1}{Z_2(s) // r_o} \end{bmatrix} \cdot \begin{bmatrix} H_{11}(s) \\ H_{12}(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{Z_1(s)} \\ 0 \end{bmatrix} \tag{4-12}$$

where $H_{11}(s)$ and $H_{12}(s)$ are the Volterra kernels. The first subscript in these two transfer functions indicates the order of the transfer function, whereas the second subscript corresponds to the numbering of the node voltages. Then the first-order kernels $H_{11}(s)$ and $H_{12}(s)$ can be expressed as

$$H_{11}(s) = \frac{1 + sC_{gd}.Z_3(s)}{(g_m + g(s)).Z_x(s)} \tag{4-13}$$

$$H_{12}(s) = \frac{sC_{gd}.Z_3(s) - g_m.Z_3(s)}{(g_m + g(s)).Z_x(s)} \tag{4-14}$$

where

$$Z_3(s) = Z_2(s) // r_o \tag{4-15}$$

$$Z_x(s) = sC_{gd}.Z_1(s).Z_3(s) \tag{4-16}$$

$$g(s) = \frac{1 + sC_{gd}.(Z_1(s) + Z_3(s)) + sC_{gs}.(Z_1(s) + Z_x(s))}{Z_x(s)} \tag{4-17}$$

Second-order kernels

For computing second-order kernels, the input signal v_{in} is replaced by a short circuit, and the second-order nonlinear current sources are applied to the linearized circuit, as shown in Fig 4.11. Applying Kirchoff's current law in Fig. 4.11 yields:

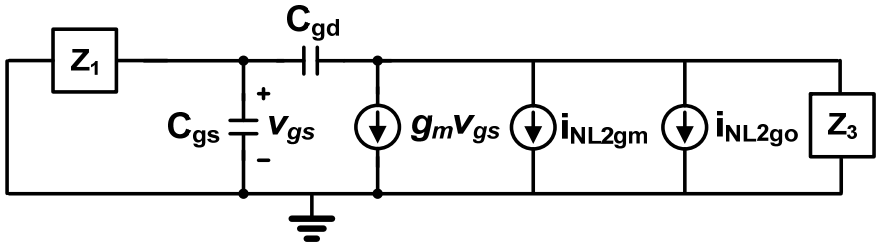


Fig. 4.11 The equivalent circuit for the computation of the second-order kernels

$$\begin{bmatrix} (\frac{1}{Z_1(s')} + s'C_{gs} + s'C_{gd}) & -s'C_{gd} \\ (g_m - s'C_{gd}) & s'C_{gd} + \frac{1}{Z_3(s')} \end{bmatrix} \cdot \begin{bmatrix} H_{21}(s_1, s_2) \\ H_{22}(s_1, s_2) \end{bmatrix} = \begin{bmatrix} 0 \\ -(i_{NL2gm} + i_{NL2go}) \end{bmatrix} \quad (4\text{-}18)$$

where $H_{21}(s_1, s_2)$ and $H_{22}(s_1, s_2)$ are the second-order kernels, and

$$s' = s_1 + s_2 \quad (4\text{-}19)$$

$$i_{NL2gm}(s_1, s_2) = \frac{g_m'}{2!} H_{11}(s_1).H_{11}(s_2) \quad (4\text{-}20)$$

$$i_{NL2go}(s_1, s_2) = \frac{g_o'}{2!} H_{12}(s_1).H_{12}(s_2) \quad (4\text{-}21)$$

Then the second-order kernels $H_{21}(s_1, s_2)$ and $H_{22}(s_1, s_2)$ can be expressed as

$$H_{21}(s_1, s_2) = \frac{-Z_1(s').Z_3(s').s'C_{gd}.(i_{NL2gm} + i_{NL2go})}{(g_m + g(s')).Z_x(s')} \quad (4\text{-}22)$$

$$H_{22}(s_1, s_2) = \frac{-Z_1(s').Z_3(s').(\frac{1}{Z_1(s')} + s'C_{gs} + s'C_{gd}).(i_{NL2gm} + i_{NL2go})}{(g_m + g(s')).Z_x(s')} \quad (4\text{-}23)$$

Third-order kernels

Just as second-order kernels, the third-order ones are computed as the response to the third-order nonlinear current sources, as shown in Fig. 4.12. Applying Kirchoff's current law in Fig. 4.12 yields:

$$\begin{bmatrix} (\frac{1}{Z_1(s'')} + s''C_{gs} + s''C_{gd}) & -s''C_{gd} \\ (g_m - s''C_{gd}) & s''C_{gd} + \frac{1}{Z_3(s'')} \end{bmatrix} \cdot \begin{bmatrix} H_{31}(s_1, s_2, s_3) \\ H_{32}(s_1, s_2, s_3) \end{bmatrix} = \begin{bmatrix} 0 \\ -(i_{NL3gm} + i_{NL3go}) \end{bmatrix} \quad (4\text{-}24)$$

where $H_{31}(s_1, s_2, s_3)$ and $H_{32}(s_1, s_2, s_3)$ are the third-order kernels, and

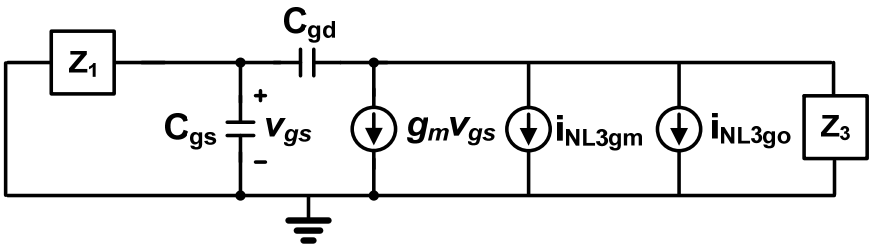


Fig. 4.12 The equivalent circuit for the computation of the third-order kernels

$$s'' = s_1 + s_2 + s_3 \tag{4-25}$$

$$i_{NL3gm}(s_1,s_2,s_3) = \frac{g_m''}{3!}.H_{11}(s_1).H_{11}(s_2).H_{11}(s_3) + \frac{g_m'}{3}[H_{11}(s_1).H_{21}(s_2,s_3) + H_{11}(s_2).H_{21}(s_1,s_3) + H_{11}(s_3).H_{21}(s_1,s_2)] \tag{4-26}$$

$$i_{NL3go}(s_1,s_2,s_3) = \frac{g_o''}{3!}.H_{12}(s_1).H_{12}(s_2).H_{12}(s_3) + \frac{g_o'}{3}[H_{12}(s_1).H_{22}(s_2,s_3) + H_{12}(s_2).H_{22}(s_1,s_3) + H_{12}(s_3).H_{22}(s_1,s_2)] \tag{4-27}$$

Then the third-order kernels $H_{31}(s_1,s_2,s_3)$ and $H_{32}(s_1,s_2,s_3)$ can be expressed as

$$H_{31}(s_1,s_2,s_3) = \frac{-Z_1(s'').Z_3(s'').s''C_{gd}.(i_{NL3gm}+i_{NL3go})}{(g_m+g(s'')).Z_x(s'')} \tag{4-28}$$

$$H_{32}(s_1,s_2,s_3) = \frac{-Z_1(s'').Z_3(s'').(\frac{1}{Z_1(s'')}+s''C_{gs}+s''C_{gd}).(i_{NL3gm}+i_{NL3go})}{(g_m+g(s'')).Z_x(s'')} \tag{4-29}$$

Of the nonlinear distortions, the third-order intermodulation is one of the most critical terms responsible for linearity degradation in general RF systems. In order to obtain the third-order intermodulation distortion, input signal is replaced by a two-tone test signal ($V_1 = A\sin(\omega_1 t)$, $V_2 = A\sin(\omega_2 t)$), and the fundamental signal at ω_1 and the distortion at $2\omega_1 - \omega_2$ must be computed by (4-14) and (4-29). By assuming $s_1 = s_2 = j\omega_1 = s$, $s_3 = -j\omega_2 \approx -s$, and $j\omega_1 - j\omega_2 = \Delta s$, then the third-order intermodulation distortion (IM_3) can be expressed as

$$IM_3 = \frac{3}{4}.A^2.\left|\frac{H_{32}(j\omega_1, j\omega_1, -j\omega_2)}{H_{12}(j\omega_1)}\right| = \frac{3}{4}.A^2.\left|\frac{H_{32}(s,s,-s)}{H_{12}(s)}\right|$$
$$= \frac{3}{4}.A^2.\frac{|1+sC_{gd}.Z_1(s)+sC_{gs}.Z_1(s)|}{|sC_{gd}-g_m|}.|i_{NL3gm}(s,s,-s)+i_{NL3go}(s,s,-s)| \quad (4\text{-}30)$$

In addition, the A_{IP3} can be obtained by assuming $IM_3 = 1$ and expressing as

$$A_{IP3}^2 = \frac{4}{3}.\frac{|sC_{gd}-g_m|}{|1+sC_{gd}.Z_1(s)+sC_{gs}.Z_1(s)|}.\frac{1}{|i_{NL3gm}(s,s,-s)+i_{NL3go}(s,s,-s)|} \quad (4\text{-}31)$$

From (4-31). The equation provides good agreement with harmonic-balance simulation. The nonlinear distortions result from the nonlinear effects of the transconductance and the output conductance and vary with different load impedance. For our case, the linearity is dominated by the device transconductance in the low load impedance.

4.5.2 Multiple Gated Transistors Method Using Complex Transconductance Analysis

The MGTR method improves linearity by cancellation of these effects due to auxiliary transistor (AT) as shown in Fig. 4.13. This negative peak of the main transistor (MT) can be cancelled by the positive peak value of a properly bias and size of AT. Because AT is biased in the sub-threshold region, this linearization method does not consume much extra power.

For typical MGTR, using DC transconductance the conventional analysis lacks of accuracy to predict the high-frequency operating condition. Essentially nonlinear distortion is frequency-dependent. An effective method using complex AC transconductance is therefore used to achieve optimized device size and bias for the auxiliary transistor.

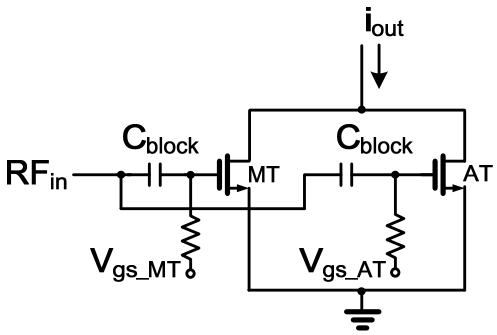


Fig. 4.13 MGTR architecture

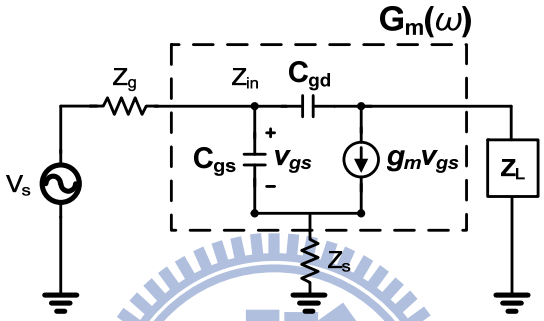


Fig. 4.14 The compact box-type equivalent circuit model

Without loss of generality in this nonlinear analysis, the superposed configuration in MGTR can be simply represented by the transconductance element $G_m(\omega)$. The transconductance $G_m(\omega)$ is defined in the same way as the ratio of the output current to the input voltage, including all the intrinsic and extrinsic frequency-dependency of MOSFET devices. It could be a complex value at high frequencies. Its nonlinearity shall be related to the load impedance and the operation frequency. Consequently the equivalent circuit model of a transconductance amplifier is shown in Fig. 4.14.

The third-order intermodulation product IMD3 in a two-tone test is derived by Volterra series analysis and expressed as

$$IMD_3 = \frac{3}{4} \cdot A^3 \cdot \frac{|Z_s(\omega)+Z_{in}(\omega)+Z_g(\omega)|\cdot|Z_L(\omega)|\cdot|i_{NL3Gm}|}{|Z_s(\omega)+Z_{in}(\omega)Z_s(\omega)G_m(\omega)+Z_g(\omega)+Z_{in}(\omega)|}, \quad (4.32)$$

where

$$|i_{NL3Gm}| = |H(\omega)|^3 \cdot |\varepsilon(\omega, \Delta\omega, 2\omega)|, \quad (4.33)$$

$$|\varepsilon(\omega,\Delta\omega,2\omega)| = \frac{G_m''(\omega)}{3!} - \frac{2}{3}\cdot\frac{G_m'(\omega)}{2!}\left[2Z_s(\Delta\omega)H(\Delta\omega)\frac{G_m'(\omega)}{2!} + Z_s(2\omega)H(2\omega)\frac{G_m'(2\omega)}{2!}\right],$$

(4.34)

Lower IMD3 can be obtained by reducing $\varepsilon(\omega,\Delta\omega,2\omega)$ in Equation (4.32). Similarly the major effort is cancellation of $G''_m(\omega)$, which is conducted in the complex domain. For the value of G''_m for MT is negative in the gate bias voltage v_{gs_MT}=0.77V and the device sizes of MT is NF=13. It can be cancelled by the positive value of AT with a proper bias voltage. This complex transconductance analysis actually suggests that the proper bias voltage v_{gs_AT}=0.46V, and the device sizes of AT are chosen as NF=11, as can be seen, G''_m appears close to zero at the gate bias voltage v_{gs_MT}=0.77V at 7GHz in the polar plot as shown in Fig. 4.15. From Fig. 4.16, we can get the magnitude of complex transconductance to matching the traditional DC transconductance analysis.

