Designs of Broadband Highly Linear CMOS LNAs for Multiradio Multimode Applications

Wei-Hung Chen

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2009-112
http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-112.html

August 10, 2009

Designs of Broadband Highly Linear CMOS LNAs for Multiradio Multimode Applications

by

Wei-Hung Chen

B.S. in Electrical Engineering, National Taiwan University, 1998
M.S. in Electrical Engineering, National Taiwan University, 2000

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Electrical Engineering

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Ali M. Niknejad, Chair
Professor Robert Meyer
Professor Paul Wright

Fall 2009

The dissertation of Wei-Hung Chen is approved:

Chair Date

Date

Date

University of California, Berkeley

Fall 2009

Designs of Broadband Highly Linear CMOS LNAs for Multiradio Multimode Applications

Wei-Hung Chen

Abstract

Designs of Broadband Highly Linear CMOS LNAs for Multiradio Multimode Applications

by

Wei-Hung Chen

Doctor of Philosophy in Electrical Engineering

University of California, Berkeley

Professor Ali M. Niknejad, Chair

With the proliferation of wireless communications, there emerges a trend towards integrating multiple wireless functionalities into one mobile device. Recently we have been observing a paradigm shift in the integrated wireless transceiver design where several narrow-band receivers tailored for dedicated applications, e.g. cellular and wireless LAN (Local Area Network), are replaced by one single circuit which is reconfigured to support different radio standards, the so-called "Universal Receiver". The front-end circuits of the universal receiver therefore have to be able to accommodate operations across a wide range of frequency bands, and different performance requirements for low noise and high linearity. Realizing such universal front-ends, e.g. Low-Noise Amplifier (LNA), with a broadband circuit appears attractive because of the reduced cost realized by area reduction due to sharing the core circuits and package pins. In a "digital" receiver architecture, most of the

signal processing is replaced by high-speed samplers and digital circuits, but a broadband LNA is still indispensable in order to relax the stringent performance requirements of the subsequent samplers. In this scenario we are actually running into the classic performance trade-off of analog circuits design among noise, linearity, and broadband impedance match. This research examines the issues associated with the implementation of conventional broadband LNAs, and presents design techniques for combined noise and distortion cancelation to achieve simultaneous low noise and high linearity. Detailed analysis are conducted, and verified experimentally with two integrated circuit prototypes fabricated in 0.13 μm and 65 nm CMOS technologies, respectively.

Professor Ali M. Niknejad
Dissertation Committee Chair

To My Parents

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to express my profound gratitude to my advisor, Professor Ali M. Niknejad, for the support and guidance he has given me throughout my years in Berkeley. I would also like to thank the other members of my dissertation committee, Professor Robert Meyer and Professor Paul Wright, for taking the time to review this dissertation. I am also grateful to Professors Robert Brodersen, Chenming Hu and Elad Alon for their helpful feedback and suggestions at various stages of this dissertation research.

It has been a privilege and an extraordinary pleasure to work in the company of many wonderful colleagues at the Berkeley Wireless Research Center. I have always found myself inspired during technical discussions with them. Even a random conversation over the lunch or happy hour is often delightful and thought-provoking. In particular, I would like to thank Gang Liu, Stanley Bo-Ting Wang, Mohan Dunga, Nuntachai Poobuapheun, Zhiming Deng, Bagher Afshar, Ehsan Adabi, Cristian Marcu, Darsen Lu, Ashkan Borna, Zhengya Zhang, Renaldi Winoto, Stanley Yuan-Shih Chen, Patrick Reynaert and Peter Haldi.

Sincere thanks also go to the dear staff members of BWRC and the EECS department. Without their assistance, my research and graduate study would have not been completed e ciently. Among them, Tom Boot is consistently the first person to greet me in the morning by resetting the alarm panel after my occasional overnight stay in the lab. In Cory hall, Ruth Gjerde is always kindly available for my various questions regarding program requirements and procedures.

In retrospect, I regard my experiences at UC Berkeley a true blessing. I would not have sailed through with a peace of mind without the company of many good friends. George and Chiu-Yu, thank you for being excellent house mates, and for being adventurous enough to appreciate my home coffee roasting. My two cousins, Hans and Jung-Sheng, I am truly thankful for being surrounded by you and your family in the Bay Area. Klaus and Linna, I would like to give a special thank to you for the genuine care I have been receiving from you and the Chou family. I owe a most sincere gratitude to my long-time mentor, I-Yun Sin, for his patience in listening and responding to my musings and contemplations since my early years.

Last but not least, I am deeply indebted to my family for their love, support and continued encouragement. Wen-shing, thank you for your love, and for bringing me the "clear and cool" realm of Manjusri's "Wutai Shan". I look forward to embarking on the next journey with you.

Chapter 1

Introduction

Wireless communications have proliferated and penetrated into daily life as a result of decades of continuous advancements in communications and semiconductor technologies. A wide variety of applications, ranging from Global Positioning System (GPS), cellular communications, Wi-Fi Local Area Network, and short-range personal communications such as Bluetooth (BT), have been commercially deployed and continue to evolve. On the emerging horizon, mobile terminals no longer limit their usage to a single purpose, but serve to provide a multitude of access to heterogeneous networks over which rich service contents are delivered by concurrent or switchable operation of/among different link communications [6]. It is not difficult to imagine a scenario in which a wireless mobile gadget is used around the globe to receive emails, search the Internet, watch hi-definition video streaming over the air, send pictures to a remote printer, track geographic position and receive route/location recommendation from www resources, just to name a few. This quest for ubiquitous wireless connectivity and the trend toward a highly integrated solution have opened up a new

big wave of challenges and opportunities for radio integrated circuits designs, as will be explained next in the introduction.

1.1 Towards Multistandard Multimode Mobile Terminals

The realization of multiple radio functionalities in a single mobile terminal is best envisioned by the increasing popularity of the so-called "Smartphone" in recent marketplace. Today's prevalent models, such as Nokia's N- series and Apple's iPhone 3G, are aimed to operate across 7 WEDGE frequency bands, i.e. quadband GSM (850/900/1800/1900 MHz) and triband UMTS (850/1900/2100 MHz), with extended support for HSDPA [7], and to allow simultaneous digital FM radio, Bluetooth, WiFi (802.11 b/g) and GPS connectivities. Some models also come equipped with support for mobile TV standards such as Digital Video Broadcasting-Handheld (DVB-H) [8]. Integration with new standards such as Long-Term Evolution (LTE) [9] and Worldwide Interoperability for Microwave Access (WiMAX) [10] are developing at a steady pace and commercial realization is expected by 2010 to 2012. Fig. 1.1 shows the projected growth of sale of Smartphones in units. It is estimated that by 2013 the demand will grow by a factor of five from the levels found between 2005 and 2007 [1]. Along with this trend, impressive momentum of growth is also ramping up for other new radio platforms, such as Mobile Internet Devices (MID)[1]. The annual shipments of MIDs will jump from 0.3 million in 2008 to 40 millions by 2012 according to [11].

[1] MIDs are a new class of emerging low-power mobile lifestyle devices. The form factor of MIDs fall in between a Smartphone and notebook/tablet computer. The inclusion of 3GPP, WiMAX, WLAN+BT+FM, GPS and Mobile TV in MIDs is roadmapped by 2012.

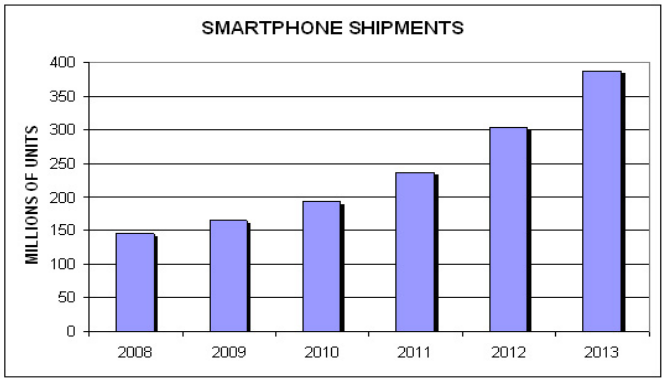


Figure 1.1: Projected sales of Smartphones [1]

In order to maximize the level of integration with a small form factor, sharing of hardware resources between multiple radio systems becomes necessary. Fig. 1.2 shows a visualization of today's state-of-the-art implementation [2]. Notice that the digital baseband and application processors are not included in this picture and the following discussion will be focused on the platform's analog/radio interfaces. The multiband 2G and 3G radios are integrated in a single chip, while several other ICs are deployed for a handful of applications. Taking a closer look at the 2G/3G portion, Fig. 1.3 reveals that multiple receiving paths are placed on the single die, with a fair number of external passive devices such as SAW filters and switches scattering on the PC board. Passive devices can be integrated with the silicon die using 3-D stack-up packaging to save the board area [2]. However, the cost of System-In-Package (SiP) is still prohibitively high for wide adoption.

[2]Common practice is two-chip solution, i.e. one front-end module for passives and power amplifiers, and the other is System-on-the-Chip (SoC)

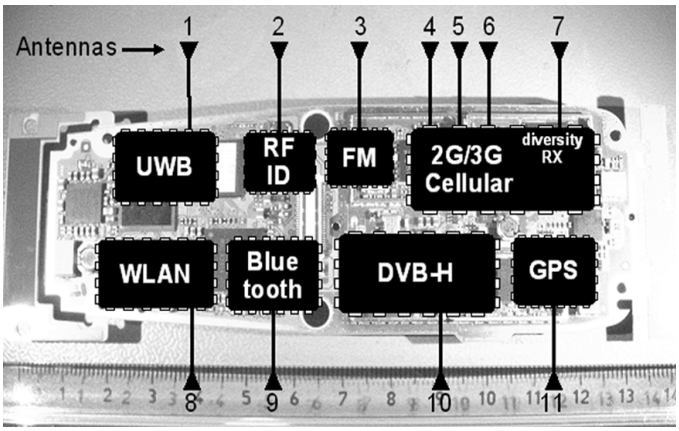


Figure 1.2: Visualization of a state-of-the-art multi-radio platform [2]

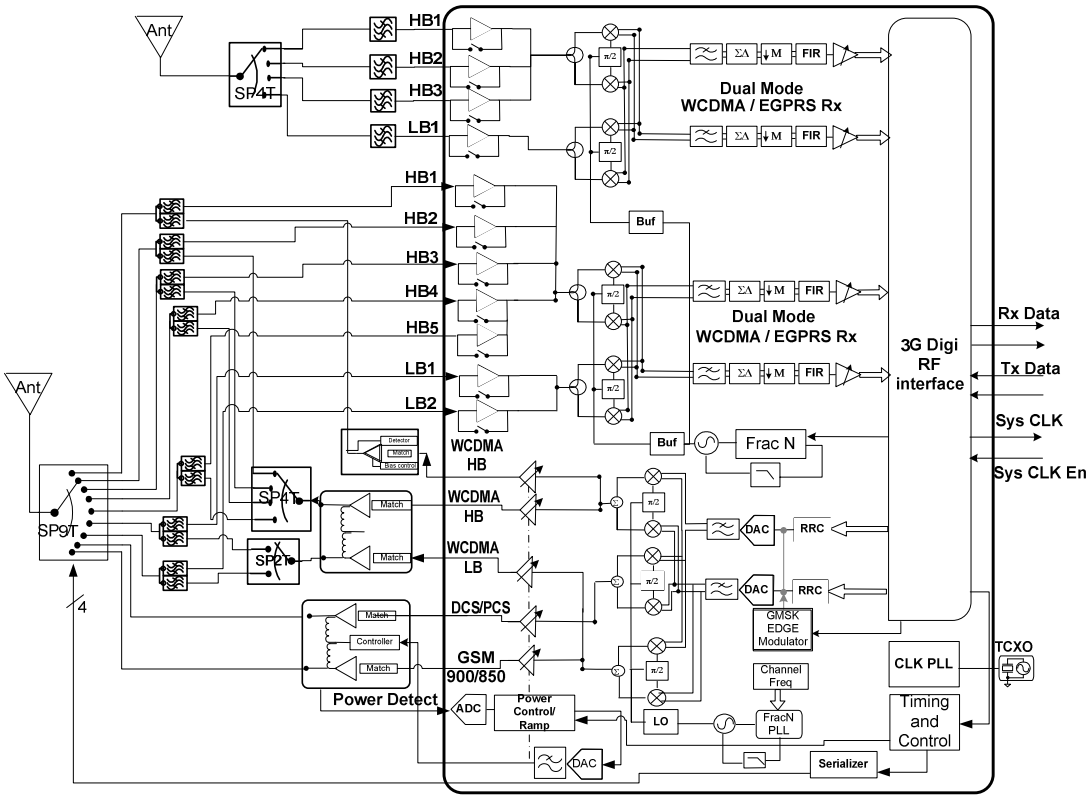


Figure 1.3: Block diagram of a WEDGE transceiver by Skyworks [3]

Another direction of reducing the number of ICs is currently taking place by integrating into a single chip WiFi, GPS, Bluetooth and FM radios [3]. Further increasing the level of radio integration into one single chip helps reduce the platform size and the overall Bill-of-Materials. This leads to a very desirable prospect for "Universal Radio", a single transceiver whose performances can be reconfigured and tailored for a plethora of different radio functionalities.

1.2 Universal Radio Receivers

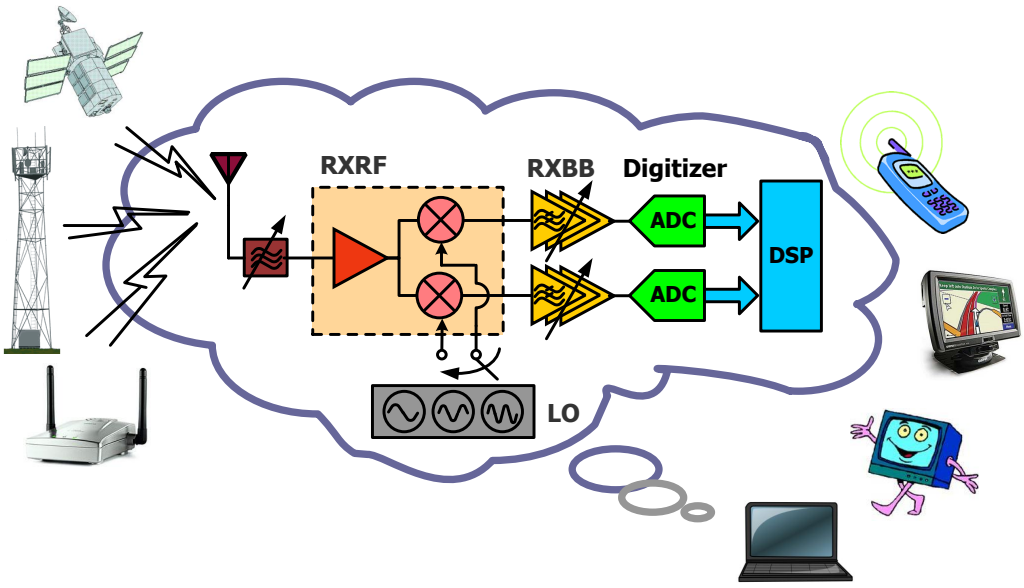


Figure 1.4: Illustration of an universal radio for multiple standards

The cartoon diagram in Fig. 1.4 illustrates the concept of an universal radio receiver. The direct-conversion architecture is preferred because it allows least circuit topology

[3] latest CSR9000 model from CSR. It is estimated that combo chips will account for nearly 60% of connectivity solutions that ship into mobile phones by 2012 [12].

change and agile frequency adaptation. Unlike Fig. 1.3 where each circuit building block is optimized and dedicated for one specific standard, the building circuits of the universal radio have to work across a wide frequency range and, at the same time, meet identical, if not more stringent, performance specifications required by each individual standard [13][14]. For instance, in the LNA, this means multiple band input impedance matching and low noise figure [4].

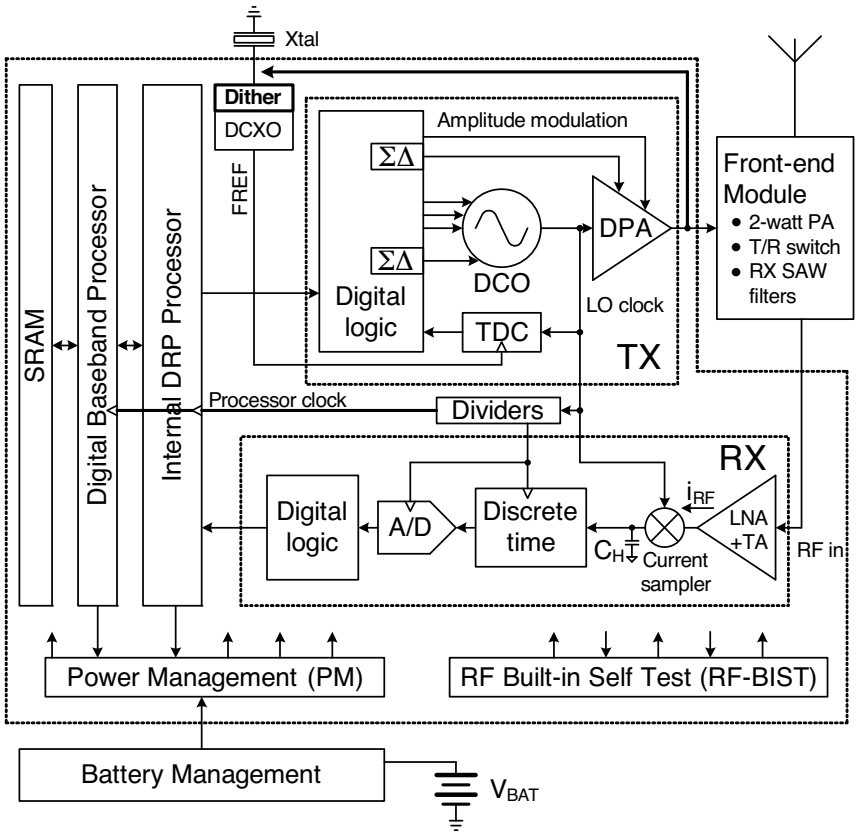


Figure 1.5: Digital RF transceiver architecture from Texas Instrument

[4] or a single broadband operation

Contrast to the conventional homodyne receivers, another architecture candidate, referred as "Digital RF", moves the task of signal conditioning, such as frequency conversion and filtering, from analog to discrete-time domain by employing a series of high-speed samplers and digital signal process techniques [15, 16, 17]. A block diagram example of a single-chip digial-RF GPRS/Blueetppth/FM transceiver is shown in Fig. 1.5. In this implementation, most of analog circuits in conventional homodyne receivers are replaced by their digital counterparts. This architecture favors fast switching nano-scale transistors and benefits by the continued technology scaling. However, due to noise folding caused by discrete-time sampling, a low noise amplifying stage has to precede the samplers in order to suppress the impact of sampled noise on the receiver sensitivity. Therefore the Low Noise Amplifier (LNA) remains indispensable for both architectures.

1.3 Blocker Considerations

Blockers are usually the strong transmitting signals from unintended uplink communications that the LNA picks up along with the desired receiving signal. The amplitude of these interference signals is usually order-of-magnitude larger than the desired signal. If not appropriately attenuated, they will dominate the LNA response and block the healthy receiving of the intended signal. The impacts of blockers include desensitizing the receiver sensitivity, compressing amplifier gain and eventually driving circuits into saturation, increasing signal distortion and reciprocally mixing with LO phase noise, all of which deteriorates link performances such as receiver bit error rate [18]. Standalone radio mitigates the blocker issue by placing high-Q filters in front of the LNA to reject the unwanted signal

energy [5]. As illustrated in Fig. 1.6, the proximity of multiple transmitting paths in a dense multiradio implementation inherently increases the chance and the complexity of blocking events. A particular case occurs when GPS functionality is incorporated with a few other

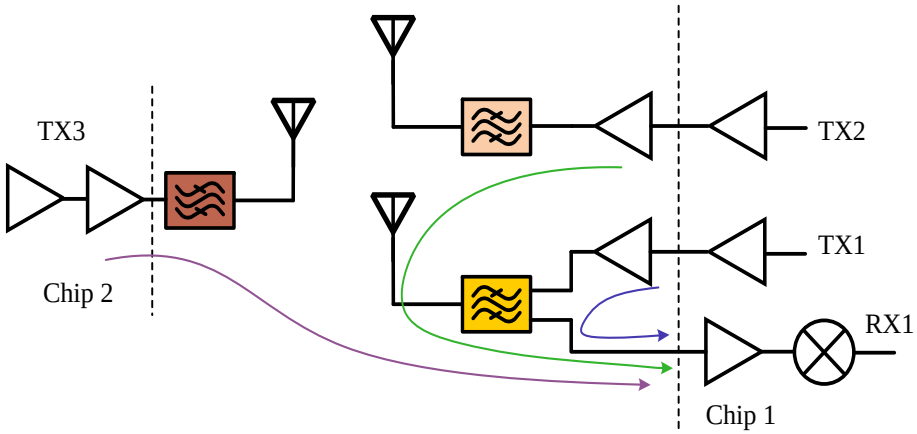


Figure 1.6: Visualization of blocker

radios. There is no LNA linearity specification for standalone GPS because only one receiving channel is expected from the satellite and a high-Q filter easily rejects any undesired interferences. If, however, situated on a platform where both WiFi and PCS transmitters are active and pumping out strong signals nearby, the residual interference signal appearing at the LNA input is still significant and incurs LNA linearity concerns [6].

The blocker scenario exacerbates as the trend of including more radio features and utilizing more frequency bands continues. A tunable RF filter, as seen in Fig. 1.4, is best suited for dynamically reconfigurable blocker rejection [19, 20]. However, existing solutions rely on MEMS technology and needs to solve practical issues such as minimizing the extra

[5] in superheterodyne receivers, additional filters are placed between the LNA and Mixer
[6] for example, IIP_3 of $+14\ dBm$, see chapter 2.

fabrication cost and improving the yield. Fig. 1.7 shows the diagram of a more promising blocker removal technique. The technique first duplicates the incoming signals, both desired and the interfering, in an auxiliary path parallel to the main LNA, down converts them to a low frequency and filter out the blocker, then subtract the blockers from at the LNA output [4, 21]. Moving around the frequency of the down mixing in the auxiliary path, the filtering null is adjusted to different blocking profiles. Up to 35 dB rejection can be achieved at the input of the mixers. This greatly relaxes the linearity requirement of the subsequent circuits. However, the front-end LNA is still bombarded by the large amplitude of the blockers and subject to stringent linearity requirement. For ″digital RF″ receivers, the filter response depends on the coefficient values of the digital filters but the same concern remains as the LNA still confronts the unattenuated blocker signals. In fact, as far as blockers are concerned, it is the front-end amplifier that sets both the noise and linearity constrain for the receiver [22].

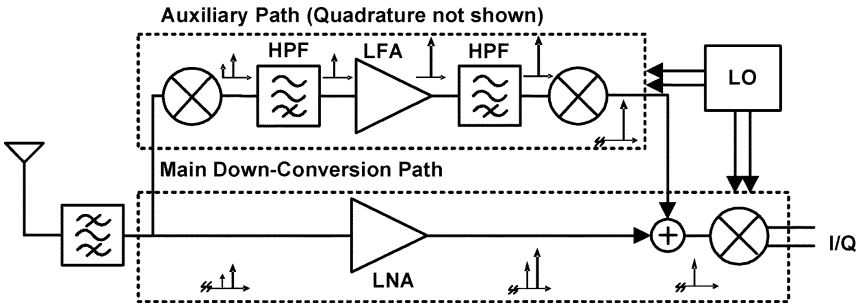


Figure 1.7: Diagram of active blocker removal [4]

1.4 Research Goal and Thesis Organization

This research focuses on the design of broadband highly linear LNAs suitable for use in universal radios. Unlike narrow-band LNA whose design methodology has been well explored in the literature [23, 24], the classic tradeoff among low noise, high linearity and broadband impedance matching has to be well addressed in the new LNA design. Moreover, the uniquely stringent linearity requirement in a multiradio scenario calls for a special attention to broadband amplifier linearization in this research. Because of application similarity, the techniques developed in this research are also applicable for broadband communication systems such as broadband cable tuner or mobile TVs.

This thesis is organized as follows. Chapter 2 begins with a review on LNA key performance metrics, along with the consideration of its implication to the system performance. Chapter 3 introduces the broadband noise cancelation technique and investigates its noise-power tradeoff. Chapter 4 switches the focus to the design of distortion cancelation for the proposed LNA circuit. The measurement results of a $0.8\sim 2.1GHz$ prototype LNA will be presented in chapter 5 . Chapter 6 reviews the deficiency of the existing distortion characterization method and presents an alternative approach. Finally, a modified noise and distortion canceling LNA prototype aiming to solve the outstanding issues left by the first prototype implementation will be presented in chapter 7. In the end, chapter 8 draws conclusions of this research and discussions for future research.

Chapter 2

LNA Fundamentals

Key LNA performances include input/output impedance matching, noise figure and linearity. The significance of these metrics is reviewed in this chapter with circuit and system examples.