For the optimal device size and bias voltage of AT minimizes the value of G''_m. Using the IIP3 contour as shown in Fig.4.17, we observe the value of IIP3 in the CS amplifier by sweeping the device size and bias voltage of AT for the condition which is the gate bias voltage v_{gs_MT}=0.77V for MT, and the device sizes of MT is finger numbers of 13. Final, we choose the proper bias voltage v_{gs_AT}=0.46V, and the device sizes of AT is finger numbers of 11 for the MGTR optimal parameters.

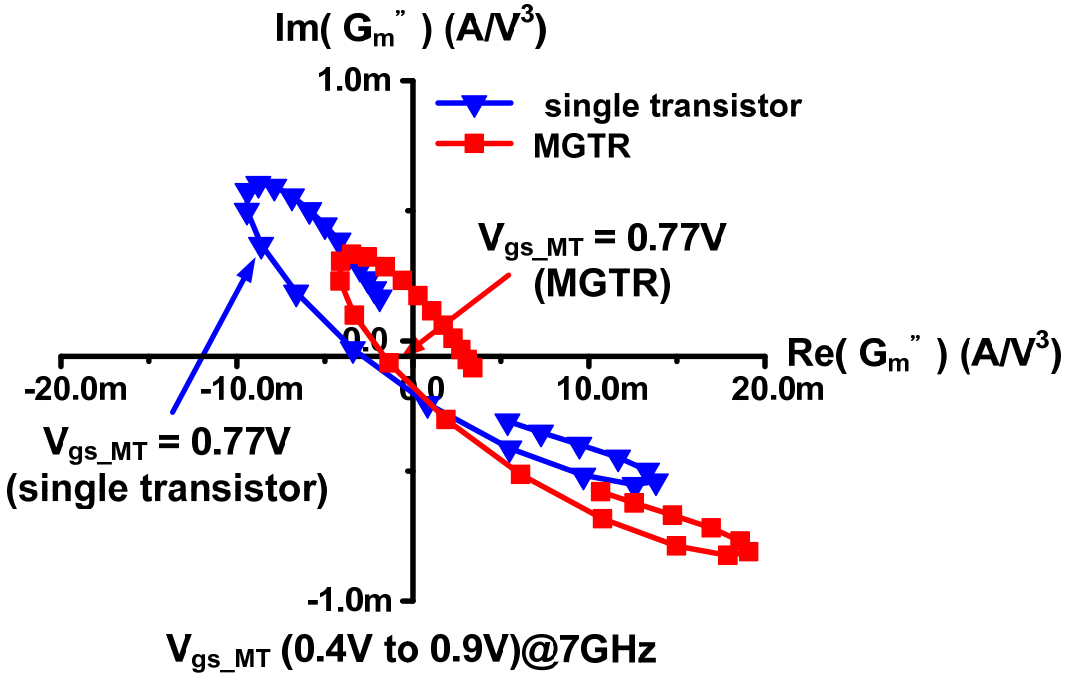


Fig. 4.15 Cancellation of complex AC G_m" in polar plot

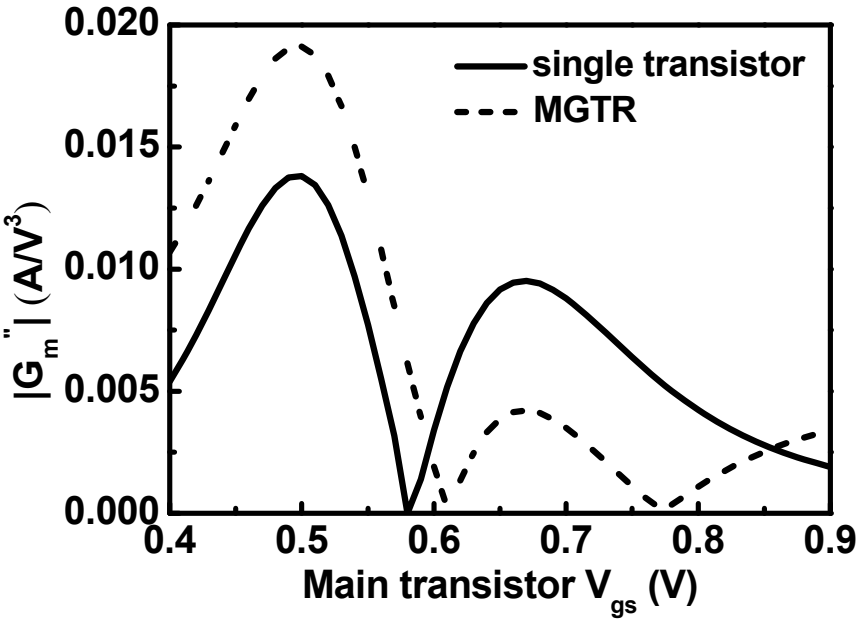


Fig. 4.16 Cancellation of AC g_m" in MGTR configuration

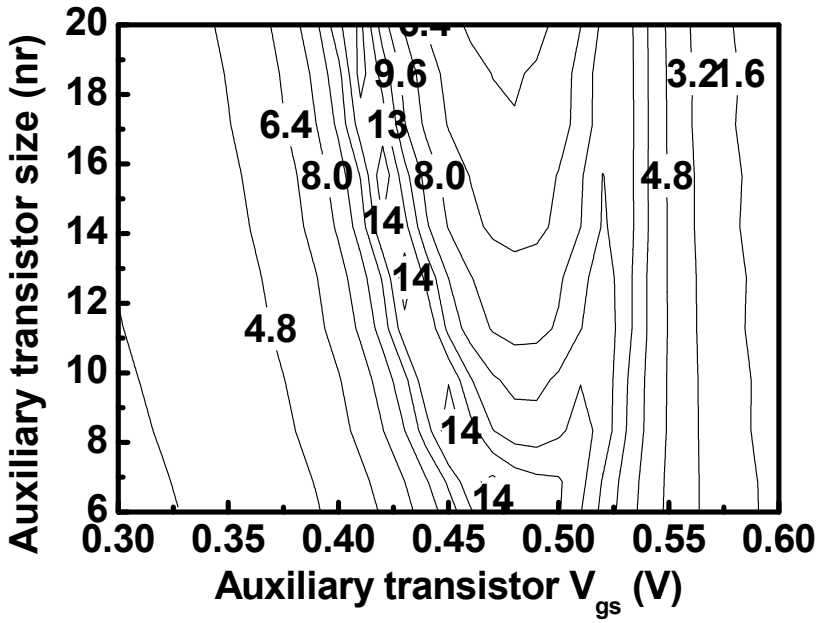


Fig. 4.17 Search for the optimal device size and bias voltage using IIP3 contour

4.5.3 Broadband Linearity Improvement By Using MGTR

For a single transistor, the third-order intermodulation of the transconductance element $G''_m(\omega)$ in a two-tone test is derived by small signal circuit analysis and expressed as

$$G''_m(\omega) = -\frac{3}{4}\left(\frac{g''_m}{3!}\right)v_{gs}^3\left(\frac{Z_L}{(1+\omega^2 C_{gs}^2 Z_s^2)(1+j\omega C_{gs} Z_L)(1+j\omega C_{gs} Z_s)}\right), \quad (4.35)$$

Similarly the major effort is canceling the IMD3 of MT, which is conducted in the complex domain. In order to improve the linearity of broadband LNA, lower output IMD3 can be obtained by reducing the negative IMD3 of MT using the positive IMD3 of AT. From Equation (4.35), we get the IMD3 frequency response of a transistor. To keep the perfect cancellation effect in broadband condition, we choose the similar

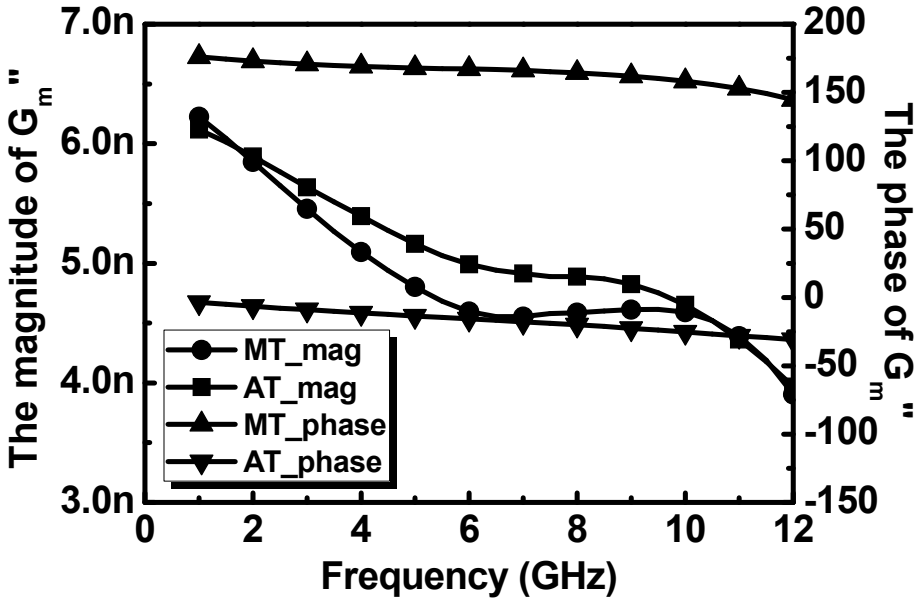


Fig. 4.18 The current of third order intermodulation of MT and AT in broadband condition

device sizes of MT and AT to get the phases of IMD3 which are always differential in operation frequency, and the magnitude of IMD3 in MT, we choose a properly bias to get the same magnitude of IMD3 in AT for broadband cancellation.

In Equation (4.35), we assume the transconductance element $G''_m(\omega)$ has magnitude and phase function depends on frequency, the IMD3 has the different magnitude variation and phase variation in operation frequency as shown in Fig. 4.18 We can get the IMD3 variation between 3GHz and 11GHz, the magnitude of third order intermodulation is similar, and the phase of third order intermodulation is always differential. Therefore, MGTR linearization method can provide good linearity improvement in broadband application. From the bias voltage v_{gs_MT}=0.77V, and the device sizes of MT and AT are chosen as NF=13 and 11, we can observe the value of IIP3 in the CS amplifier by sweeping transistor size and bias voltage of AT in difference frequency as shown in Fig. 4.19. Using the IIP3 contour, we can choose the optimal bias voltage to get the perfect cancellation in broadband condition.

Final, this complex transconductance analysis in broadband condition actually suggests that are chosen v_{gs_AT} is 0.46V to get perfect cancellation in broadband condition. The device sizes of MT and AT are finger numbers of 13 and 11. We see the linearity which be improved by MGTR technique as shown in Fig. 4.20. Finally, the cancellation effect is as good as broadband application between 3GHz and 11GHz. MGTR linearization method can provide good linearity improvement in broadband application as shown in Fig. 4.21.