2.1 Impedance Matching

Referring to Fig. 1.4, Low-Noise Amplifier (LNA) is the first receiver circuit on the chip to accept and to amplify the incoming signal from the off-chip antenna. The signal has to route along the trace on the PC board, through components such as filter or duplexer, and via the package pin before it reaches the LNA input. Because of the substantial distance by wavelength it has to travel, the signal energy may bounce back to the antenna if the LNA input impedance does not match with antenna and transmission line's characteristic impedance. The signal energy needs to be received by the LNA for further amplification and processing before it reaches the ADC. In some architectures where LNA output needs

to route out of the chip[1], the LNA output impedance also needs to be matched.

Impedance matching is characterized by scattering parameter, S_{11}, by

$$S_{11} = \frac{Z_{in} - Z_0}{Z_{in} + Z_0} \tag{2.1}$$

and specified in dB. Z_0 is the transmission line characteristic impedance, typically 50 Ω. The ratio of the input signal power and the power absorbed by the LNA is related to S_{11} by

$$\frac{\text{LNA input power}}{\text{available power from the signal source}} = 1 - S_{11}^2 \tag{2.2}$$

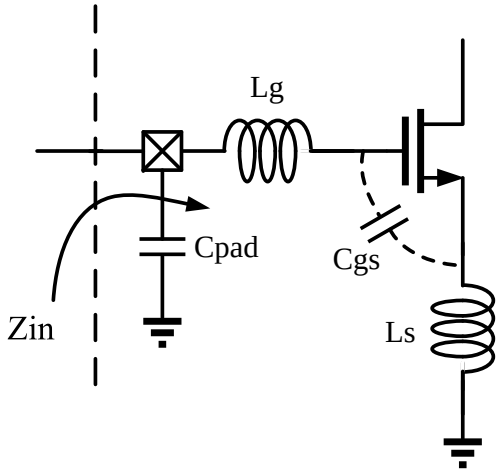
In practice, S_{11} of less than -10 dB, i.e. $> 90\%$ signal energy absorption, is considered acceptable. Impedance matching is LNA's primary requirement because of its preceding role in the receiver chain. This imposes a constraint on the available amplifier topologies, and distinguishes front-end amplifier design from generic amplifier designs.

A commonly-used LNA schematic for narrow-band application is shown in Fig. 2.1(a). From [23] the input impedance ignoring pad parasitics is

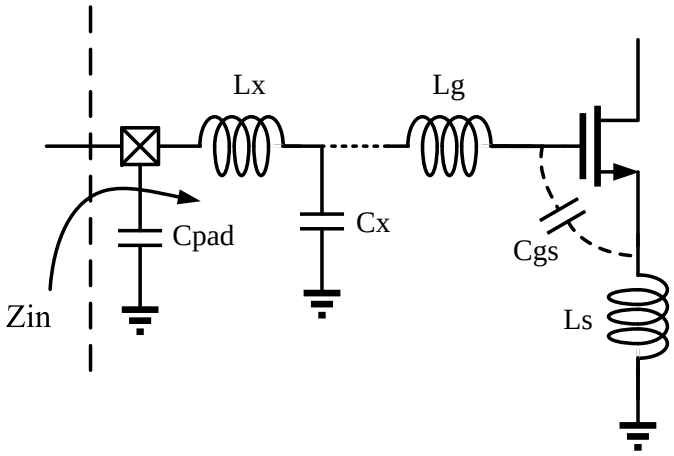
$$Z_{in} = sL_g + \frac{1}{sC_{gs}} + sL_s + \frac{g_m}{C_{gs}} \times L_s \tag{2.3}$$

Although no resistive device in the circuit, a real impedance forms due to a 90^o phase lag of the gate current compensated by a 90^o phase lead of the voltage across L_s by g_m current. g_m, thus C_{gs}, are usually predominated by the amplifier gain or power consumption requirement. The source inductor, L_s, is then chosen to match this real-valued impedance to source resistance R_s. The gate inductor, L_g, is to tune out the residual imaginary

[1] Interstage off-chip filtering in super heterodyne receivers.



(a) single stage



(b) multiple stage

Figure 2.1: Impedance matching with source degeneration inductor

impedance left by C_{gs} and L_s at resonant frequency ω_0.

$$\omega_0 = \frac{1}{\sqrt{C_{gs}(L_s + L_g)}} \tag{2.4}$$

The quality factor of the series RLC network is

$$Q = \frac{1}{2R_s\omega_0 C_{gs}} \tag{2.5}$$

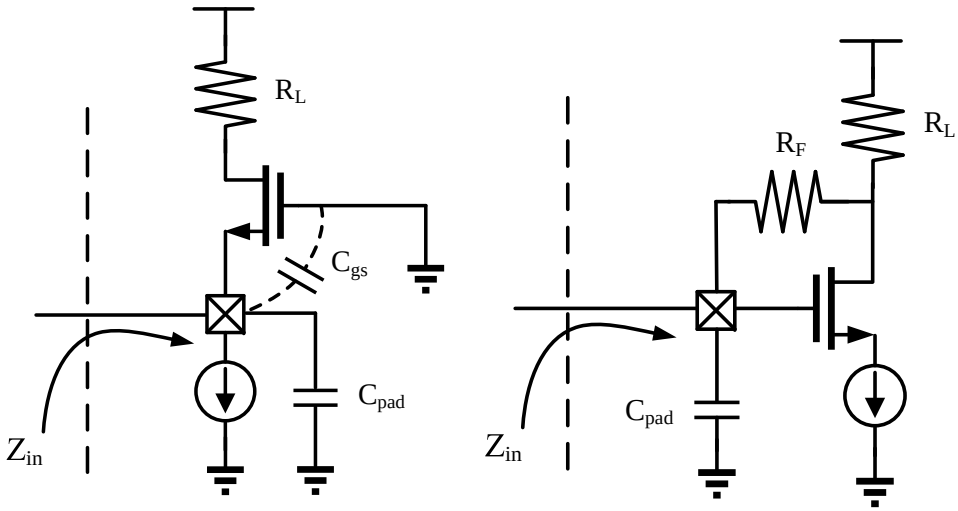
, and the matching bandwidth is inversely related to Q by

$$BW = \frac{\omega_0}{2Q} \tag{2.6}$$

Given $R_s = 50\ \Omega$ and f_0 in the GHz range, Q of Eq. 2.5 is greater than 1 for practical C_{gs} values, and the bandwidth is limited.

To accommodate multi-band operation, more than one narrow-band LNA is usually placed in parallel, as previously seen in Fig. 1.3. This is a straightforward implementation but at cost of die area, due to duplicity of similar circuits, and pin counts. In addition, the number of the inductors also multiplies, leading to even greater area consumption. In an analogy to filter synthesis, the bandwidth of the resistance conversion from $\frac{g_m}{C_{gs}} \times L_s$ in Eq. 2.3 to R_s is in inverse proportion to the network Q, or, in proportion to the number of the conversion stages [25]. As shown in Fig. 2.1(b), a much wider bandwidth is realized by inserting more LC stages. An Ultra Wideband (UWB) performance, say from 3.1 to 10.6 GHz is achieved in [26] with two more inductors. This implementation reduces the number of pins, but the number of inductors remains high.

Contrary to the area overhead of inductors, inductorless broadband LNA configuration achieves very compact silicon implementation and receives a high interest recently



(a) common-gate (b) shunt-shunt feedback

Figure 2.2: Broadband inductorless LNA topology

[27, 28]. This is partly attribute to the availability of high f_T transistors, as a result of continual technology scaling, so that the need of using inductors to compensate for capacitive parasitics for bandwidth extension abates. Two well-known broadband topologies are common-gate and feedback amplifiers. Fig. 2.2(a) and Fig. 2.2(b) show the schematics, respectively. It is straightforward to show the input impedance of both circuits without parasitics is

$$\begin{aligned} Z_{in,CG} &= \frac{r_o + R_L}{1 + g_m r_o} \approx \frac{1}{g_m} \quad \text{and} \\ Z_{in,FB} &= \frac{R_F + R_L \parallel r_o}{1 + g_m \times (R_L \parallel r_o)} \approx \frac{R_F}{1 + g_m \times (R_L \parallel r_o)} \end{aligned} \tag{2.7}$$

, where r_o is the drain output resistance. Both input impedances can be easily matched to source resistance R_s by varying g_m and the resistance values. Even at a high frequency when the effect of parasitic capacitance kicks in, the bandwidth of the common-gate configuration

is improved by including the existing bond wire inductor to form a low Q network.

$$Q = \frac{R_s \omega C_{gs}}{2} < 1 \tag{2.8}$$

The Q value is smaller than that of Eq. 2.5, indicating its broad bandwidth. For shunt-shunt amplifier, the bandwidth is readily extended by the nature of feedback configuration.

Circuit examples in Fig. 2.1 and Fig. 2.2 can be combined to derive versatile broadband topologies. For instance, common-gate + shunt- shunt feedback [29, 30], common-source degeneration + shunt-shunt feedback [31], and common-source degeneration + transformer feedback [32] all appear. The down side of the common-gate and shunt-shunt feedback amplifiers is their inferior noise performance, as will be discussed shortly.

2.2 Noise Figure

Noise factor, F, is the metric to evaluate Signal-to-Noise Ratio (SNR) degradation by a noisy circuit after the signal passes through it.

$$F = \frac{SNR_{in}}{SNR_{out}} \tag{2.9}$$

In practice, noise figure, NF, the decibel value of noise factor, is more commonly used. An alternative expression more transparent to circuits design is by distinguishing the source output noise from the circuit's own noise, and is

$$\begin{aligned} F &= \frac{\text{total output noise}}{\text{output noise due to source}} \\ &= 1 + \frac{\text{output noise due to the circuit}}{\text{output noise due to source}} \end{aligned} \tag{2.10}$$

In a cascade system where each building block not only amplifies the incoming signal, but also contributes noise of its own circuit, the system noise figure is described by the well-known Friis formula,

$$F_{all} = 1 + F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + ... + \frac{F_i - 1}{\prod_{j=1}^{i-1} G_j} \tag{2.11}$$

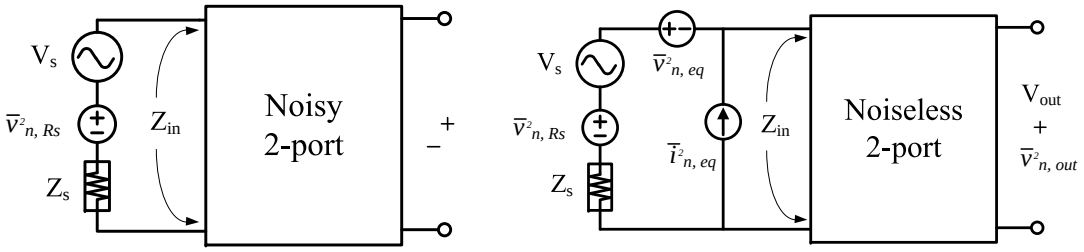
, where F_i, G_i are the noise factor and gain, respectively, of the ith stage down in the chain [33]. As Eq. 2.11 suggests, the LNA dominates, if not overwhelms, the receiver noise figure by dividing the noise of the subsequent stages with LNA gain.

Receiver sensitivity defines the minimum input signal that can be discerned in the backdrop of a chain of noisy circuits. Based on the definition of Eq. 2.9,

$$P_{sens.} = SNR_{R_x} + P_{n,s} + NF_{R_x} \tag{2.12}$$

SNR_{R_x} is the signal-to-noise ratio at receiver output. $P_{n,s}$ and NF_{R_x} are the source noise power and the receiver noise figure, respectively, referred to antenna, all in dB or dBm. The receiver sensitivity and output SNR are metrics closely related to link performances such as bit rate and communication distance. With higher SNR, for instance, the receiver can support the use of more complex signal modulation schemes and boost the data rate. With lower sensitivity level, the furthest distance that the receiver can operate enhances. Therefore the receiver, thus LNA, noise figure needs to be as low as possible.

Fig. 2.3 illustrates the way to represent the noise behavior for any 2-ports. An input-referred noise voltage source and an input-referred noise current source suffice to characterize 2-port's noise behavior at any source impedance [34]. The noise factor is found



(a) noisy 2-port (b) noiseless 2-port with input referred noise sources

Figure 2.3: Two-port noise representation

as

$$F = 1 + \frac{\bar{v}^2_{n,eq} + \bar{i}^2_{n,eq} \times (Z_s \parallel Z_{in})}{\bar{v}^2_{n,R_s}} \tag{2.13}$$

, where R_s is the real part of the source impedance Z_s. It shows that noise figure is affected by the noise voltage $\bar{v}^2_{n,eq}$, and the split noise current $\bar{i}^2_{n,eq}$ into Z_s. The noise figure depends both on the ratio between Z_s and Z_{in}, and the 2-port's own noise properties. This is different from input impedance match, which depends only on the ratio of Z_s and Z_{in}. In other words, the optimal source impedance for noise match does not necessarily correspond to power match at the same time.