Fig. 4.19 Search for the optimal bias voltage using IIP3 contour in broadband condition

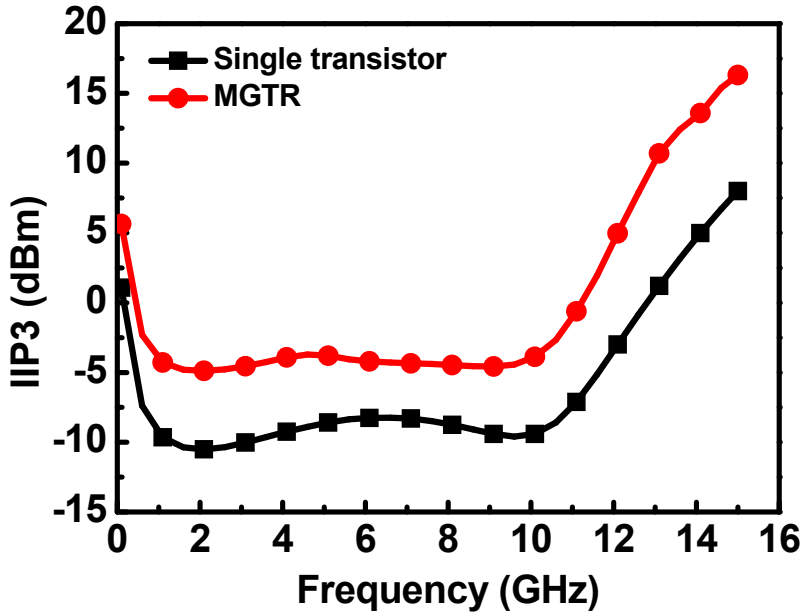


Fig. 4.20 The UWB LNA IIP3 be improved in broadband condition

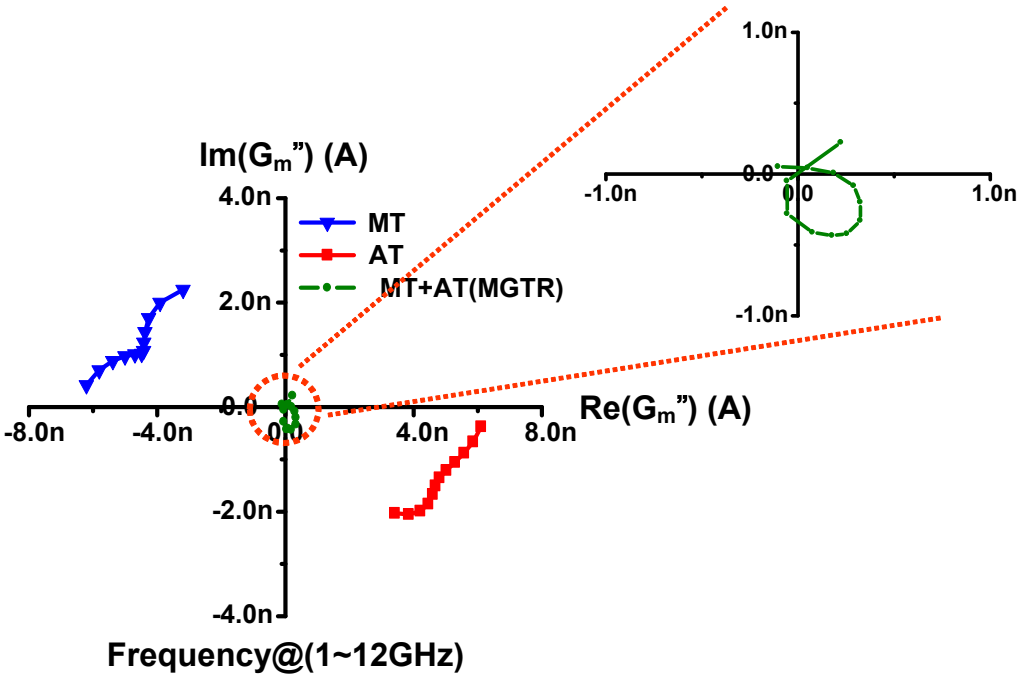


Fig. 4.21 The MGTR cancellation effect in broadband condition

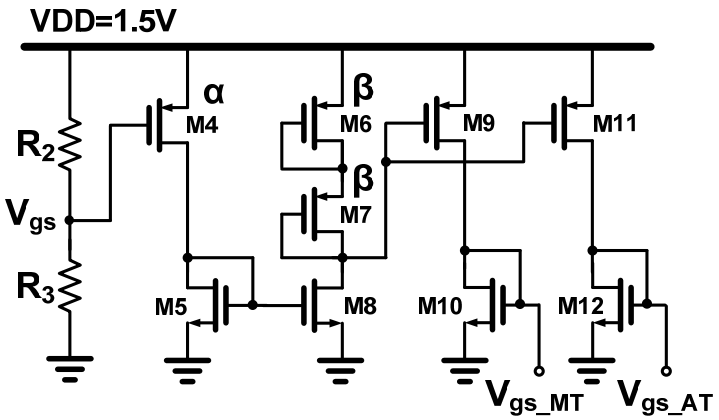


Fig. 4.22 CMOS constant current reference

4.6 CMOS Constant Current Reference

According to the broadband complex transconductance analysis, the MGTR technique can use linearity improvement in broadband condition. Simultaneously, we observe the cancellation effect which is sensitive to the process corners. In various process corners, it makes the variation of voltage and current, so the distortion cancellation effect is affected by the various process corners. Therefore, we need a constant current reference to biasing MT and AT for keeping the cancellation effect, the following discuss will present a CMOS constant current reference designing flow.

In this work, a CMOS constant current reference over process variations is presented as shown in Fig. 4.21. During processing, variations in the process parameters like, t_{ox} (gate oxide thickness), N_{ch} (channel doping), and others affect the reference current. Therefore, the core of the proposed technique is a CMOS circuit, which generates the reference current as the drain current of a MOS device. This drain current is kept relatively constant by exploiting the physical relationship between K' and V_T in process variations.

The drain current of MOS device in saturation region can be expressed as

$$I = K'\left(\frac{W}{2L}\right)(V_{GS} - V_T)^2, \tag{4.36}$$

Due to process variations, K' and V_T, which affect the drain current. The normalized variation in the drain current can be expressed as

$$\frac{\Delta I}{I} = \frac{\Delta K'}{K'} - \frac{2\Delta V_T}{(V_{GS}-V_T)}, \tag{4.37}$$

K' and V_T are primarily dependent on two process parameters, t_{ox} and N_{ch}, and they can be expressed as

$$K' = \frac{\mu\varepsilon_{ox}}{t_{ox}}, \tag{4.38}$$

$$V_T = V_{FB} + \varphi_S + \frac{t_{ox}\sqrt{2\varepsilon_{si}qN_{ch}(\varphi_S - V_{BS})}}{\varepsilon_{ox}}, \tag{4.39}$$

Across the process corners, K' and V_T vary due to variations in t_{ox} and N_{ch}. In Equation (4.38) and (4.39), an increase in K' is accompanied by a decrease in V_T when t_{ox} decreases, or vice versa. Therefore, this correlation between K' and V_T can be interpreted as an inverse relationship between them. At this point, it is important to have some understanding about the modeling process corners in typical CMOS processes.

In Equation (4.37), we assume ($V_{gs} - V_T$) to be equal to V_T. Even though the inverse relationship of K' with V_T causes a wide variation in the drain current, the same inverse relationship can also be used to our benefit if we assume for the moment that the expression of the drain current can be expressed as

$$I = K'\left(\frac{W}{2L}\right)(V_T + V_{ds}(sat))^2, \tag{4.40}$$

In Equation (4.40), assuming $V_{ds}(sat)$ can be kept constant, the variation in the drain can be express as

$$\frac{\Delta I}{I} = \frac{\Delta K'}{K'} + \frac{2\Delta V_T}{(V_T+V_{ds}(sat))}, \tag{4.41}$$

For the process corners, the inverse relationship between K' and V_T will minimize the variation in the drain current given by Equation (4.41). The variation in

the drain current can be minimized to the order of K' variations by choosing a large value of $V_{ds}(sat)$. The following discussion will assume that the inverse relationship of K' and V_T exists across process variations.

In this circuit, there are 3 design parameters: V_{gs}, α and β. The reference current can be expressed as

$$I_{ref} = K_p' \left(\frac{W}{2L}\right)_{M9,11} (V_{GS} - V_{TP})^2, \tag{4.42}$$

Assuming an ideal NMOS current mirror, the reference current can be modified as

$$I_{ref} = K_p' \left(\frac{W}{2L}\right)_{M9,11} \left(V_{TP} + 2\sqrt{\frac{2I}{K_p'(\beta)}}\right)^2, \tag{4.43}$$

$$= K_p' \left(\frac{W}{2L}\right)_{M9,11} \left((1 - 2\sqrt{\frac{\alpha}{\beta}})V_{TP} + 2\sqrt{\frac{\alpha}{\beta}}(V_{GS})\right)^2, \tag{4.44}$$

In Equation (4.44), V_{gs} is a constant voltage, α and β are the ratios (W/L) of the PMOS devices. In order for the reference current to be constant across process corners

$$\frac{dI_{ref}}{dK_p'} = 0, \tag{4.45}$$

Substituting Equation (4.44) in (4.45) gives

$$\frac{dV_{TP}}{[(1-2\sqrt{\frac{\alpha}{\beta}})V_{TP}+2\sqrt{\frac{\alpha}{\beta}}(V_{GS})]} = \frac{dK_p'}{2(1-2\sqrt{\frac{\alpha}{\beta}})K_p'}, \tag{4.46}$$

Both sides of Equation (4.46) can be integrated, and by eliminating the constant of integration using the nominal values of K'_p and V_{TP}, we will get

$$\frac{K_p'}{K_{p.nom}'} = \left[\frac{[(1-2\sqrt{\frac{\alpha}{\beta}})V_{TP.nom}+2\sqrt{\frac{\alpha}{\beta}}(V_{GS})]}{[(1-2\sqrt{\frac{\alpha}{\beta}})V_{TP}+2\sqrt{\frac{\alpha}{\beta}}(V_{GS})]}\right]^2, \tag{4.47}$$

$K'_{p,nom}$, and $V_{TP,nom}$ are the nominal (typical) values of K'_p and V_{TP} in any process variations. Equation (4.47) represents the desired inverse relationship between K'_p and V_{TP} for a constant reference current. For the moment, we can get the optimal ratio of α and β in Equation (4.47). Final, the V_{gs} of the CMOS constant current reference is 1.3V, the ratio of α and β is 5:12. In the simulations, the variation of the CMOS constant current reference was ±0.8% of its nominal value in process corners.

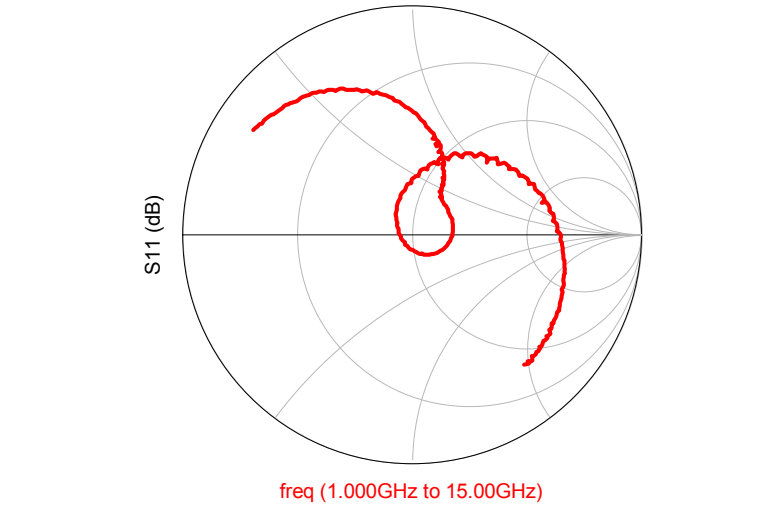
4.7 Chip Implementation and Measured Result

The chip micrograph is shown in Fig. 4.22. The parasitic shunt capacitance of the C_C is employed as the C_1 for matching network. Fig. 4.23 shows the measured input impedance matching S_{11}. Fig. 4.24 shows the power gain S_{21}. The measurement and simulation results have a 5-dB difference, which was found due to the inaccurate SPICE model and metal conductivity in EM simulation. The measured noise figure (NF) are shown in Fig. 4.25. The measured in-band noise figure is 7.3dB. The measured NF higher than the simulated one is due to the degraded power gain. Fig. 4.26 shows the measured IIP3 of the MGTR UWB LNA at two conditions. One is the MGTR operation condition, and the other is single transistor condition that the AT gate bias voltage is ground. As can be seen, linearity improvement achieves more than 5dB without extra power consumption. The linearity performance is improved with a measured IIP3 of about -5dBm in-band. The power consumption without output buffer is 8.26mW from a 1.5V supply. The performance is summarized and compared with other in Table. 4.1.

Fig. 4.23 Chip micrograph of the UWB LNA.

Table. 4.1 Measure performance summary

	This Work	[1]	[2]	[3]	[4]
Technology	0.18um CMOS	0.18um CMOS	0.18um CMOS	0.13um CMOS	90nm CMOS
Power Supply	1.5V	1.8V	1.5V	1.2V	1.0V
BW (GHz)	3.1 ~ 10.6	2.3~9.2	3.1 ~ 10.6	3.1 ~ 10.6	3.1~10.6
$S_{11,MAX}$ (dB)	<-8	-9.4	-11	-9.9	<-10
S_{21} (dB)	8.1	9.3	11	13.7	7.8~12.3
Noise Figure	7.31dB	8 dB	5.1 dB	3.0dB	2.7~3.3dB
Power (w/o buffer)	8.26 mW	9 mW	9 mW	9 mW	2.5mW
IIP3	>-5dBm	-16dBm	-12dBm	-8.5dBm	-6.4dBm



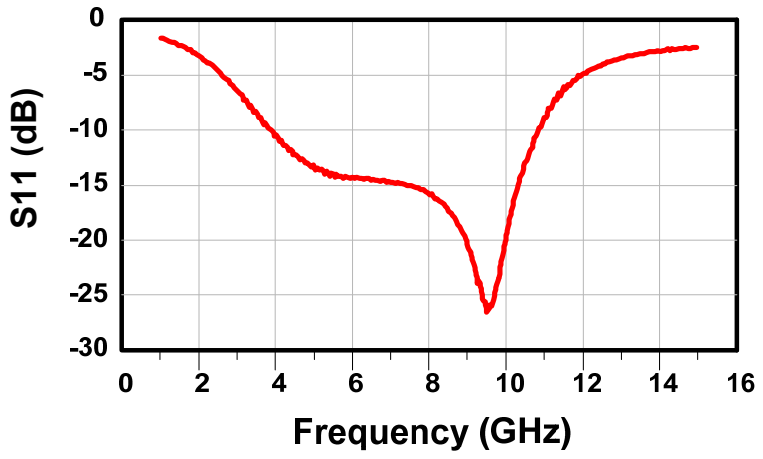


Fig. 4.24 Measured input impedance matching S_{11}

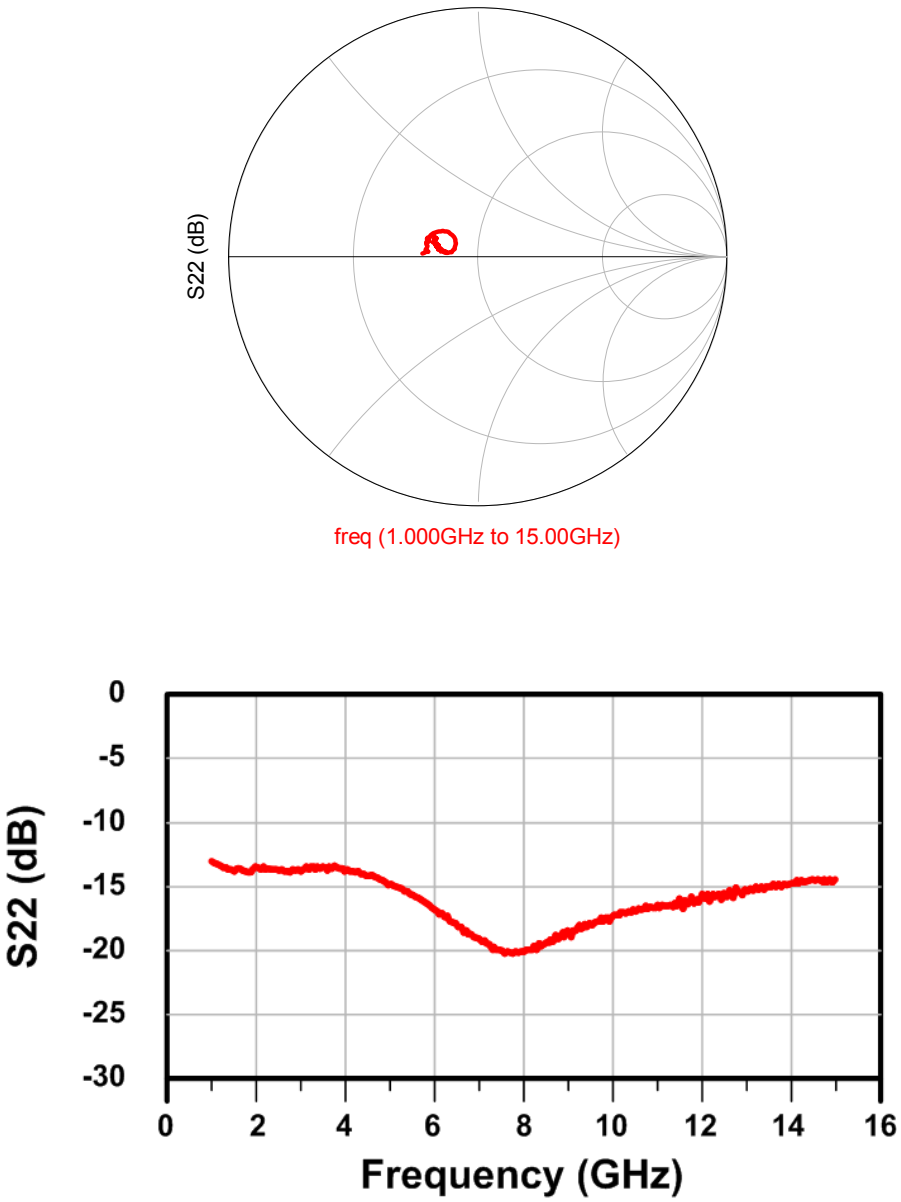


Fig. 4.25 Measured output impedance matching S_{22}

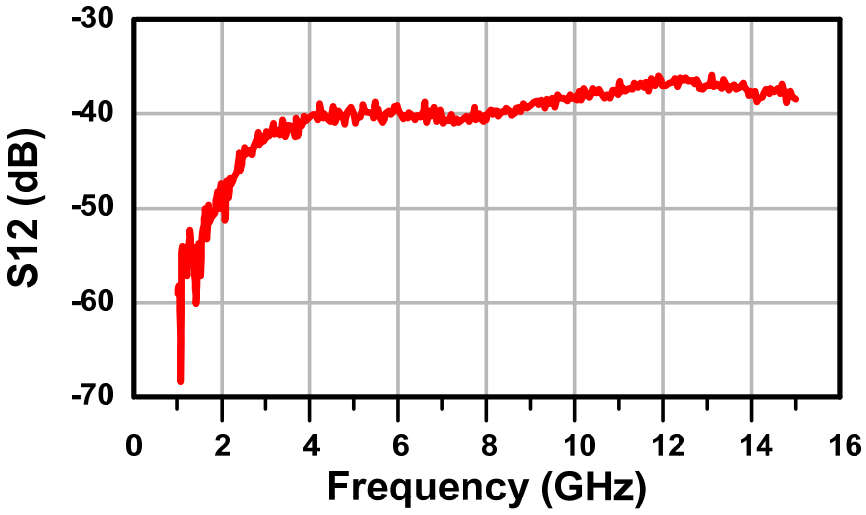


Fig. 4.26 Measured S_{12}

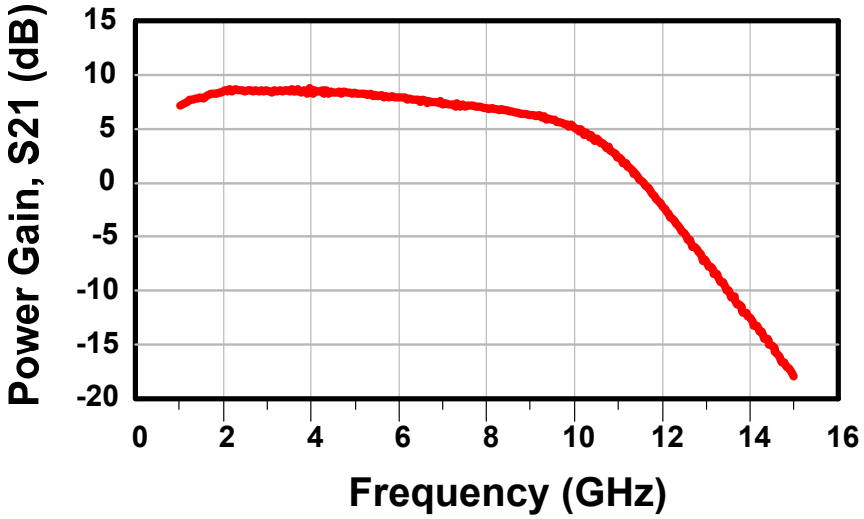


Fig. 4.27 Measured power gain S_{21}.

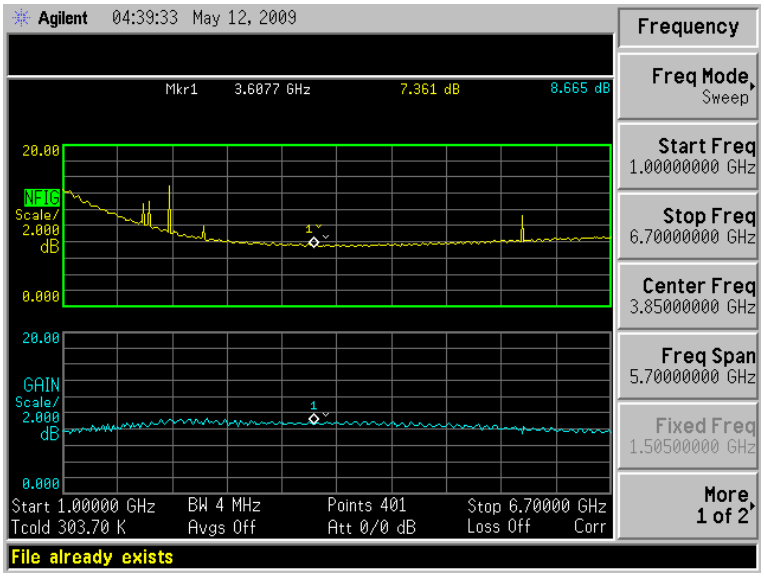


Fig. 4.28 Measured noise figure

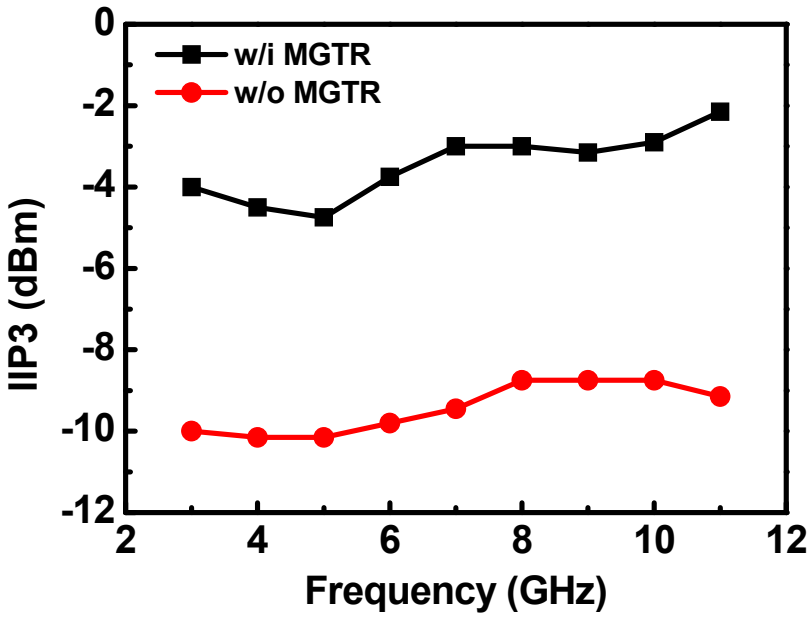


Fig. 4.29 IIP3 measurement of w/i MGTR and w/o MGTR broadband LNA

4.8 Summary

The proposed broadband simultaneous noise and impedance matching technique is employing dual reactive feedbacks and a high-order input matching network to cover a wide bandwidth.

A low-power high-linearity UWB LNA, intended for use in the receiver path of ultra-wideband wireless system, is designed in standard 0.18um CMOS technology. The MGTR technique is adopted to improve the linearity of the broadband LNA. A common source equivalent circuit using AC complex transconductance analysis provides broadband linearity improvement, IIP3 contour gives the optimal design parameters. Measured data show that the improvement of the linearity is more than 5 dB in broadband condition.

Chapter 5

A Low Power, High Linearity UWB Receiver for Ultra-Wide Band Wireless System

5.1 Introduction

This chapter is aimed at the low power, high linearity UWB RF front-end circuit in wireless receivers. For the concern of low power consumption, Direct Conversion Receiver is chosen as the system architecture. It provides great possibility of better form factor, lower cost, less power consumption, and high linearity. In this work, a low power, high linearity UWB direct down-conversion front-end circuit is implemented. The front-end circuit includes a low noise amplifier, an active balun, and a direct down-conversion mixer, as shown in Fig. 5.1. The UWB LNA is the same in the chapter 4.