Specify the noise sources in Fig. 2.3 by their equivalent noise resistance or conductance.

$$\begin{aligned} \bar{v}^2_{n,eq} &= R_n \times 4KT\Delta f \\ \bar{i}^2_{n,eq} &= \frac{1}{G_u} \times 4KT\Delta f \\ \bar{v}^2_{n,R_s} &= R_s \times 4KT\Delta f \end{aligned} \tag{2.14}$$

It can be shown that the noise figure can be expressed as

$$F = F_{min} + \frac{R_n}{G_s} \times \left((G_s - G_{opt})^2 + (B_s - B_{opt})^2\right) \tag{2.15}$$

, where the minimum noise figure

$$F_{min} = 1 + 2R_n \times (G_{opt} + G_c) \tag{2.16}$$

corresponds to the noise figure at an optimum source admittance

$$\begin{aligned} Y_{s,opt} &= G_{opt} + jB_{opt} \quad \text{and} \\ B_{opt} &= -B_c \\ G_{opt} &= \sqrt{(\frac{G_u}{R_n} + G_c^2)} \end{aligned} \tag{2.17}$$

G_c and B_c are the correlated conductance/admittance between $\bar{v}_{n,eq}^2$ and $\bar{i}_{n,eq}^2$, respectively [35, 36].

By properly designing the impedance matching network, the inductor source-degenerated LNA typically achieves a noise figure around 2 dB, or less, and it improves with technology scaling due to smaller NF_{min} [37]. However, for the broadband common-gate and feedback circuits in Fig. 2.2, consider transistor's drain current thermal noise,

$$\bar{i}_{n,drain}^2 = 4KT\frac{\gamma}{\alpha}g_m \times \Delta f \tag{2.18}$$

, $K = 1.38 \times 10^{-23}$, γ is transistor thermal noise coefficient and $\alpha = \frac{g_m}{g_{d0}}$. g_{d0} is the drain conductance at zero drain voltage. α is roughly equal to 1 and γ is greater than 1 in short-channel transistors [38, 39]. The noise factor of both circuits due to $\bar{i}_{n,drain}^2$ alone is

$$F = 1 + \frac{\gamma}{\alpha}\frac{1}{g_m R_s} \tag{2.19}$$

Under the circumstance of input impedance match, i.e. $Z_{in} = R_s$, Eq. 2.19 is already larger than 3 dB. Factoring in other noises in the circuit, it is very common that such LNA has a noise figure exceeding 5 dB.

Table 2.2 summarizes typical receiver input sensitivity and noise figure for a few well-deployed communication standards. Notice the listed NF values in Table 2.2 refers to noise figure at the air interface of the receiver, i.e. the antenna. It includes the noise figure of the passive devices between the LNA and antenna[2], as well as noise of the remaining receiver circuits. Accounting for these extra NF degradations, a noise figure of 3 dB is expected for LNA in most applications [40, 41].

	WCDMA	WLAN[1]	GPS	WiMAX[2]	BlueTooth
$P_{sens.}$ (dBm)	-102	-65	-136	-65	-70
SNR (dB)	5.2	28	7	24	21
NF (dB)	9	7.5	2	7	23

[1] 802.11a @ 54Mb/s mode

[2] 64 QAM

Table 2.1: Sensitivity and noise figure comparison

2.3 Linearity

The incoming signal of a radio receiver may vary in magnitude by several orders as a result of moving away from or close to the signal transmitting source. Unlike noise

[2]Typical loss of off-chip components is between 1.5 to 3 dB

figure which translates to receiver's minimum detectable signal as shown in Eq. 2.12, the linearity performance determines the largest signal level that a receiver can accommodate without failing the Signal-to-Noise+Distortion ratio (SNDR).

2.3.1 Taylor Series Distortion Characterization

Assume the nonlinear relationship between the system's output and input is described by

$$\begin{aligned} S_{out} &= S_{out,DC} + \sum_{i=1}^{\infty} a_i \times (S_{in} - S_{in,DC})^i \quad \text{, where} \\ a_i &= \frac{1}{i!} \frac{d^i S_{out}}{dS_{in}^i}\bigg|_{S_{in,DC}} \end{aligned} \tag{2.20}$$

Disregarding the static terms, the AC signals, s_{in} and s_{out}, are expressed as

$$s_{out} = a_1 \times s_{in} + a_2 \times s_{in}^2 + a_3 \times s_{in}^3 + \text{Higher order terms } (HOTs). \tag{2.21}$$

For now, neglect higher order terms and assume a sinusoidal input,

$$s_{in} = A \sin(\omega_1 t) \tag{2.22}$$

Plugging it into Eq. 2.21, the output becomes

$$\begin{aligned} s_{out} &= a_1 A \sin(\omega_1 t) + a_2 A^2 \sin(\omega_1 t)^2 + a_3 A^3 \sin(\omega_1 t)^3 \\ &= (a_1 A + \frac{3}{4} a_3 A^3) \sin(\omega_1 t) + \frac{1}{2} a_2 A^2 \sin(2\omega_1 t) + \frac{1}{4} a_3 A^3 \sin(3\omega_1 t) \end{aligned} \tag{2.23}$$

The output magnitude at ω_1 deviates from the linear term $a_1 A$ by a small amount of $\frac{3}{4} a_3 A^3$ if A is small. The nonlinear coefficients produce additional terms at $2\omega_1$ and $3\omega_1$. Their relative magnitude to the fundamental output is referred to as the second- and third-order

harmonic distortion.

$$\begin{aligned} HD_2 &\approx \frac{\frac{1}{2}a_2A^2}{a_1A} = \frac{a_2}{2a_1}A \\ HD_3 &\approx \frac{\frac{1}{4}a_3A^3}{a_1A} = \frac{a_3}{4a_1}A^2 \end{aligned} \tag{2.24}$$

The harmonic distortion increases with input signal amplitude. Due to negative a_3 in most circuits, the output signal at ω_1 experiences gain compression as input magnitude increases. This is characterized by P_{1dB} that defines the input level with the signal gain 1 dB, or 11%, smaller than a_1.

$$\begin{aligned} a_1 + \frac{3a_3}{4}A^2 &= a_1 \times 10^{(\frac{1}{20})} \\ A_{P_{1dB}} &= \sqrt{\left|\frac{4a_1}{3a_3}\right|} \end{aligned} \tag{2.25}$$

For inputs of more than one frequency, say, two of ω_1 and ω_2,

$$s_{in} = A_1\sin(\omega_1 t) + A_2\sin(\omega_2 t) \tag{2.26}$$

The results of the output become

$$\begin{aligned} s_{out} =& (a_1A_1 + \frac{3}{4}a_3A_1^3 + \frac{3}{2}a_3A_1A_2^2)\sin(\omega_1 t) + (a_1A_2 + \frac{3}{4}a_3A_2^3 + \frac{3}{2}a_3A_1A_2^2)\sin(\omega_1 t) \\ &+\frac{1}{2}a_2A_1^2\sin(2\omega_1 t) + \frac{1}{2}a_2A_2^2\sin(2\omega_2 t) \\ &+\frac{1}{4}a_3A_1^3\sin(3\omega_1 t) + \frac{1}{4}a_3A_2^3\sin(3\omega_2 t) \\ &+\Big(a_2A_1A_2\sin(\omega_1 t + \omega_2 t) + a_2A_1A_2\sin(\omega_1 t - \omega_2 t) + \\ &\quad \frac{3}{4}a_3A_1^2A_2\sin(2\omega_1 t + \omega_2 t) + \frac{3}{4}a_3A_1^2A_2\sin(2\omega_1 t - \omega_2 t) + \\ &\quad \frac{3}{4}a_3A_2^2A_1\sin(2\omega_2 t + \omega_1 t) + \frac{3}{4}a_3A_2^2A_1\sin(2\omega_2 t - \omega_1 t)\Big) \end{aligned} \tag{2.27}$$

In addition to the harmonic distortions, the terms in the capital bracket contain distortions at different frequency combination of ω_1 and ω_2. They are the intermodulation distortion, defined by

$$\begin{aligned} IM_2 &= \frac{a_2 A_1 A_2}{a_1 A_1} \\ IM_3 &= \frac{\frac{3}{4} a_3 A_1^2 A_2}{a_1 A_1} \end{aligned} \tag{2.28}$$

In a 2-tone test, A_1 and A_2 are made equal. If ω_1 is close to ω_2, the intermodulation product at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are also near the fundamental tones and hard to distinguish and to eliminate in real systems due to finite filter Q. Fig. 2.4 shows a plot of output v.s. input signal power at both frequencies. On dB-dB scale and in the small signal regime, the curves feature a slope of 1 and 3, respectively, corresponding to the order of the Taylor coefficients. The curves eventually saturate due to gain compression. Input Interception Point (IIP) defines the input level at which the extrapolated curves intercept and the intermodulation product equals the fundamental output.

$$\begin{aligned} IIP_2 &= \frac{a_1}{a_2} \\ IIP_3 &= \sqrt{\left|\frac{4a_1}{3a_3}\right|} \end{aligned} \tag{2.29}$$

Comparing Eq. 2.29 and Eq. 2.25, IIP_3 is linked with P_{1dB} by

$$A_{P_{1dB}} = IIP_3 - 9.6\ dB \tag{2.30}$$

Moreover, it is shown that the intercept point of a cascaded system is

$$\frac{1}{IIP_x^2} \approx \frac{1}{IIP_{x,1}^2} + \frac{a_{1,1}^2}{IIP_{x,2}^2} + \frac{a_{1,1}^2 a_{1,2}^2}{IIP_{x,3}^2} + \ldots + \frac{\prod_{j=1}^{i-1} a_{1,j}^2}{IIP_{x,i}^2} \tag{2.31}$$

, where $IIP_{x,i}$ and $a_{1,i}$ are the input intercept point of the xth order, and the a_1 coefficient at the ith stage, respectively [33]. Contrary to the cascade noise figure, the linearity of

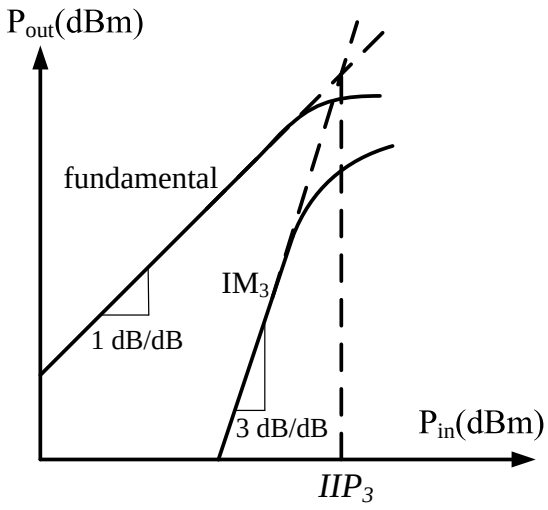


Figure 2.4: dBplot

the last stage dictates because the signal magnitude reaches its maximum[3]. Some radio systems employ non-constant-envelop signal modulation scheme to achieve higher data rate. The constant-envelop 2-tone IIP_3 test may not reflect the linearity requirement for such systems faithfully. However, it has been shown that linearity requirement of both schemes are statistically correlated and 2-tone IIP_3 still serves as a good metric [42, 43, 44, 45].

2.3.2 Example Scenario

Radio receivers specify IIP_2 and IIP_3 based on the worst interference scenario. For instance, GSM system specifies a $-49\ dBm$ continuous wave and a $-49\ dBm$ GMSK modulated blocker at 8 MHz and 1.6 MHz away from the receiving channel. The frequency of the IM_3 product is right on top of the receiving signal band. Based on the required SNR and minimum receiving signal, the blocker IM_3 has to be lower than $-111\ dBm$. This leads

[3]under the assumption that the signal causing the major distortions is not filtered down the chain.

to a receiver linearity requirement of [46]

$$IIP_{3,GSM} = -49 + 0.5 \times IM_3 = -18\ dBm. \tag{2.32}$$

In duplex systems such as WCDMA, the transmitter and receiver are separated by a duplexer and are active at the same time. Due to the finite isolation of the duplexer, a portion of the transmitter signal also appears at the LNA input. The frequency band of WCDMA TX and RX is $1920 - 1980\ MHz$ and $2110 - 2170\ MHz$, respectively. A PCS signal is midway between the TX and RX bands, and the IM_3 caused by a combination of PCS and TX leakage creates an unwanted distortion signal in the RX band. Consider PA maximum output of +28dBm and 55 dB isolation at the duplexer, the TX leakage shows up at LNA input as large as $-27\ dBm$. Assume maximum PCS blocker power of $-40\ dBm$ after the attenuation of antenna and duplexer, the equivalent input power for IIP_3 calculation becomes $(2 \times (-40) - 27)/3 = -35.7\ dBm$. In order to fulfill a WCDMA sensitivity level of -102 dBm,

$$IIP_{3,\mathrm{WCDMA}} = -35.7 + 0.5 \times (-35.7 + 102) = -2.55\ dBm. \tag{2.33}$$

The third example applies to a peculiar case where a GPS receiver is accompanied by a WiFi and a PCS transceiver, as is common in a multiradio platform. Assume the power of these two blockers is as large as $+30\ dBm$ and $+20\ dBm$ respectively. They experience a typical antenna isolation of 15 dB. The off-the-shelf GPS filter provides another 50 dB and 27 dB attenuation, respectively. The interferers at GPS LNA input will be $-35\ dBm$ and $-22\ dBm$. Given a $-120\ dBm$ sensitivity, this leads to a very high IIP_3 for the front-end.

$$IIP_{3,\text{GPS}} = -30.7 + 0.5 \times (-30.7 + 120) = +14\ dBm. \tag{2.34}$$

2.3.3 MOSFET linearity

The transconductor is one of the most fundamental building blocks in analog circuits. The linearity of a transconductor serves as a good indicator of scaled-CMOS linearity. In the saturation mode, the following simplified drain current expression is used

$$I_{ds}(V_{gs}) = \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}(V_{gs} - V_{th})^2 \times \frac{1}{1 + (\theta + \frac{\mu_0}{v_{sat}L}) \cdot (V_{gs} - V_{th})} \tag{2.35}$$

to estimate the transconductor nonlinearity [47]. θ and v_{sat} account for the main nonlinearity due to mobility reduction and velocity saturation, respectively. Apply Taylor series representation to Eq. 2.35,

$$\begin{aligned}
i_{ds} &= g_m \times v_{gs} + g'_m \times v_{gs}^2 + g''_m \times v_{gs}^3 \quad \text{, where} \\
g_m &= \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}(V_{gs} - V_{th}) \times \left(2 + (\theta + \frac{\mu_0}{v_{sat}L}) \times (V_{gs} - V_{th})\right) \\
g'_m &= \mu_0 C_{ox}\frac{W}{L} \times \frac{1}{2\left(2 + (\theta + \frac{\mu_0}{v_{sat}L}) \cdot (V_{gs} - V_{th})\right)^3} \\
g''_m &= -\mu_0 C_{ox}\frac{W}{L} \times \frac{\theta + \frac{\mu_0}{v_{sat}L}}{2\left(2 + (\theta + \frac{\mu_0}{v_{sat}L}) \cdot (V_{gs} - V_{th})\right)^4}
\end{aligned} \tag{2.36}$$

IIP_3 due to g_m nonlinearity is found to be

$$IIP_3 = \sqrt{8 \times \frac{\left(1 + (\theta + \frac{\mu_0}{v_{sat}L}) \cdot (V_{gs} - V_{th})\right)^2 \times \left(2 + (\theta + \frac{\mu_0}{v_{sat}L}) \cdot (V_{gs} - V_{th})\right) \times (V_{gs} - V_{th})}{\theta + \frac{\mu_0}{v_{sat}L}}} \tag{2.37}$$

Fig. 2.5 shows the results of Eq. 2.37 by filling in parameter values of different technologies from several foundries. Similar to the conclusion in [48], the IIP_3 does not change

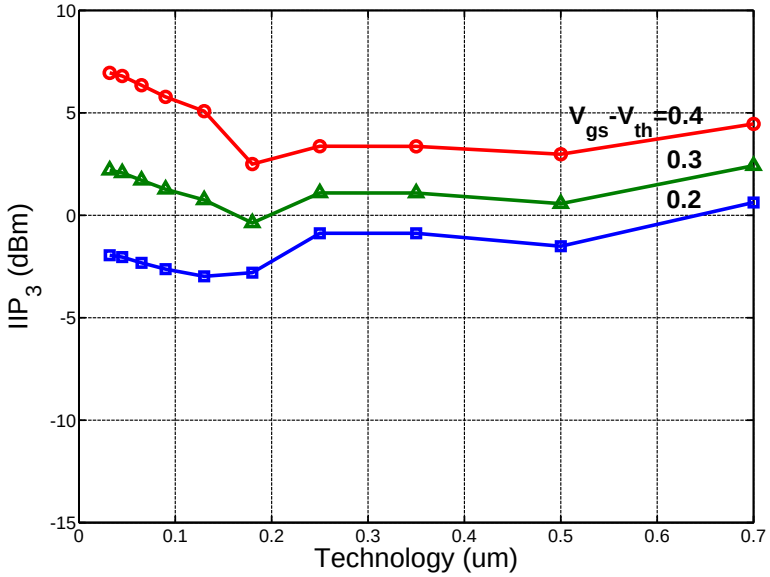


Figure 2.5: IIP_3 trend of scaled-CMOS

significantly across the technology nodes. But it shows strong dependence on the overdrive voltage, which is affected by technology scaling due to decreasing supply voltage.

Chapter 3

Noise Cancelation Design

The main issue concerning the broadband amplifier topologies in Fig. 2.2 is their inferior noise performance for radio applications. In [5], a noise cancelation technique was applied to a shunt-shunt feedback amplifier to reduce the amplifier noise figure below 3 dB, while maintaining broadband impedance matching from 2 to 1000 MHz and voltage gain of 20 dB. Later in [49] came the debut of a noise canceling front-end LNA in a "digital RF" receiver implementation. Despite the many publications of broadband noise-canceling amplifiers to date [50][51], a clear discussion and optimization of the scheme's noise-power tradeoff is not yet available in the literature. This chapter will take on this task and analyze the noise cancelation with the common-gate configuration in details [52].

3.1 Review of Bruccoleri's Thermal Noise Cancelation

Fig. 3.1 illustrates the original noise cancelation schematic proposed by Bruccoleri in [5]. The shunt-shunt feedback provides broadband impedance matching, but induces

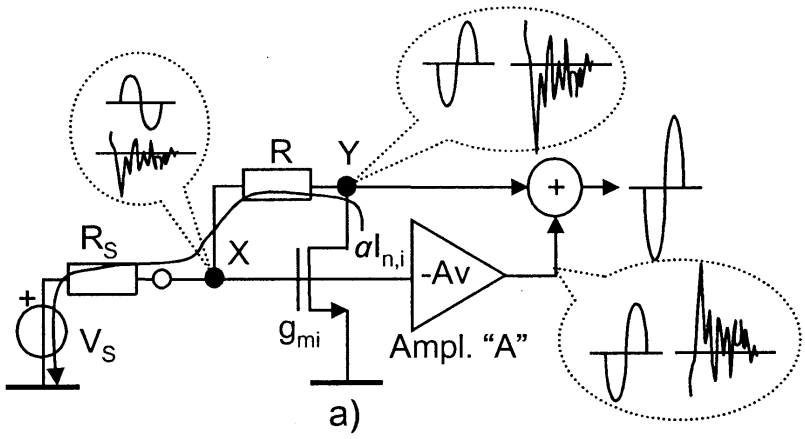


Figure 3.1: Bruccoleri's approach in [5]

high noise figure. Apply Kirchoff's current law (KCL), the circuit's response to the input signal and transistor's drain current noise can be understood as follows. First, the input signal, V_S, is amplified and inverted at node Y. Modeling the drain noise by a current source, $\overline{i}^2_{n,i}$, between its drain and source nodes, part of the noise current flows out of the transistor due to finite transistor r_o. This current embarks on a loop across resistors R and R_S, and returns to transistor from the grounded source node. Around the loop, the noise current establishes two voltages at node X and Y. These two voltages are fully correlated and have the same polarity. A second amplifier multiplies both the signal and noise voltages at node X with a gain of $-A$. The results are summed with voltages at node Y. The summation network is designed in such a way that for the desired input signal, the voltages add up at the output constructively. On the other hand, because of the negative gain of the second amplifier and the same noise voltage polarity at X and Y, the noise voltages are subtracted from each other at the summing node. By choosing the resistance and gain values to match the path gain for the concerned drain noise, a complete removal of $\overline{i}^2_{n,i}$ is achieved. Notice

the noise source in this model does not limit itself to drain current thermal noise only, but any nonideality that can be modeled as a current between the drain and source nodes. For instance, the drain distortion current is canceled in the same manner. At high frequencies, the parasitic capacitive impedances around the loop causes phase difference other than 180° and undermines the cancelation. However, the degradation of noise cancelation is less severe than the shrinkage of amplifier bandwidth because the phase change seen by input signal as it passes through intermediate nodes accumulates but the relative phase change of noise voltage between the nodes does not increase as much. As a rule of thumb, as long as the frequency is still within the amplifier 3 dB bandwidth, the noise cancelation remains effective.

3.2 CG-CS Noise canceling LNA

Now let us move on to the analysis of the proposed common-gate common-source circuit shown in Fig. 3.2. For clarity, only the small signal schematic is shown here. The principle of noise cancelation in this new circuit is explained as follows. The input signal, say, a current denoted by solid line, undergoes feed-forward voltage amplification by transistor M_3 and M_4. On the other hand, the channel current noise of transistor M_1, the dotted line, undergoes subtraction at output node due to two correlated but out-of-phase noise voltages at V_x and V_y. For completeness, Z_{in} is used for notation but bear in mind that only the resistive impedance is considered. The input impedance Z_{in} and voltage gain A_v are

$$Z_{in} = \frac{R_1 + r_{o1}}{1 + g_{m1} r_{o1}} \quad \text{, and} \tag{3.1}$$

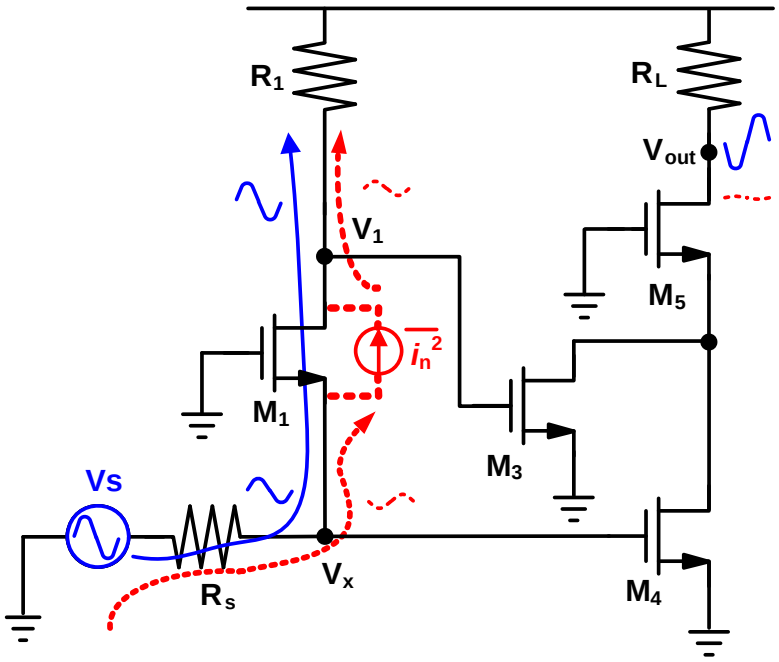


Figure 3.2: Simplified small-signal schematic of the proposed LNA.

$$A_v = \frac{V_{out}}{V_x} = \left(g_{m4} + g_{m3} \left(\frac{1 + g_{m1} r_{o1}}{1 + \frac{r_{o1}}{R_1}} \right) \right) \times R_L \tag{3.2}$$

In Eq. 3.2, the output conductance of transistors M_3 and M_4 are ignored due to the cascode transistor M_5, while it is important to consider the output conductance of M_1 as its value is low at deep sub-micron technologies. For the noise discussion, assume the thermal noise caused by the resistors and MOSFET channel current are dominant. From Fig. 3.2 we can easily determine the output noise voltage, $\bar{v}^2_{out}$,due to various devices as

$$\bar{v}^2_{out,R_L} = 4KTR_L \times \Delta f \tag{3.3}$$

$$\bar{v}^2_{out,R_1} = 4KTR_1 \times {g_{m_3}}^2 \times {R_L}^2 \times \Delta f \tag{3.4}$$

$$\bar{v}^2_{out,M_3} = \frac{\gamma}{\alpha} 4KTg_{m_3} \times {R_L}^2 \times \Delta f \tag{3.5}$$

$$\bar{v}^2_{out,M_4} = \frac{\gamma}{\alpha} 4KTg_{m_4} \times {R_L}^2 \times \Delta f \tag{3.6}$$

$$\bar{v}^2_{out,R_s} = 4KTR_s \times (R_s \parallel Z_{in})^2 \times {A_v}^2 \times \Delta f \tag{3.7}$$

$$\bar{v}^2_{out,M_1} = \frac{\gamma}{\alpha} 4KTg_{m_1} \times (R_s \parallel \frac{r_{o1}}{1+\frac{R_1}{R_s}} \parallel \frac{1}{g_{m_1}})^2 \times (\frac{R_1}{R_s} g_{m_3} - g_{m_4})^2 \times {R_L}^2 \times \Delta f \tag{3.8}$$