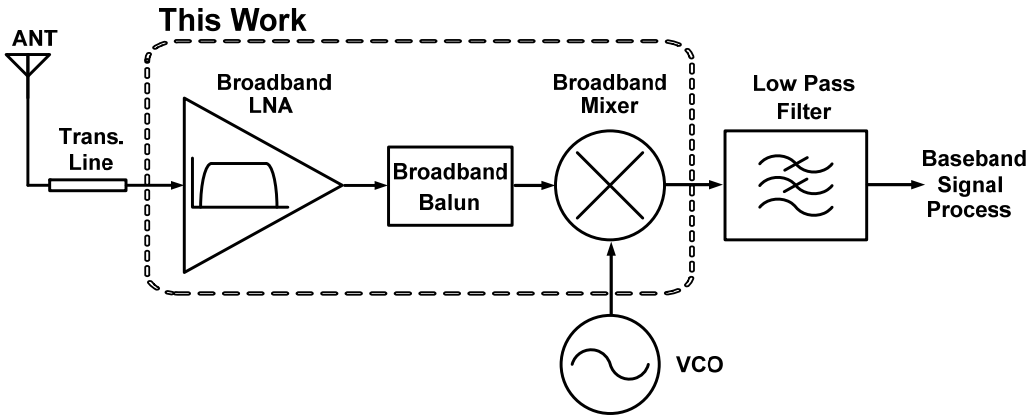


Fig. 5.1 UWB receiver front-end architecture.

Mixer is an essential part of RF front-end circuits. Double balanced type mixer is better than single-ended one for its better port-to-port isolation and even-order terms rejection. The received signal from the antenna is usually single ended. An active

balun is needed to transform the signal from the proceeding stage into a differential form to benefit from the double-balanced structure. To avoid unwanted signal loss, an active balun is usually adopted. However, the active balun contributes limited gain and consumes some power. Therefore, in this work the active balun and transconductance stage are combined to a single stage to largely save power consumption as shown in Fig. 5.2. It consists of common gate and common source transistors which is claimed to have averagely good performance over other inspected types. The transconductance stage has been analyzed in the chapter. The NMOS switching stage with large size has been selected for performance consideration.

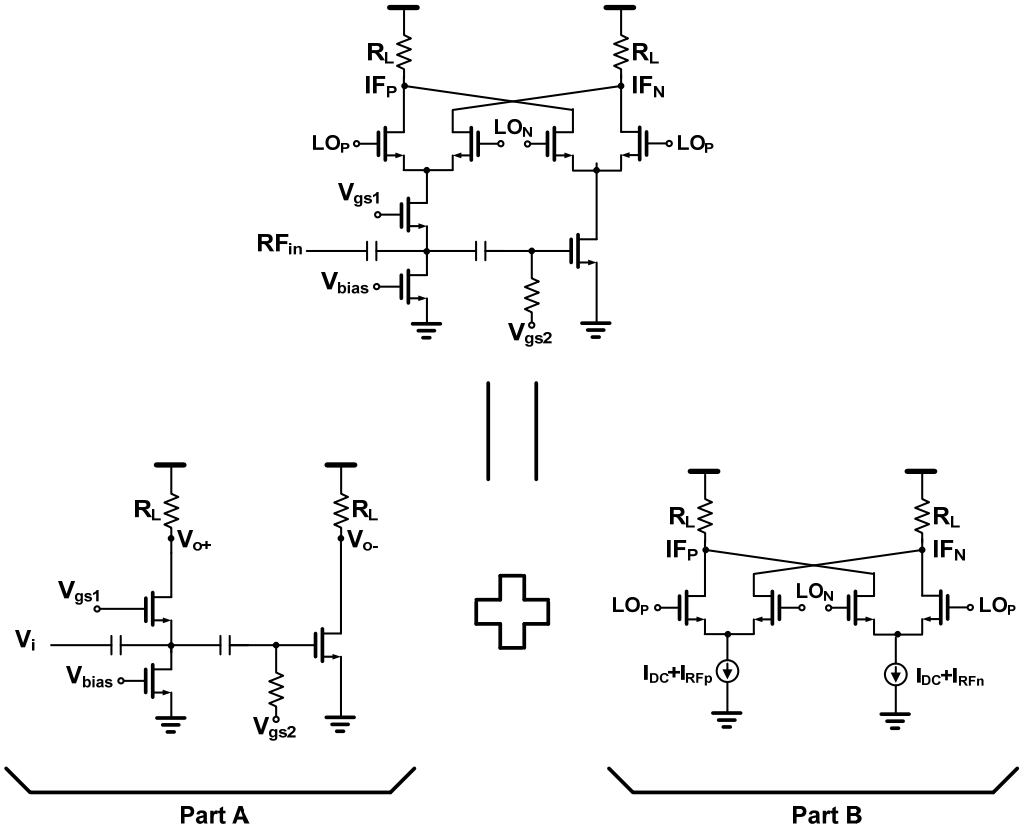


Fig. 5.2 The active balun mixer

5.2 Principle of the Mixer Circuit Design

In this chapter, the design principle of the low power, high linearity UWB front-end circuit is introduced. Fig. 5.3 shows the schematic of the low power, high linearity broadband down conversion mixer circuit. The principle emphasizes on the features of each block, which include low noise amplifier, active balun, and direct down-conversion mixer.

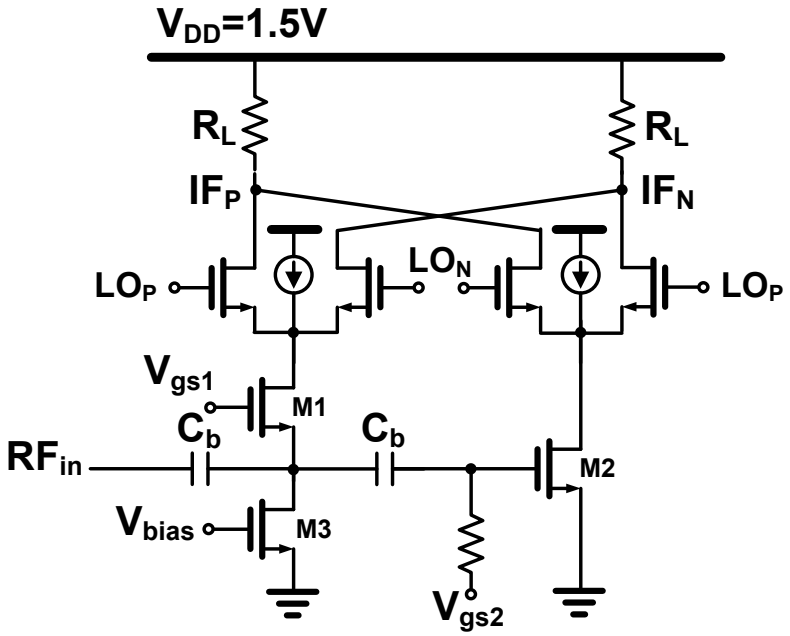


Fig. 5.3 The schematic of the broadband down conversion mixer circuit

5.2.1 Transconductance stage

The transconductance stage consists of two transistors, a common gate transistor M_1 and a common source transistor M_2 as shown in Fig. 5.4. The RF input signal is transformed into differential current by M_1 and M_2, respectively. The output differential current is then connected to the switching stage for current commutation, which loads the transconductor stage and makes the current into voltage V_{o1} and V_{o2} at drain nodes. In the analysis, the parasitic capacitance C_{gd1} and C_{gd2} are not

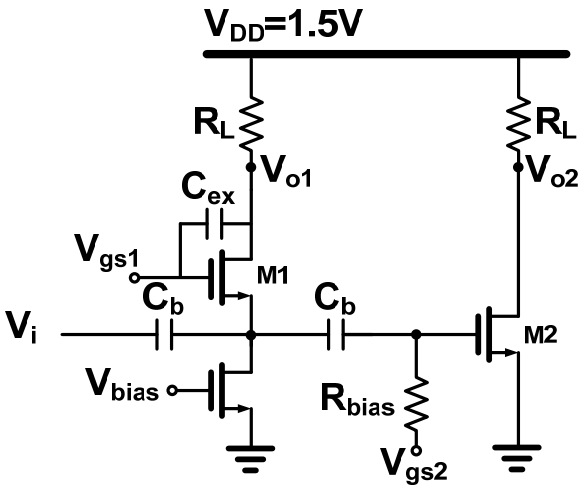


Fig. 5.4 Common-gate common-source transconductance stage

neglected and r_{o1} and r_{o2} are also taken into consideration.

Applying KCL to the small signal models shown in Fig. 5.5, the equations of output voltages to input voltage V_i is obtained as:

$$\frac{V_{o1}}{V_i} = \frac{Z_L / r_{o1} + g_{m1} Z_L}{1 + Z_L / r_{o1} + sC_{gd1} Z_L} \tag{5.1}$$

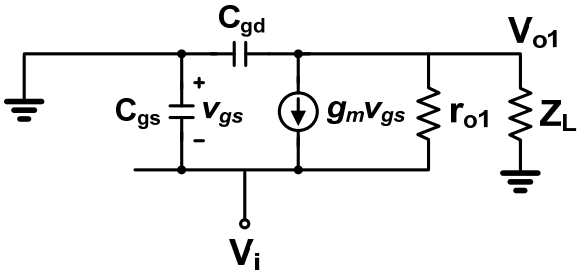
$$\frac{V_{o2}}{V_i} = \frac{-g_{m2} + sC_{gd2} + sZ_{s2}\left(\frac{C_{gs2} + C_{gd2}}{r_{o2}} + g_{m2} C_{gd2} + sC_{gd2} C_{gs2}\right)}{\frac{1}{r_{o2}} + \frac{1}{Z_L} + sC_{gd2} + \frac{Z_{s2}}{Z_L}\left(\frac{1}{r_{o2}} + g_{m2}\right) + sZ_{s2}\left(\frac{C_{gs2} + C_{gd2}}{r_{o2}} + g_{m2} C_{gd2} + sC_{gd2} C_{gs2} + C_{gs2} / Z_L\right)} \tag{5.2}$$

Where Z_L models the loading impedance of the switching pairs and V_i' is assumed close to V_i at the operating frequency for the large capacitance C_b which is added to do the dc blocking. The large resistor R_b is used to give dc bias voltage and neglected in the analysis.

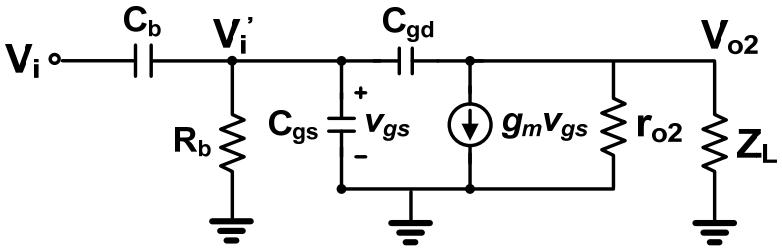
Suppose that Z_{s2} is equal to zero in (5.2), then (5.2) would be:

$$\frac{V_{o2}}{V_i} = \frac{-(g_{m2} - sC_{gd2}) Z_L}{1 + Z_L / r_{o2} + sC_{gd2} Z_L} \tag{5.3}$$

(5.1) and (5.3) must be equal in magnitude and out of phase at the operation frequency. Assume a pure resistance loading, Z_L=R_L, and equal size for the two transistors. The



(a)



(b)

Fig. 5.5 Common-gate common-source transconductance stage
(a) small signal model of common gate
(b) small signal model of common source

magnitude and phase of the two equations is extracted:

$$\left|\frac{V_{o1}}{V_i}\right| = \frac{(1/r_{o1}+g_{m1})R_L}{\sqrt{(1+R_L/r_{o1})^2+(wC_{gd1}R_L)^2}}, \left|\frac{V_{o2}}{V_i}\right| = \frac{\sqrt{g_{m2}^2+(wC_{gd2})^2}R_L}{\sqrt{(1+R_L/r_{o2})^2+(wC_{gd2}R_L)^2}} \quad (5.4)$$

$$Magnitude\_balance: \frac{g_{m1}}{\sqrt{1+(\omega C_{gd1}R)^2}} = \frac{\sqrt{(g_{m2})^2+(\omega C_{gd2})^2}}{\sqrt{1+(\omega C_{gd2}R)^2}}$$

$$Arg\left(\frac{V_{o1}}{V_i}\right) = -\angle\tan^{-1}\frac{wC_{gd1}R_L}{1+R_L/r_{o1}}, Arg\left(\frac{V_{o2}}{V_i}\right) = \pi - \angle\tan^{-1}\frac{wC_{gd2}R_L}{1+R_L/r_{o2}} - \angle\tan^{-1}\frac{wC_{gd2}}{g_{m2}} \quad (5.5)$$

$$Phase\_balance: \angle\tan^{-1}(\omega C_{gd1}R) = \angle\tan^{-1}(\omega C_{gd2}R) + \angle\tan^{-1}(\omega C_{gd2}\frac{1}{g_{m2}})$$

From (5.4) and (5.5), it is found that g_{m1} does not exist in (5.5) that is g_{m1} could adjust the magnitude balance condition without any impact on the phase difference. Moreover, from (5.5), the phase difference of the two equations could be adjusted by one coefficient C_{gd1} by adding a parallel capacitance C_{ex} to the gate to drain capacitance of M_1. Since C_{gd1} will affect both the phase and gain difference, the C_{ex} is

defined to make the phase balanced first, then to adjust g_{m1} to meet the magnitude condition. The bias voltages of CG and CS stage are separated, so g_{m1} can be adjusted by gate voltage independently. The value of g_{m1} and g_{m2} would not be far apart, and the values of r_{o1} and r_{o2} are expected to be almost equal.

The input admittance Y_{in} of the mixer can also be derived from the small signal model shown in Fig. 5.4:

$$Y_{in\_CG} = \frac{1}{Z_{s1}} + g_{m1} + sC_{gs1} + \frac{1}{r_{o1}} - \frac{1/r_{o1}^2 + g_{m1}/r_{o1}}{sC_{gd1} + 1/r_{o1} + 1/Z_L}$$

$$Y_{in\_CS} = sC_{gs2} + sC_{gd2}\frac{g_{m2} + 1/r_{o2} + 1/Z_L}{sC_{gd2} + 1/r_{o2} + 1/Z_L} \quad (5.6)$$

$$Y_{in} = Y_{in\_CG} + Y_{in\_CS}$$

Where Y_{in_CG} is the input admittance of the common gate transistor M_1 only and Y_{in_CS} is the input admittance seen into the gate of the common source transistor M_2. The real part of the input impedance is mainly the parallel combination of $1/g_{m1}$ and Z_{s1}. By Z_{s1}, the gm value can be much more released, that is the power consumption can be much lower when making the input impedance matches to the source resistance.

Fig. 5.6 shows the gain error in broadband condition. The gain error at the operation frequency is around -0.14~0.3dB. And the phase error is always 178~180degree, as shown in Fig. 5.7.

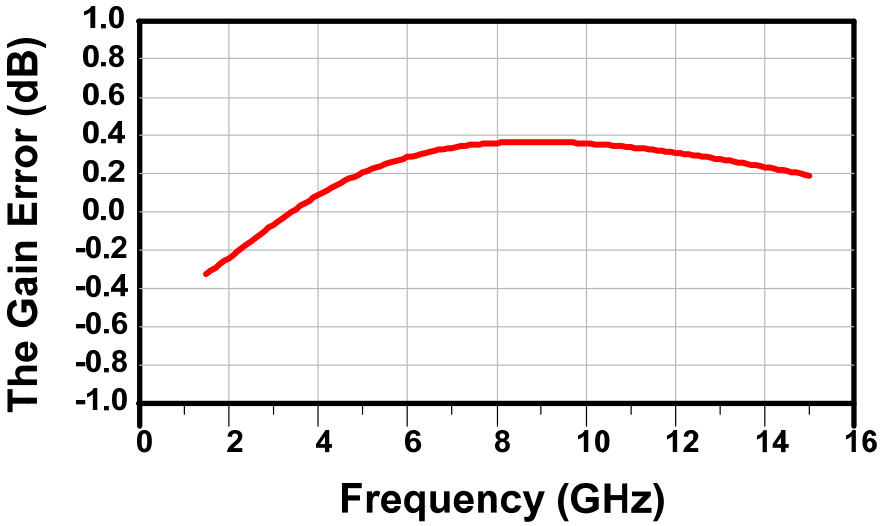


Fig. 5.6 Simulation result of gain error

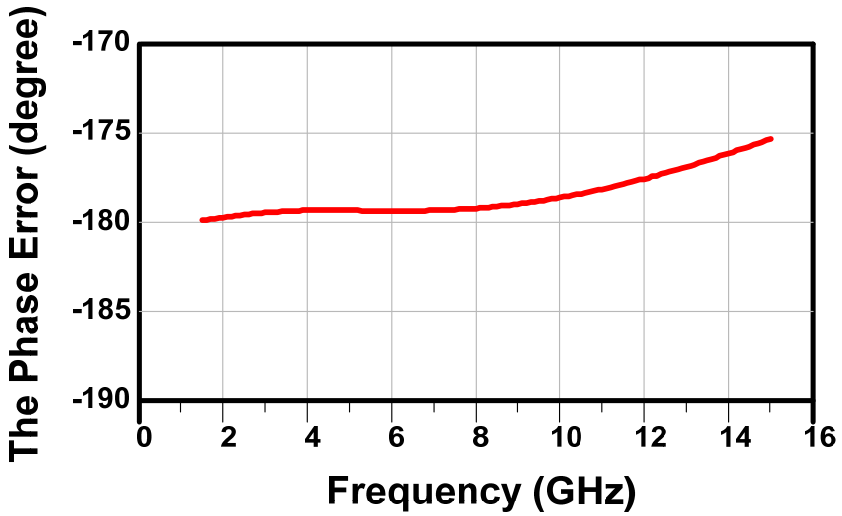


Fig. 5.7 Simulation result of phase error

5.2.2 Mixing stage

In this work, the proposed mixer utilizes current commutation for frequency mixing. A mixing stage is constructed to transform the incoming RF signal to a lower frequency as shown in Fig. 5.8. The non-ideal switching character and noise contribution will alleviate the circuit performance. To make the switching behavior more ideal, MOSFET of larger size is chosen and biasing point is set near threshold voltage. As pointed out in the beginning of this paper, the effort to convert signal into

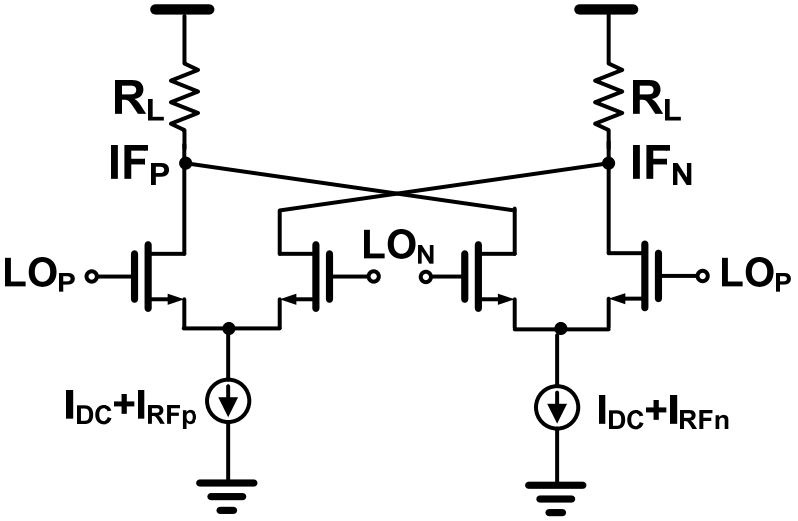


Fig. 5.8 Mixing stage

differential form is to make mixer in double-balanced structure. The even-order distortion and the LO-IF feedthrough are eliminated in the double-balanced mixer. Moreover, large MOS device size is chosen for its lower flicker noise. The flicker noise of MOSFET is appeared in low frequency range around DC, much lower than LO frequency, it can be effectively modeled as interference at the gate terminal of switching component. This slowly varying offset voltage disturbs the switching time, advancing or retarding the time of zero crossing. Mixed with the LO signal, the low frequency noise is up-converted to frequency around the LO frequency which degrades the function of mixing. Final, the mixer topology is chosen as NMOS pairs.

5.2.3 Current injection method

In order to increase the mixer linearity and gain, there are only two or three possibilities: to increase the current flowing through the trans-conductors or to increase the load impedance or both of them. At a given current there is a limitation for increasing the load resistors because of the voltage drop along these resistors. On the other hand, higher current of the trans-conductors improves the gain and the linearity of the mixer but forces more current flowing through LO switches. The latter

effect causes a non ideal switching, which means that both switching transistors are simultaneously in on state for a longer time. Thus more and more RF current is lost as a common-mode signal. For a given LO amplitude, the switching time of the transistors can be reduced by reducing the drain current flowing through them. This can be achieved by the current injection method.

Current injection method is employed to increase the linearity of this mixer. Two dc currents are injected at the drains of the bottom stage transistors. The value of this DC current is optimized, so that it provides the best linearity, gain, power consumption, and bandwidth for the mixer. The other consideration made in designing the current source is that, its output impedance should be large, so that it does not attenuate the RF signal passing through the bottom stage of the mixer to the top transistors. The noise contribution of the switching stage is also reduced, because of the lower DC-current flowing through the switches. So we use the current injection method to improve the mixer linearity and gain.

5.3 UWB receiver Simulation Results and Comparison

The direct down-conversion mixer transforms the radio-frequency (RF) signal into base-band directly and needs high linearity to avoid the distortion of the signal. Following the previous discussion, a low-power and high-linearity UWB receiver is designed, as shown in Fig. 5.9 and Fig. 5.10.

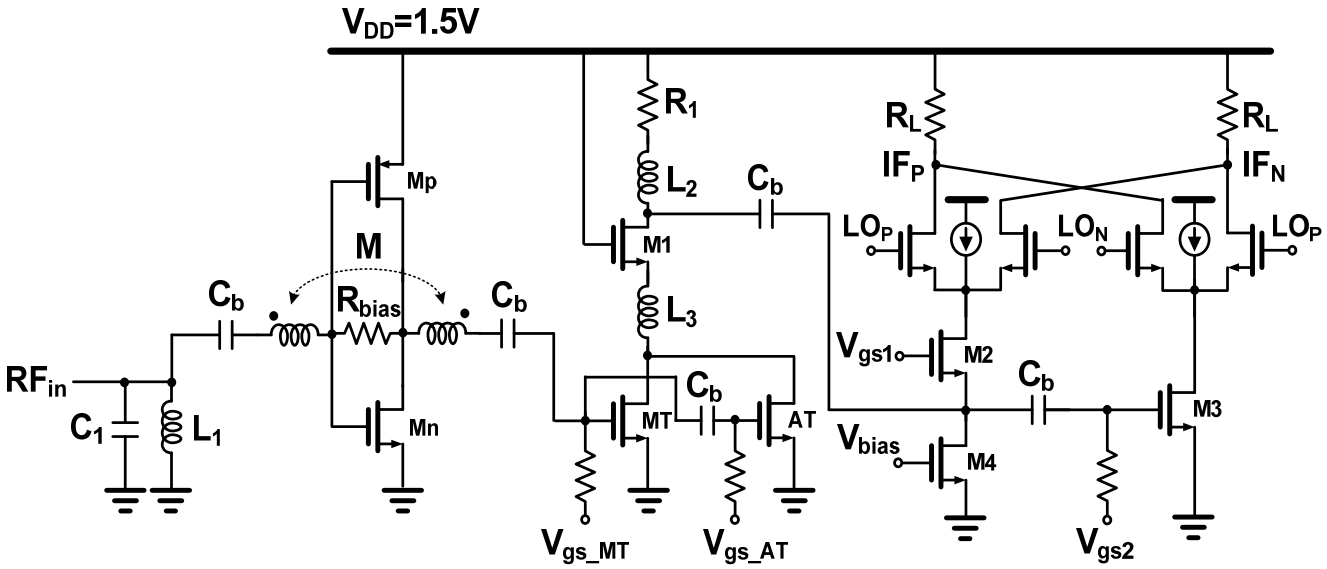


Fig. 5.9. UWB receiver architecture

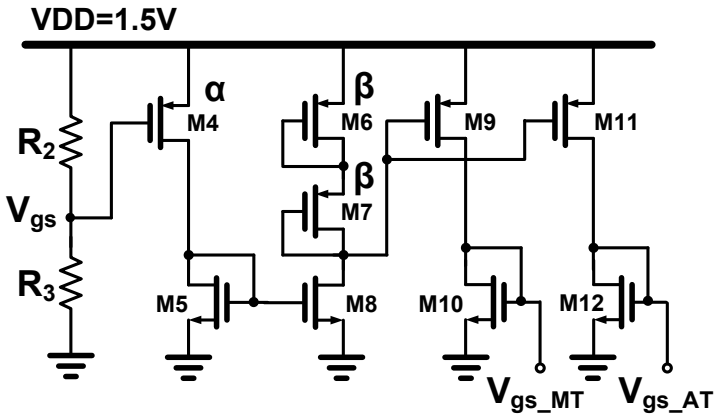
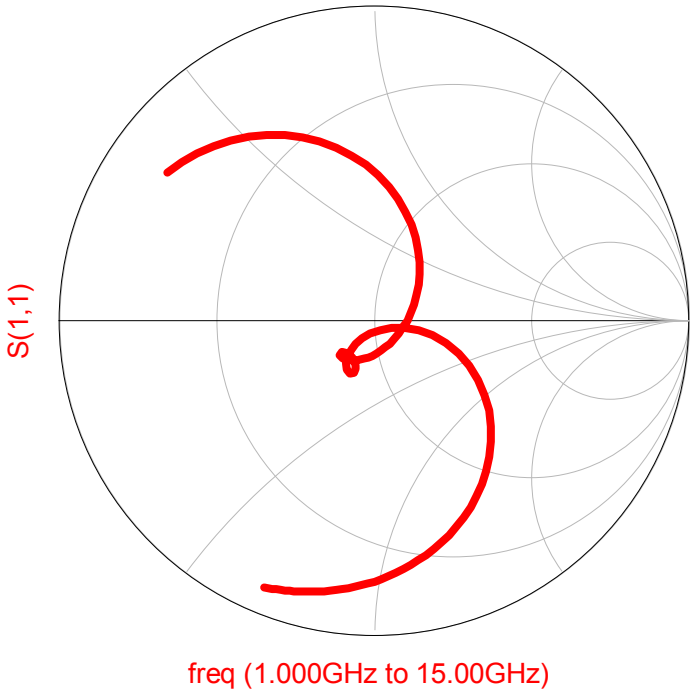


Fig. 5.10. CMOS constant current reference

The simulation input return loss is plotted in Fig. 5.11. The simulation result of conversion gain versus radio frequency (RF) is shown in Fig. 5.12 and DSB noise figure is 6.9dB at IF of 100MHz.