The first congregated terms in Eq. 3.8, $R_T = (R_s \parallel \frac{r_{o1}}{1+\frac{R_1}{R_s}} \parallel \frac{1}{g_{m1}})$, can be viewed as the equivalent impedance appearing to M_1 noise current at the common-gate node. The term associated with the second parenthesis indicates that noise voltage at V_1 and V_X due to M_1 is ratioed by $-\frac{R_s}{R_1}$. Neglecting noise of the cascode transistor M_5, the LNA noise factor is defined by

$$F = 1 + \frac{\bar{v}^2_{out,M_1} + \bar{v}^2_{out,R_L} + \bar{v}^2_{out,M_3} + \bar{v}^2_{out,M_4}}{\bar{v}^2_{out,R_s}} \tag{3.9}$$

Substituting Eq. 3.3 to Eq. 3.8 into Eq. 3.9,

$$F = 1 + \frac{g^2_{m3} R_1 + \frac{\gamma}{\alpha}\left(g_{m1} R_T^2 \left(g_{m3}\frac{R_1}{R_s} - g_{m4}\right)^2 + g_{m3} + g_{m4}\right) + \frac{1}{R_L}}{R_L^{-2} R_s^{-1} \cdot (R_s \| Z_{in})^2 \cdot A_v^2} \tag{3.10}$$

The terms in the numerator of Eq. 3.10 refer to output noise due to R_1, M_1, M_3, M_4 and

R_L respectively. It is clearly seen that M_1's noise, now becomes a result of g_{m3} and g_{m4} subtraction. A residual factor

$$\delta = \frac{R_s}{R_1}\frac{g_{m4}}{g_{m3}} - 1 \tag{3.11}$$

is used to define the degree of M_1 noise cancelation. $\delta = -1$ means disabling M_4 and no cancelation. $\delta = 0$ represents full cancelation of M_1 noise. δ can be any value greater than -1.

The input impedance matching, S_{11}, is infinitesimal if the LNA input impedance is perfectly matched to $R_s = 50\ \Omega$. In practice, a small impedance mismatch is acceptable that relaxes the selection of g_m and R_1. To capture the minimum input match, we introduce an error factor,

$$\varepsilon_{rr} = \frac{Z_{in}}{R_s} - 1 = \frac{2S_{11}}{1 - S_{11}} \tag{3.12}$$

For instance, ε_{rr}=0.22 corresponds to S_{11}=-20 dB.

3.3 Noise contour analysis

The above equations provide the design guideline for meeting A_v, Z_{in} and F requirements. However it is difficult to directly apply these equations to designing the LNA because there are too many parameters involved at the same time. In practice, a NF contour analysis by sweeping the parameter values is more commonly used in order to apprehend the design insights. Fig. 3.3 shows such NF contour. The parametric analysis is performed based on Eq. 3.10 by sweeping g_{m1}, g_{m3} and g_{m4}, in the form of δ, with the following constrains: $R_s = 50\ \Omega$, $S_{11} = 20\ dB$, $A_v = 18\ dB$, $g_m r_o = 12$, $\frac{\gamma}{\alpha}$=2.5 and total transconductance of 100 mS. During the parametric analysis, resistance R_1 and R_L are

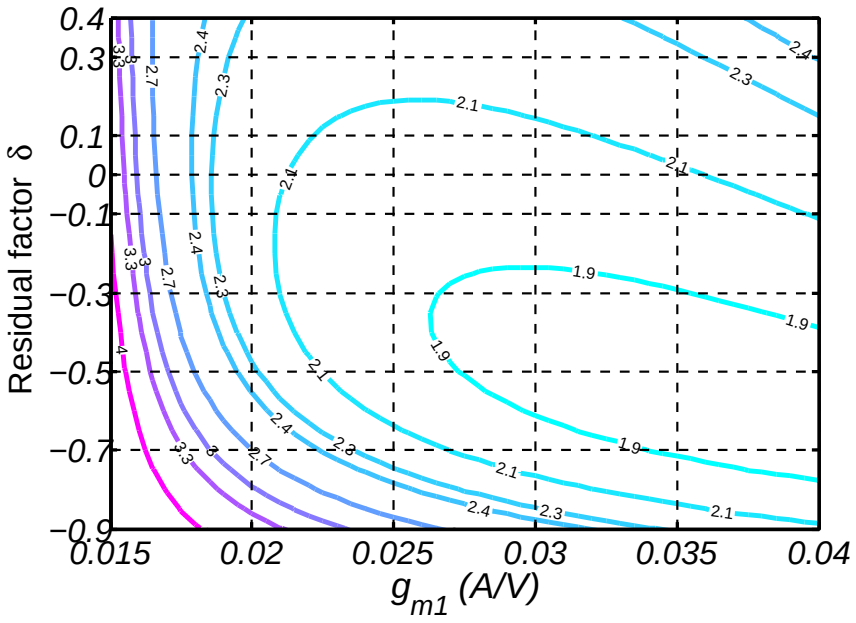


Figure 3.3: NF contour plot with different g_{m3} and g_{m4}

adjusted as transistor g_m varies in order to satisfy the S_{11} and A_v constrains. The values of R_1 and R_L are not transparent in the contour plot and needs to be carefully checked for headroom issue. The LNA bias current, to first order, is related to the overall g_m by $\frac{gm}{I} = \frac{2}{V_{dsat}}$. If transistor bias is chosen such that $V_{dsat} = 0.2\ V$, g_m of 100 mS translates into a current budget of 10 mA. Fig. 3.3 reveals several interesting characteristics. First, curves in the bottom-left triangular region shows exacerbated gradient and should be avoided for design robustness. The lowest noise figure appears to occur for larger g_{m1}. g_{m1} selection has to be checked with the value of the corresponding R_1 under S_{11} of 20 dB to see if the IR drop exceeds the available voltage headroom. From Eq. 3.1, the value of R_1 is in proportion to g_m and can not be made arbitrarily large. Moreover, for practical selection of g_{m1}, e.g. $g_{m1} = 30\ mS$, the lowest noise figure does not occur at full cancelation of M_1 noise, i.e.

$\delta = 0$. This is because the common-gate common-source circuit is configured to cancel the noise from the common-gate transistor only. Any other devices in Fig. 3.2 also contribute noise and is not cancelled. A full cancelation of M_1 noise, though desirable, may lead to sub-optimal design for the remaining noise sources and shift the location of NF_{min} away from $\delta = 0$ in Fig. 3.3. Here then comes a natural question: is it possible, or what is the condition, to realize a noise cancelation design whose NF_{min} matches with $\delta = 0$? This turns out to be an issue related to both circuit configuration and power consumption, and is explored shortly.

3.4 Power-noise figure tradeoff

The above contour-based methodology facilitates LNA design by graphically visualizing the impacts of varying the design parameters from the change of the contour curves. The only drawback is that it does not reveal the noise behavior of individual devices, nor does it point to a formulated design optimization. More specifically, it is unclear from the contour analysis how to allocate individual devices to achieve lowest overall noise figure. Moreover, how noise figure is affected by the power consumption of the circuit remains an interesting question. A more analytical approach is introduced as follows to address these issues.

3.4.1 Power-constrained noise optimization

Let us combine Eq. 3.1, Eq. 3.2 and Eq. 3.12 such that

$$\begin{aligned} A_v &= \left(g_{m4} + g_{m3}\frac{R_1}{Z_{in}}\right) \times R_L \\ &= \left(1 + \frac{1}{(1+\varepsilon_{rr})(1+\delta)}\right) \times g_{m4}R_L \end{aligned} \tag{3.13}$$

The next step is to relate Eq. 3.13 with Eq. 3.10 so that the relative output noise, i.e. the output noise caused by individual devices divided by that of the source resistance R_s, is identified.

$$F_{R_1} = \frac{R_s}{R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.14}$$

$$F_{M_1} = \frac{\frac{\gamma}{\alpha} g_{m1} R_T^2 \delta^2}{R_s} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.15}$$

$$F_{M_4} = \frac{\frac{\gamma}{\alpha}(1+\delta)}{g_{m3}R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.16}$$

$$F_{M_3} = \frac{\frac{\gamma}{\alpha}}{g_{m3}R_1} \times \frac{R_s}{R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.17}$$

$$F_{R_L} = \frac{1}{g_{m3}^2 R_L R_1} \times \frac{R_s}{R_1} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.18}$$

$$F = 1 + F_{R_1} + F_{M_1} + F_{M_3} + F_{M_4} + F_{R_L} \tag{3.19}$$

$$\Theta(\varepsilon_{rr}, \delta) = \frac{(2+\varepsilon_{rr})^2}{\left((1+\varepsilon_{rr})(1+\delta)+1\right)^2} \tag{3.20}$$

Eq. 3.14 to Eq. 3.20 describe individual device's noise behavior in terms of ε_{rr} and δ. In reality, it is ok to consider ε_{rr} as a constant rather than a variable because practical S_{11} has to be less than $-10\ dB$ and from Eq. 3.12 this means the value of ε_{rr} almost constant at 0.22. Since R_T is expressed in a way similar to Z_{in}, the same variable ε_{rr} can be used to represent R_T as well.

$$\begin{aligned} R_T &= \left(R_s \parallel \frac{r_{o1}}{1+\frac{R_1}{R_s}} \parallel \frac{1}{g_{m1}} \right) \\ &= \frac{R_s r_{o1}}{(1+g_{m1}r_{o1})(R_s+Z_{in})} \\ &= \frac{r_{o1}}{(1+g_{m1}r_{o1})(2+\varepsilon_{rr})} \\ &\approx \frac{1}{g_{m1}(2+\varepsilon_{rr})} \end{aligned} \tag{3.21}$$

F_{M_1} in Eq. 3.15 is then replaced by

$$F_{M_1} = \frac{\frac{\gamma}{\alpha}}{g_{m1}R_s} \times \frac{\delta^2}{(2+\varepsilon_{rr})^2} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.22}$$

In this way, we formulate Eq. 3.14 to Eq. 3.20 consistently with functions of ε_{rr} and δ. For the same ε_{rr}, δ, g_{m3} and R_L, it is clear to see that increasing g_{m1}, thus increasing R_1 for the same S_{11}, has the effect of lowering the relative noise of all devices. Therefore a large g_{m1} is always favored. From 2-port noise representation, both R_s and R_1 noise currents flow through R_1, and M_1 input equivalent noise voltage is inversely related to g_m. Therefore both g_{m1} and R_1 should be maximized. Furthermore, large g_{m1} and R_1 boost the voltage gain and suppress the noise contribution of the second stage devices, i.e. M_3, M_4 and R_L. The maximum g_{m1} and R_1, however, are limited by the available voltage headroom. For instance, $R_1 = 464\ \Omega$ is paired with $g_{m1} = 25\ mS$ given ε_{rr}=0.22 and $g_{m1}r_{o1}$=22. This results in a voltage drop across resistor R_1 as large as $1.16\ V$ assuming

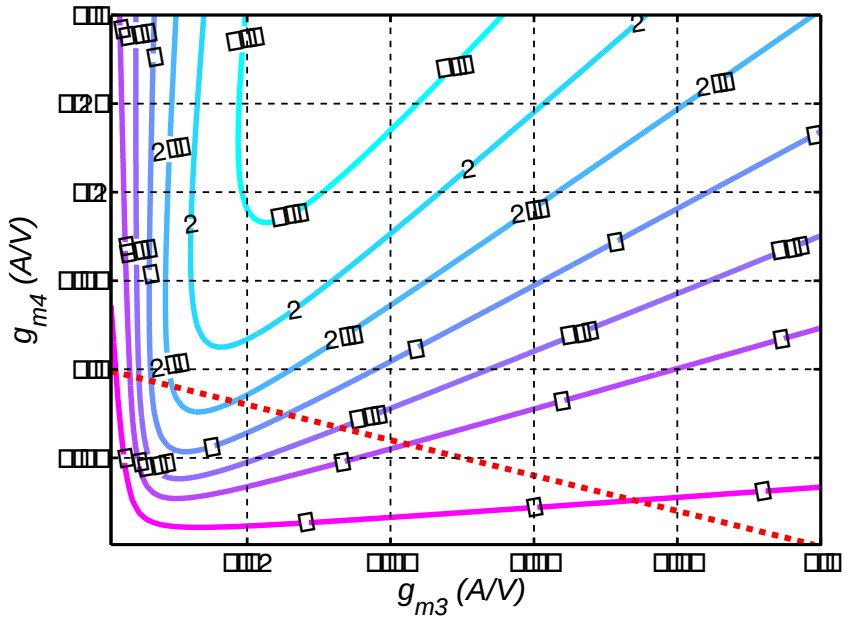


Figure 3.4: NF contour plot with different g_{m3} and g_{m4} at $g_{m1} = 25mS$, R_1=464 Ω, $g_m r_o$=22, $\frac{\gamma}{\alpha}$=2.5

the transistor overdrive voltage is 0.2 V. For nano-scale CMOS technology, this already reaches the maximum g_{m1} value since the supply voltage is around 1 V only.