The conversion gain of UWB front-end circuit is 20.3~21.1dB. Linearity analysis is conducted by the two-tone test. Two-tone test is done for measuring third-order intermodulation distortion. The simulation IIP3 is about -4.7~-5.8dB as shown in Fig. 5.13, and the power consumption of active mode is 10.4mW. Complete simulation results are summarized in Table 5.1 together with simulation results for comparison.



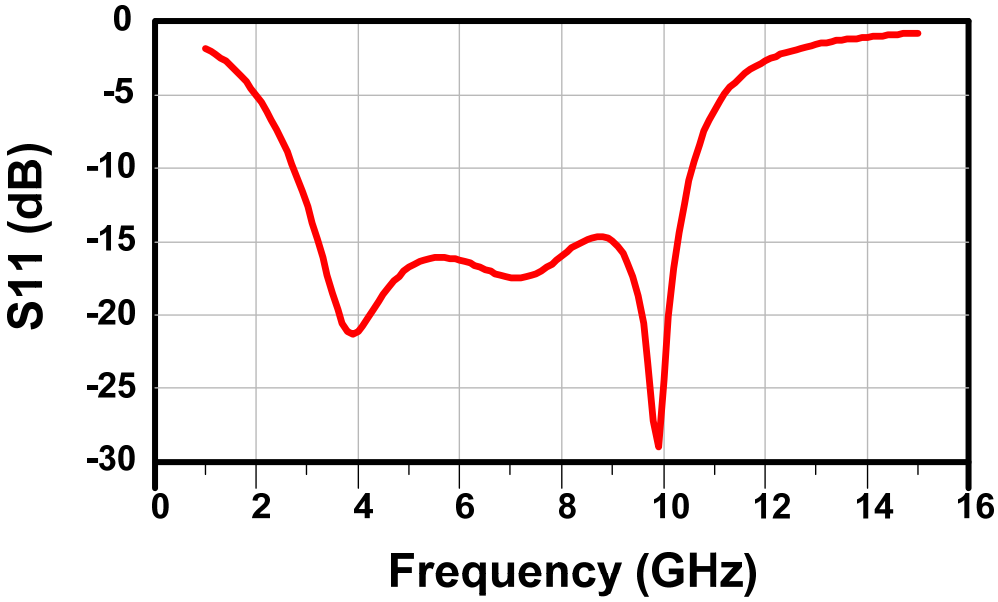


Fig. 5.11. The simulation of input return loss S11

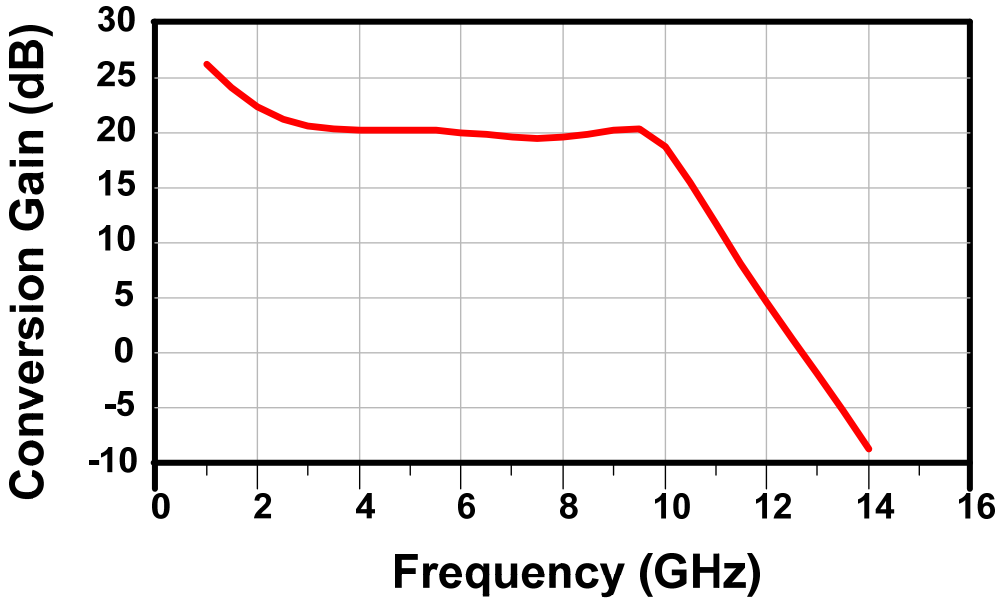


Fig. 5.12. The simulation of conversion gain

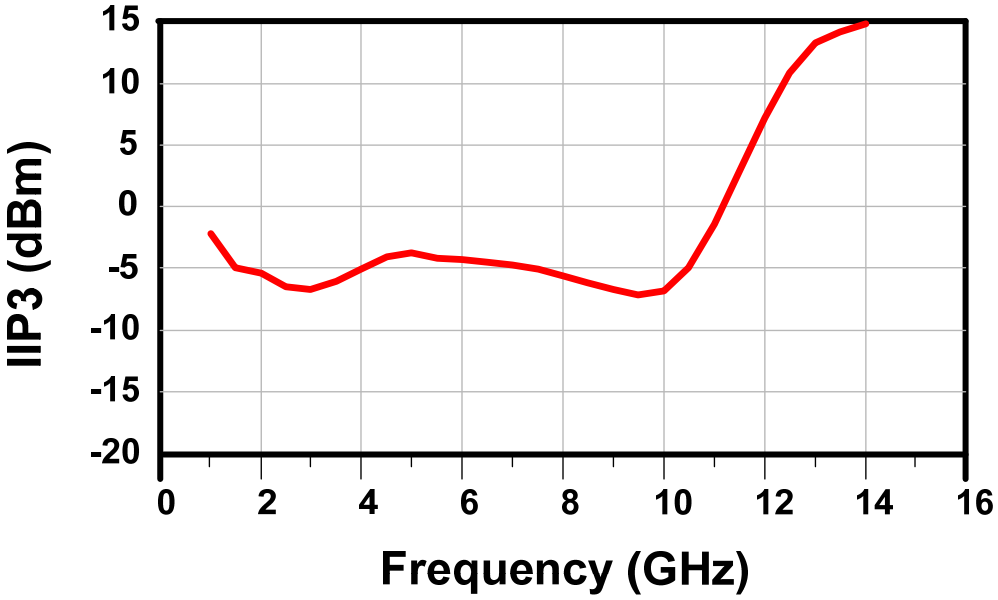


Fig. 5.13. The simulation of IIP3

Table 5.1 Summary of performance and comparison

	This work (sim)	[13]	[14] 07APMC	[15] 08ISSCC	[16] 07EuMa
RF Bandwidth	3.1~10.6 GHz	3.1~10.6 GHz	3.1~8 GHz	0.5~7 GHz	3.1~10.6 GHz
Technology	0.18um CMOS	0.13um CMOS	0.18um CMOS	65nm CMOS	0.13um CMOS
VDD	1.5V	1.5V	1.5V	1.2V	1.5V
Power	10.4mW	48mW	19.25mW	16mW	42mW
Conversion gain(dB)	20.3~21.1	22.9~26.4	19~21.5	18	19.5~23.3
Noise Figure(dB)	6.9	4.8~7.7	4.3~6.2	5.5	5.2~9.2
S11(dB)	<-10	<-10	<-10	<-10	<-10
IIP3(dBm)	-4.7~-5.8	-11.5	-17	-3	-10.4
P1dB(dBm)	-15	-21	N/A	N/A	-22.7

$$FOM = \frac{Gain * IIP3 * BW}{Power * (NF - 1)}$$

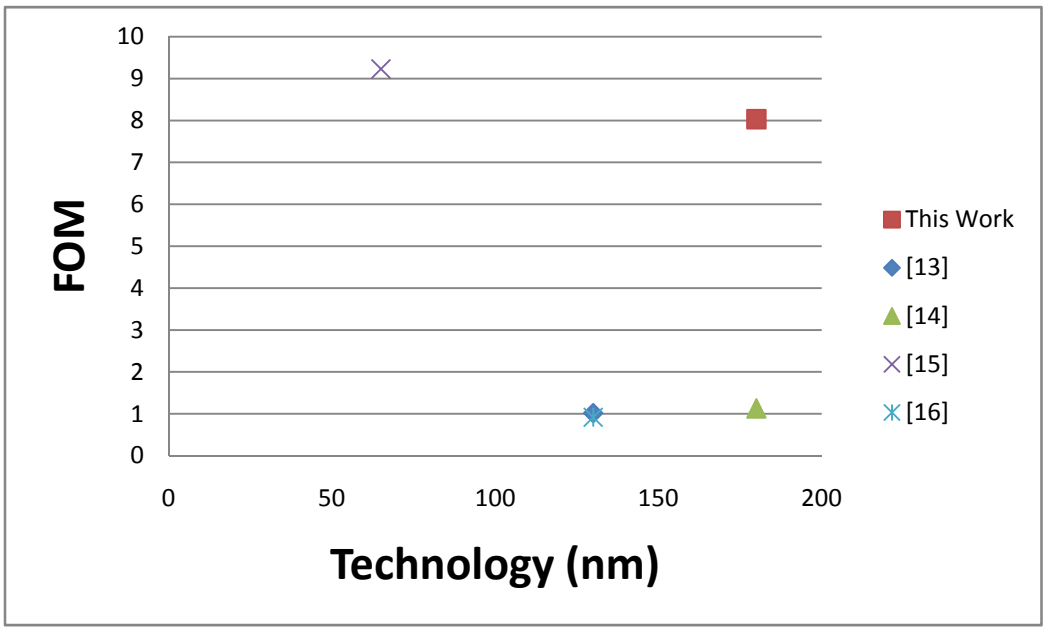


Fig. 5.14 The FOM of UWB receiver performance

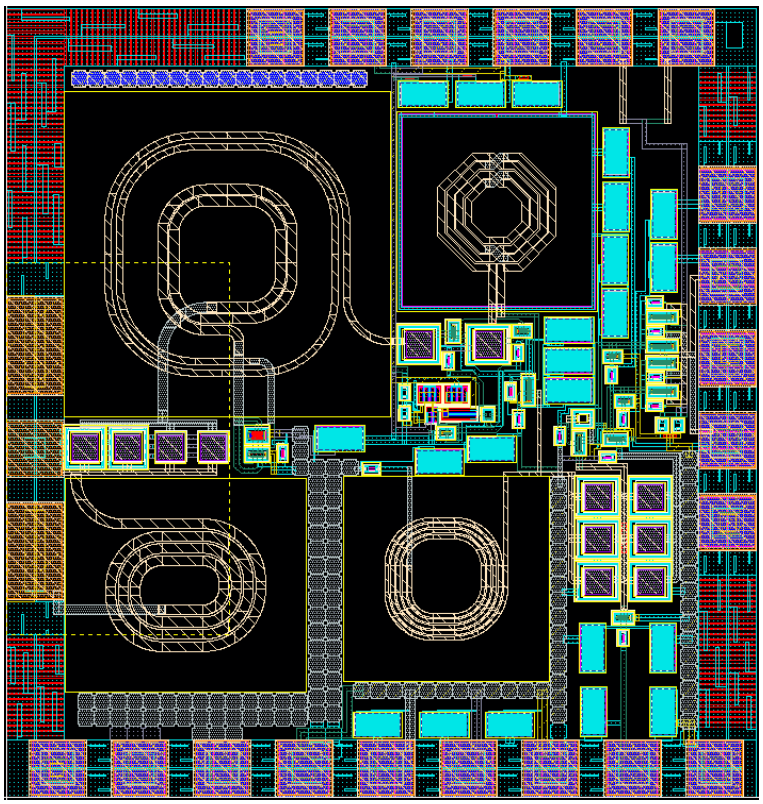


Fig. 5.15 Chip layout of UWB receiver front-end circuit

5.5 Chip Implementation And Measurement Considerations

5.5.1 Circuit Implementations

Following the above-mentioned analysis, a UWB front-end circuit is designed, as shown in Fig. 5.9 and Fig. 5.10. The front-end circuit is fabricated using 0.18um RF CMOS technology. The full chip layout of circuit is shown in Fig. 5.15. The total die area including bonding pads is 0.913 mm by 0.962 mm.

The RF input is placed on the left side, the LO input is placed on the right side, and the IF input is placed on the top side of the chip. The placement of pads is considered for the on-board measurement. In order to minimize the effect of the substrate noise on the system, a solid ground plane, constructed using a low resistive metal material, is placed between the signal pads and the substrate. Besides, there are many ground pads to minimize the effect of the bond-wire.

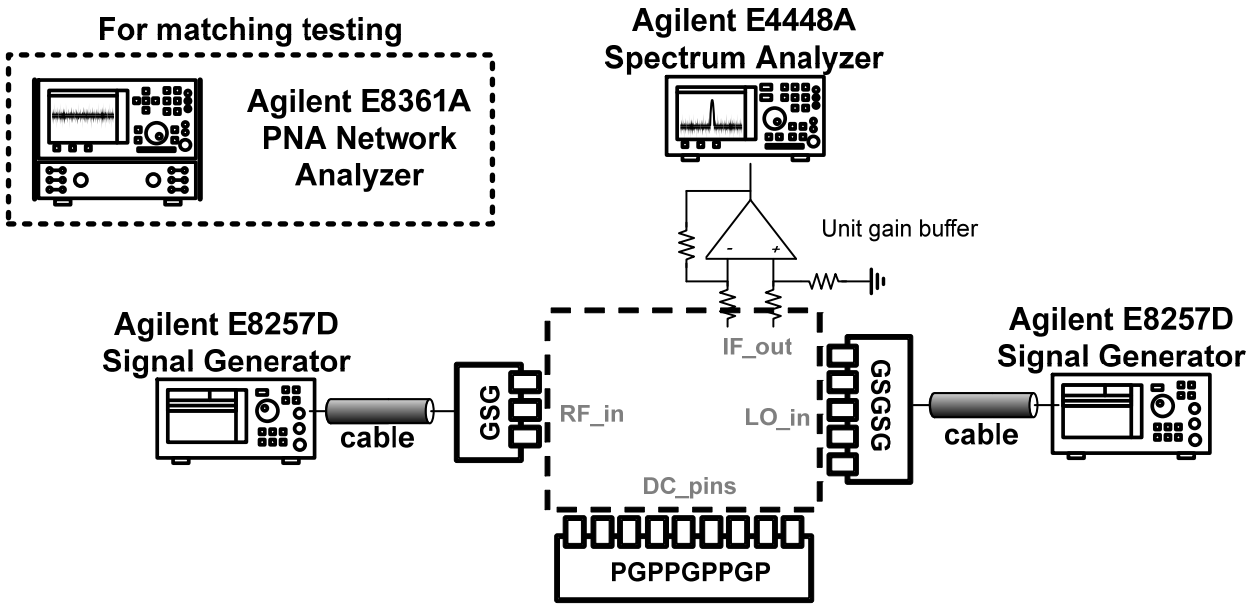


Fig. 5.16 Measurement diagram including unit gain output buffer

5.5.2 Measurement Considerations

Measurement is conducted by mounting the mixer die on FR4 board. Input testing signal is transformed into a differential form at LO ports by GSGSG differential probe. Fig. 5.16 shows the measurement diagram. A unit gain output buffer is used to transform the differential signal into the single-ended form, and provides high input impedance to reduce loading effect.

5.6 Summary

A low-power, high linearity UWB front-end circuit, intended for use in the receiver path of a wireless local area network, is designed in standard 0.18um CMOS technology. The circuit architecture is chosen available for the application of low power consumption and high linearity. It is composed of a low noise amplifier, a active balun, and a direct down-conversion mixer. The linearity improvement is using MGTR circuit. Simulation results shows that the UWB front-end circuit achieves conversion gain of 20.3~21.1dB, input return loss of -10dB, input third-order intercept point (IIP3) of -4.7~-5.8dBm, and input power consuming only 10.4mW.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In Chapter 2, some architectures of the receiver, noise sources, and the theoretical MOSFET noise model are introduced. In Chapter 3, the design consideration of LNA and mixer is introduced. By analyzing and improving these, a low power, high linearity front end circuit are design and verified.

In Chapter 4, a MGTR UWB LNA is analyzed and verified in standard 0.18um CMOS technology. The MGTR technique is adopted to improve the linearity of the common source amplifier in the broadband condition. A compact equivalent circuit using AC complex transconductance analysis truly gives the optimal design parameters. Measurement results show that the improvement of the linearity is more than 5dB without extra power consumption.

In Chapter 5, a low power, high linearity UWB front end circuit, intended to use in the receiver path of UWB system, is designed in standard 0.18um CMOS technology. It is composed of a low noise amplifier, an active balun, and a Gilbert cell mixer. The UWB LNA provides higher linearity by using MGTR technology, and the single-ended signal is transformed into a differential form by the active balun. Simulation results shows that the front-end circuit achieves conversion gain of 20.3~21.1dB, input return loss of -10dB, input third-order intercept point (IIP3) of -4.7~-5.8dBm, and power consuming only 10.4mW.

In this thesis, two chips, UWB LNA and receiver were designed and analyzed. These two circuits, fabricated in TSMC 0.18um RF CMOS technology, are proposed

for the application of UWB communication system. One is a low power, high linearity UWB LNA which is composed of two amplifier stages with MGTR linearity improvement circuit. The IMD3 of the distortion cancellation for the common source amplifier stage have been analyzed and designed. The measured voltage conversion gain is 8.3dB and the total power consumption of it is 8.26mW from a 1.5V power supply.

Another circuit is a UWB receiver front-end circuit whose topology is chosen to suit the application of low power consumption and high linearity. A broadband active balun converts single input signal into differential output. Besides, a broadband mixer is adopted and carefully designed. The simulation results show the voltage conversion gain is 20.3~21.1dB and the total power consumption is only 10.4 mw from a 1.5V supply.

6.2 Future Work

About the work, the linearity can be improved by minimized the distortion of the third order intermodulation in broadband condition. The MGTR technique is an effective way to cancel the $G_m''(\omega)$ of the device. But the second order intermodulation may become important to degrade the IIP2 of LNA in broadband condition, it is especially obvious when the $G_m''(\omega)$ is small. Therefore, we can keep the original performance, such as gain, noise figure, etc. and improve the linearity and fixing IIP2 at the same time. There should be some other methods to look for a better compromise. If this method is realized together with MGTR technique or used to cancel the $G_m'(\omega)$, the improvement of the linearity will be obvious.

References

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

[2] C.-T. Fu and C.-N. Kuo, “3~11-GHz CMOS UWB LNA using dual feedback for broadband matching,” *IEEE RFIC Symp. Dig.*, 2006, pp.67-70.

[3] 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.

[4] Takao Kihara, Toshimasa Matsuoka, and Kenji Taniguchi, “A 1.0 V, 2.5 mW, Transformer Noise-Canceling UWB CMOS LNA,” *IEEE RFIC Symp. Dig.*, 2008, pp.493-496.

[5] S. Tanka, F. Behbahani, and A. Abidi, “A linearization technique for CMOS RF power amplifier,” *in Symp. VLSI Circuit Dig. Tech. Paper*, 1997, pp.93-94.

[6] Tae Wook Kim, Bonkee Kim, and Kwyro Lee, “Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors,” *IEEE Journal of Solid-State Circuit*, vol. 39, no. 1, pp. 223-229, January 2004.

[7] Wei-Chia Zhan, Chien-Nan Kuo, and Jyh-Chyurn Guo, “Low-Power and High-Linearity Mixer Design Using Complex Transconductance Equivalent Circuit,” *IEEE Asian Solid-State Circuits Conference,* 2008, pp. 365-368.

[8] Chang-Tsung Fu and Chien-Nan Kuo, "3~11-GHz CMOS UWB LNA Using Dual Feedback for Ultra-wideband Matching," *IEEE Radio Frequency Integrated Circuits Symp*, San Francisco, CA, June, 2006, pp. 53-56.

[9] Chang-Tsung Fu, Chun-Lin Ko, and Chien-Nan Kuo, “A 2.4 to 5.4 GHz Low Power CMOS Reconfigurable LNA for Multistandard Wireless Receiver,” *IEEE Radio Frequency Integrated Circuits Symp.*, Honolulu, Hawaii, June, 2007, pp. 65-68.

[10] Chang-Tsung Fu, Chun-Lin Ko, Chien-Nan Kuo, and Ying-Zong Juang, "A 2.4–5.4-GHz Wide Tuning-Range CMOS Reconfigurable Low-Noise

Amplifier ," *IEEE Trans. Microw. Theory Tech.*, vol. 56,no. 1, pp. 2754-2763, Dec. 2008.

[11] Tae Wook Kim, Bonkee Kim, Ilku Nam, Beomkyu Ko, and Kwyro Lee, "A low-power highly linear cascoded multiple-gated transistor CMOS RF amplifier with 10 dB IP3 improvement," *IEEE Microwave and Wireless Component Letters*, vol. 13, no. 6, pp. 205-207, September 2003.

[12] Jongsik Kim, Tae Wook Kim*, Minsu Jeong*, Boeun Kim*, and Hyunchol Shin, "A 2.4-GHz CMOS Driver Amplifier Based on Multiple-Gated Transistor and Resistive Source Degeneration for Mobile WiMAX," *IEEE Asian Solid-State Circuits Conference,* 2006, pp. 255-258.

[13] Bo Shi and Michael, Yan Wah Chia, "A CMOS Receiver Front-End for 3.1-10.6 GHz Ultra-Wideband Radio," *Proceedings of the 3rd European Radar Conference*, 2006, pp. 350-353.

[14] Zi-Hao Hsiung, Hui-I Wu, and Christina F Jou, "A 3-8 GHz RF Receiver Front-End for Multi-Band UWB Wireless System," *Proceedings of Asia-Pacific Microwave Conference*, 2007, pp. 1-4.

[15] Stephan Blaakmeer, Eric Klumperink, Domine Leenaerts, Brain Nautal, "A Wideband Balun LNA I/Q-Mixer combination in 65nm CMOS," *IEEE International Solid-State Circuits Conference*, 2008, pp. 326-617.

[16] Yen-Horng Chen, Chih-Wei Wang, Ching-Feng Lee, Tzu-Yi Yang, Chih-Fan Liao, Gin-Kou Ma,Shen-Iuan Liu, "A 0.18μm CMOS Receiver for 3.1 to 10.6GHz MB-OFDM UWB Communication systems," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, 2006, pp. 261-264.

[17] Shuzuo Lou, Hui Zheng and Howard C. Luong, "A 1.5-V CMOS Receiver Front-End for 9-Band MBOFDM UWB System," *IEEE Custom Intergrated Circuits Conference (CICC)*, 2006, pp. 804-804.

[18] Behzad Razavi, Turgut Aytur, Christopher Lam, Fei-Ran Yang, Kuang-Yu Li, Ran-Hong Yan, Han-Chang Kang, Cheng-Chung Hsu and Chao-Cheng Lee, "A UWB CMOS Transceiver," *IEEE Journal of Solid-State Circuit,* vol. 40, no. 12, pp. 2555-2562, Dec 2005.

[19] A. Ismail, A. Abidi, ”A 3.1 to 8.2GHz Direct Conversion Receiver for MB-OFDM UWB Communications,” *IEEE International Solid-State Circuits Conference*, 2005, pp. 208-593.

[20] S. Lee, J. Bergervoet, K.S. Harish, D. Leenaerts, R. Roovers, R. van de Beek, G. van der Weide, ”A Broadband Receive Chain in 65nm CMOS,” *IEEE International Solid-State Circuits Conference*, 2007, pp. 418-612.

[21] Bo Shi, Michael Yan Wah Chia, ”An Ultra-Wideband CMOS Receiver Front-End,” *IEEE Proceedings of the 37th European Microwave Conference (EuMA)*, 2007, pp. 284-287.

Vita

Chuan-Hao Lai

Birthday: 1984/09/14

Birthplace: Taipei, Taiwan

Education:

2003/09 ~ 2007/06

B.S. Degree in Department of Mechatronics Engineering, National Changhua University of Education

2007/09 ~ 2009/08

M.S. Degree in Department of Electronics Engineering & Institute of Electronics, National Chiao Tung University