Once g_{m1} and R_1 are determined, the remaining task is to decide on an optimum ratio for g_{m3} and g_{m4} as this affects M_1 noise cancelation as well as noise contribution of noncanceled sources. Fig. 3.4 shows another contour by varying g_{m3} and g_{m4} at g_{m1} = 25 mS and R_1 = 464 Ω. The straight dash line in Fig. 3.4 refers to a constant current consumption of 10 mA for M_3+M_4. The dash line intercepts the contour at different g_{m3}, i.e. different residual cancelation ratio δ. An optimum ratio, δ_{opt}, occurs at the turn-around of the contour curves and the LNA NF reaches its local minimum. By moving the straight line up and down and checking its interception with the contour, the lowest NF at

different power consumption is determined [1]. Mathematically the optimal value of δ exists if derivative of Eq. 3.19 is zero.

$$\delta_{opt} \equiv \frac{\partial F_{R_1}}{\partial \delta} + \frac{\partial F_{M_1}}{\partial \delta} + \frac{\partial F_{M_4}}{\partial \delta} + \frac{\partial F_{M_3}}{\partial \delta} + \frac{\partial F_{R_L}}{\partial \delta} = 0 \tag{3.23}$$

For now we will pursue more design insights before resorting this approach to find δ_{opt}. From Fig. 3.4, it is clear that LNA noise figure strongly depends on the total g_m. To reflect the bias current dependence, we introduce Σg_m, the total g_m, as a new parameter. Σg_m is related to LNA current with a constant overdrive voltage of, say, 0.2 V. Rewrite from Eq. 3.16 to Eq. 3.18 with Σg_m,

$$F_{M_4} = \frac{\frac{\gamma}{\alpha}(1+\delta)}{R_1} \times \frac{1+(1+\delta)\frac{R_1}{R_s}}{(\Sigma g_m - g_{m1})} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.24}$$

$$F_{M_3} = \frac{\frac{\gamma}{\alpha}R_s}{R_1^2} \times \frac{1+(1+\delta)\frac{R_1}{R_s}}{(\Sigma g_m - g_{m1})} \times \Theta(\varepsilon_{rr}, \delta) \tag{3.25}$$

$$F_{R_L} = \frac{R_s}{R_L R_1^2} \times \left(\frac{1+(1+\delta)\frac{R_1}{R_s}}{(\Sigma g_m - g_{m1})} \right)^2 \times \Theta(\varepsilon_{rr}, \delta) \tag{3.26}$$

If we pick Σg_m =125 mS and follow the trace on the straight line from the right to the left, i.e. from δ=-1 to $\delta = \infty$, the trend of relative output noise of individual devices is illustrated in Fig. 3.5. There are some distinct features shown in this figure. First, F_{M_1} curve drops from large value as δ begins departing from -1, i.e. no M_1 noise cancelation, reach zero at $\delta = 0$, then bounces back up for δ larger than 0. On the other hand, F_{R_1} and F_{M_4} shows a trend of monotonic increase and decrease, respectively. We also notice that

[1] assume transistors still in saturation region

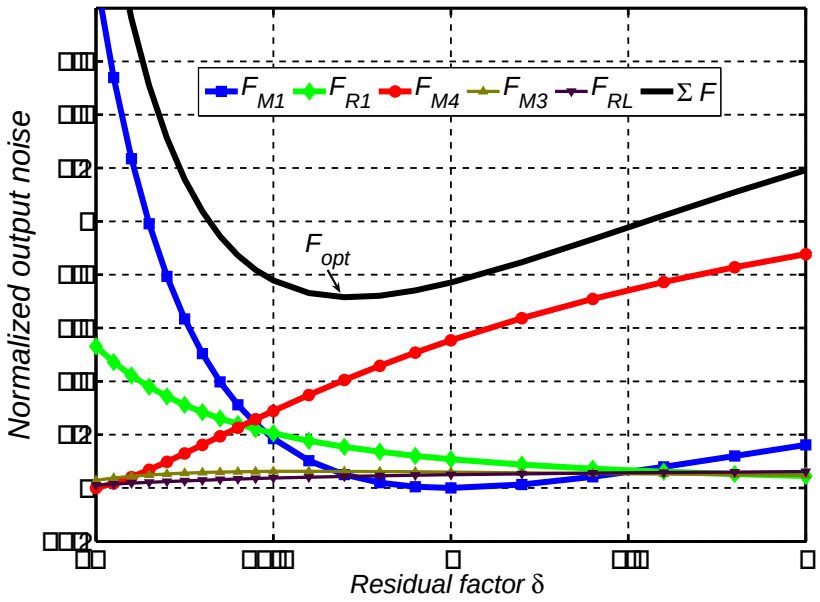


Figure 3.5: Relative output noise with different *delta* and fixed LNA current=12.5 mA/V

F_{M_3} and F_{R_L} stay relatively constant as opposed to the others. Because F_{M_3} and F_{R_L} are the noise contribution from the second stage, their magnitude is suppressed by a reasonable voltage gain of $\frac{R_1}{R_s}$. Therefore, they are be excluded from NF_{opt} discussion.

Now let us explain the tendency of the F_{M_1}, F_{R_1} and F_{M_4} curves in Fig. 3.5 in more detail. Starting with the non-cancelation case, i.e. M_4 off and δ=-1, the LNA NF is dominated by M_1 and partly by R_1 due to a typical $\frac{R_s}{R_1}$ greater than 5. As M_4 gradually turns on, both F_{M_1} and F_{R_1} decrease because M_4 amplifies only the R_s noise, but not M_1 and R_1. According to dividing Eq. 3.14 with Eq. 3.15, F_{M_1} decreases faster than F_{R_1} by

$$\begin{aligned} \frac{F_{M_1}}{F_{R_1}} &= \frac{\frac{\gamma}{\alpha} g_{m1} R_1 R_T^2}{R_s^2} \delta^2 \\ &= \frac{\frac{\gamma}{\alpha} R_1}{g_{m1} R_s^2} \frac{\delta^2}{(1+\varepsilon_{rr})^2} \end{aligned} \tag{3.27}$$

This is because M_1 and R_1 noise experiences the same M_3 amplification whereas M_1 noise experiences cancelation through the M_4 path. Along with the decreasing F_{M_1} and F_{R_1}, F_{M_4} keeps increasing according to Eq. 3.24. This is an interesting phenomenon since the transistor noise figure is expected to improve as its bias current increases. From R_s noise's perspective, M_3 and M_4 form two amplifying paths in parallel. Under fixed g_m assumption, g_{m4} increases by the same amount as the loss in g_{m3}. The R_s noise sees more reduction along the M_3 path than the increase along the M_4 path due to the voltage gain of $\frac{R_1}{R_s}$ before M_3. On the other hand, output noise due to M_4 keeps increasing as g_{m4} increases. The opposite trend of output noise due to R_s and M_4 leads to a monotonic increase of F_{M_4}.

As indicated by Eq. 3.23, δ_{opt} is the point where the falling of F_{M_1} and F_{R_1} and rising of F_{M_4} reach an equilibrium. Once g_{m1} and R_1 are chosen at its maximum value, the curves of F_{M_1} and F_{R_1} depends solely on the cancelation ratio δ. The remaining way to move δ_{opt}, thus NF_{opt}, in Fig. 3.5 is by changing the curve of F_{M_4}. From Eq. 3.24, F_{M_4} is inversely related to g_{m3} and this is done by assigning different bias current, i.e. power consumption. For instance, with larger Σg_m to begin with, F_{M_4} becomes smaller in all ranges of δ and shifts both the interception of F_{M_4} with F_{M_1} and F_{R_1}, and δ_{opt} to the right. This results in a lower F, and vice versa.

An empirical expression of δ_{opt} exists by simplifying Eq. 3.23 and taking only the sum of $\frac{\partial F_{R1}}{\partial \delta}$, $\frac{\partial F_{M1}}{\partial \delta}$ and $\frac{\partial F_{M4}}{\partial \delta}$ to be zero.

$$\delta_{opt} = \frac{\frac{2R_s}{R_1}(1+\varepsilon_{rr}) - \frac{\frac{\gamma}{\alpha}}{R_1(\Sigma g_m - g_{m1})}(\frac{2R_1}{R_s} - \varepsilon_{rr})}{\frac{2\frac{\gamma}{\alpha}g_{m1}R_T^2}{R_s}(2+\varepsilon_{rr}) + \frac{\frac{\gamma}{\alpha}}{R_1(\Sigma g_m - g_{m1})}(\frac{2R_1}{R_s} - \varepsilon_{rr} - 1)} \tag{3.28}$$

δ_{opt} increases as Σg_m increases. Moreover, we can find the LNA NF_{opt} by substituting δ_{opt}

back to Eq. 3.9.

$$F_{opt} = 1 + (\frac{R_s}{R_1} + \frac{\frac{\gamma}{\alpha} g_{m1} R_T^2 \delta_{opt}^2}{R_s} + \frac{\frac{\gamma}{\alpha}}{R_1(\Sigma g_m - g_{m1})} \times (1 + \delta_{opt} + \frac{R_1}{R_s}(1 + \delta_{opt})^2)) \times \Theta(\varepsilon_{rr}, \delta_{opt}) \tag{3.29}$$

Eq. 3.28 and Eq. 3.29 are very useful results because they allow us to choose the parameter values to achieve the lowest LNA noise figure at the given bias current. Eq. 3.28 and Eq. 3.29 are compared with the δ_{opt} and NF_{opt} extracted from Fig. 3.4 of the previous contour analysis. Excellent matching between both results is seen in Fig. 3.6. The case of conventional common-gate circuit is also plotted in Fig. 3.6 and shows the distinctive benefit of noise cancelation scheme.

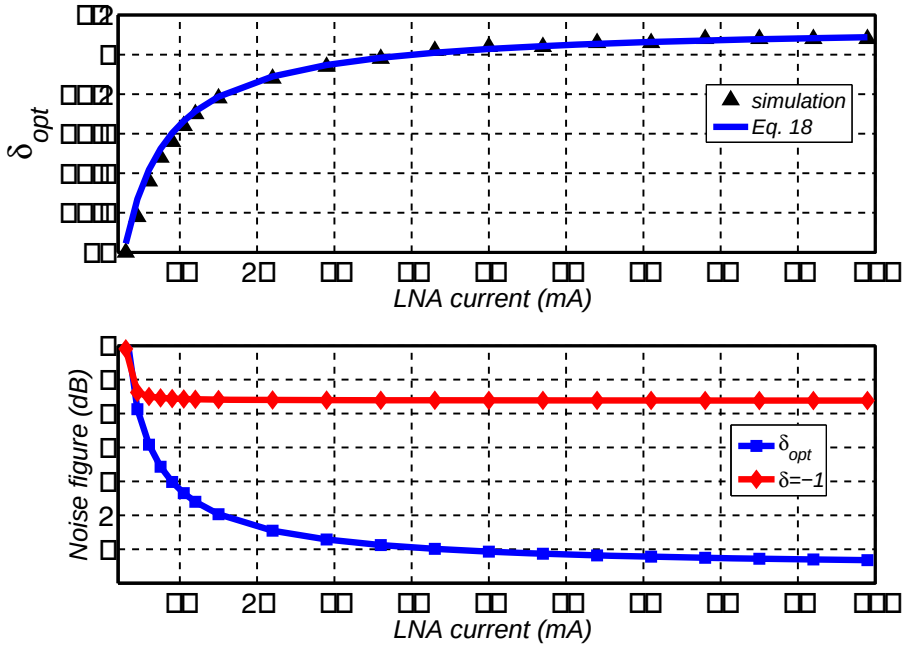


Figure 3.6: NF and δ_{opt} at different LNA current

3.4.2 Power-efficient noise optimization

In Fig. 3.6, lower noise figure is traded off higher power consumption. The benefit of noise cancelation is most appreciable when the LNA bias current is still small then becomes marginal at high bias level. To characterize this dynamics with a quantitative expression, consider the first checkpoint, δ_1, at the crossover of F_{M_1} and F_{R_1}. This implies a substantial M_1 noise reduction has been achieved.

$$\delta_1 = -\frac{R_s}{R_T}\sqrt{\frac{1}{\frac{\gamma}{\alpha}g_{m1}R_1}} = -0.526 \tag{3.30}$$

Substituting δ_1 into Eq. 3.28, the corresponding bias level in terms of total g_m is

$$\Sigma g_{m,\delta_1} = \ 67mA/V \tag{3.31}$$

which matches with bias level of 6.7 mA in Fig. 3.6. For bias current exceeding this level, the noise from M_1 and R_1 has become less dominant and the added power consumption is mainly used to reduce M_4 noise contribution in a fashion similar to $\frac{1}{g_m}$ dependence. In this regards, the power efficiency is the same as the conventional approach.

It is then interesting to predict the lowest achievable NF if the constraint on power is lifted. Continually increasing g_{m3} and g_{m4}, the curve of F_{M_4}, F_{M_3} and F_{R_L} approach to zero asymptotically. In the end, the LNA NF is limited by R_1 and M_1 only. An estimate of δ for such scenario is found by making $\frac{\partial F_{M_1}}{\partial \delta}+\frac{\partial F_{R_1}}{\partial \delta}=0$ such that

$$\delta_3 = \frac{R_s^2}{\frac{\gamma}{\alpha}g_{m1}R_1R_T^2} \times \frac{1+\varepsilon_{rr}}{2+\varepsilon_{rr}} = +0.15 \tag{3.32}$$

Notice that δ_3 is larger than $\delta = 0$ where $\frac{\partial F_{M_1}}{\partial \delta}$ is positive. The corresponding NF_{opt} ends up to be less than 1 dB as shown in Fig. 3.6. This noise figure value is unfortunately

impractical because of the large bias current and DC headroom. A realistic perspective is by considering a practical current level, say, 200 mS for 20 mA, that fits in the available supply headroom. A sub-2 dB LNA noise figure can still be achieved at this bias level.

3.5 Discussions

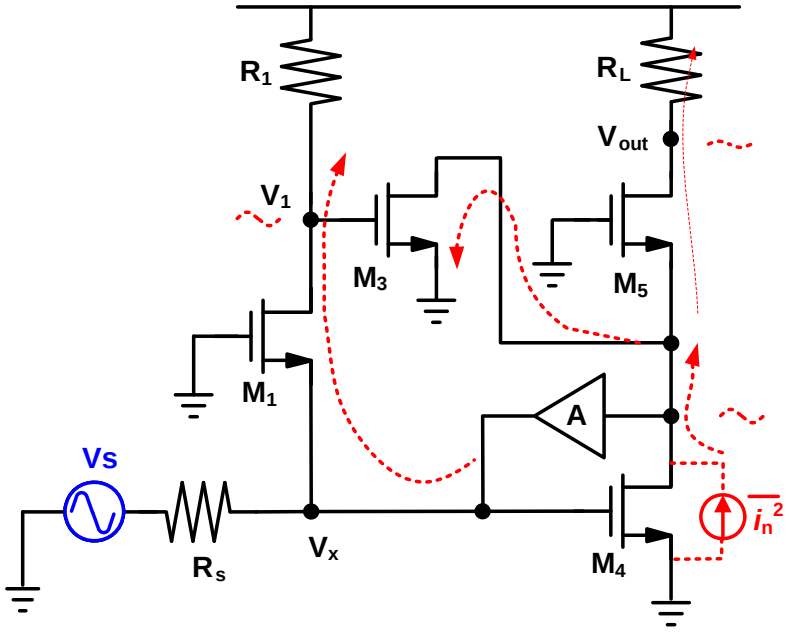


Figure 3.7: LNA with M_4 noise cancelation

We have analyzed the power-noise figure tradeoff of the common-gate common-source noise canceling configuration. From Fig. 3.6, it is identified that F_{M_4} poses as the major noise contribution in the case of low LNA noise figure. Achieving sub-2 dB noise figure requires substantially large bias current and may not be feasible for some very low power applications. Another solution is by applying a second noise cancelation around

transistor M_4 while keeping the remaining circuit intact. This is shown in Fig. 3.7. M_4 noise is first sensed at V_1 through a local feedback, inverted by M_3, and, in turn, subtracted at the drain of M_3 and M_4. Qualitatively speaking, adding a feedback around transistor M_4 only reduces M_4's effective g_m, but does not disturb the previous noise analysis for other devices.

Chapter 4

Distortion Cancelation Design

According to Eq. 2.37, transistor linearity improves by high overdrive voltage. This calls for the adoption of elevated supply voltage, which becomes less feasible in the era of low supply voltage of scaled CMOS technology. On the other hand, circuit or system linearizations can be used to compensate the deteriorating device linearity. A common circuit technique is by adding feedback around the original circuit such that the magnitude of actual signal driving the devices is smaller than before so as to avoid the undesirable nonlinear operation. With noise cancelation, feedback amplifiers also achieve broadband impedance match and low noise figure [53].

Another category of improving the linearity is by cancelation. Unlike noise whose characteristics is random and hard to manipulate, circuit distortions are deterministic, leading to easier and multiple ways of cancelation scheme. Mixed-mode or digital approaches employ adaptive algorithm to characterize the receiver nonlinearity [54, 55]. The receiver distortion is predicted and subtracted from the output, assuming that the scheme responds

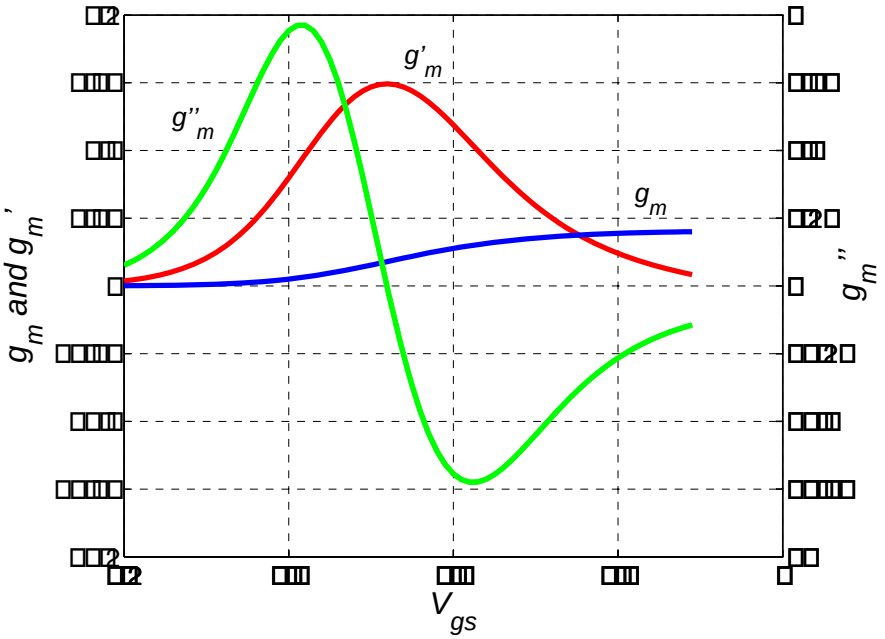
fast enough to the blocker profile. The latency of the FIR algorithm is the key for this scheme. The cancelation of distortion can also be realized in the analog domain. Some techniques have been available for a long time. Differential signalling, for example, is be among the most widely implemented to cancel even-order nonlinearity using matched devices. Many distortion cancelation techniques are developed for power amplifier designs as the circuit has to deal with large signal swing which easily drives the circuit into highly nonlinear regime. Feedforward [56] cancelation and LINC (Linear amplification using nonlinear components) [57] amplifiers are two well-known ones. However for low-noise amplifiers, the signal magnitude is small and the noise added by the circuit is a major concern. Transmitter linearization approaches utilizing complex configuration inherently introduce more noise and are less suitable. The work in [58] uses a second LNA in parallel to the main LNA and manages to match the third-order distortion between the two paths for complete cancelation. The noise figure degradation is minor due to the gain difference between two paths. However, the assumption that a linearly scaled input signal is available at LNA input is less practical. Other LNA works exploit the device's nonlinear properties with matched transistor bias and size [59, 60, 61, 62]. Excellent IIP_3 is reported, but they are tailored for narrow-band applications only. A viable distortion cancelation for broadband LNA has not been devised yet. In this chapter, we will develop distortion cancelation design for the proposed broadband noise canceling LNA.

4.1 All-Range MOSFET Linearity

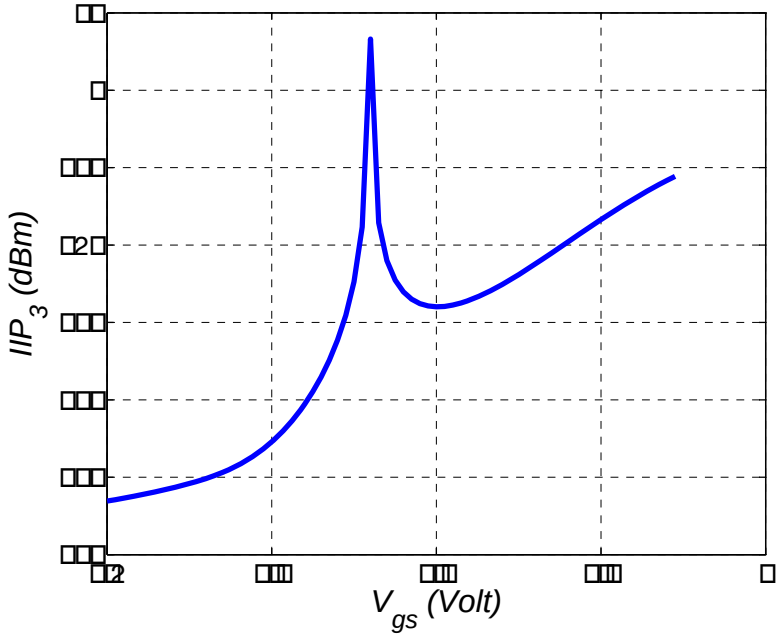
Fig. 4.1(a) shows the simulated g_m, g'_m and g''_m of a $6.52\mu m/0.13\mu m$ NMOS transistor at a constant V_{ds} and a sweeping V_{gs}. At low V_{gs}, drain current is caused by the diffusion of majority carriers in the substrate. The behavior in this weak inversion region is similar to BJT, and g_m increases exponentially. At high V_{gs}, the inversion layer of minority carriers fully forms and changes the current transport mechanism to the drift of these carriers in the channel. Since the current is proportional to the carrier speed, eventually g_m saturates because of the mobility reduction and velocity saturation in high field condition. The drain current in this strong inversion region is previously described by Eq. 2.35. As transistor transits from weak to strong inversion, the exponential growth of g_m gives away to the trend of saturation. Therefore in the intermediate region slightly above the threshold voltage, the growth of g_m reaches its maximum. This results in a bell-shaped g'_m curve. g''_m, the derivative of g'_m, at this particular turn-around becomes zero [63].

4.1.1 IM_3 sweet spot

According to the definition in Eq.2.29, zero g''_m corresponds to an excellent third-order linearity. This is demonstrated by the peaking at a sweet-spot bias in Fig. 4.1(b). Vigilant readers may wonder why infinite IIP_3 is not seen in Fig. 4.1. The reasons are two fold. First, it is very difficult, if not impossible, to pinpoint the exact g''_m sweet spot in the simulation. A minuscule offset from sweet spot results in finite peak value. Secondly, as a_3 nullifies IM_3 at the sweet spot, the impact of higher order nonlinearities such as a_5 and a_7, become dominant and is not taken into account in the simplified definition of Eq. 2.29. A closer discussion probing into this issue is presented in chapter 6.



(a) g_m and its derivatives



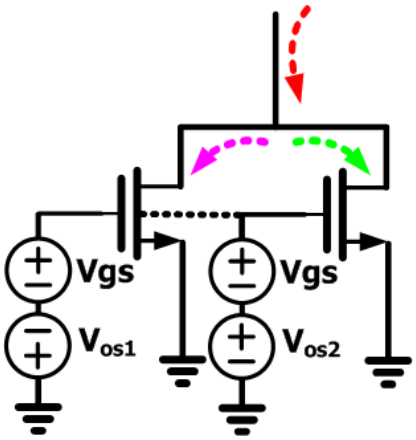
(b) IIP_3

Figure 4.1: MOSFET linearity at all V_{gs}

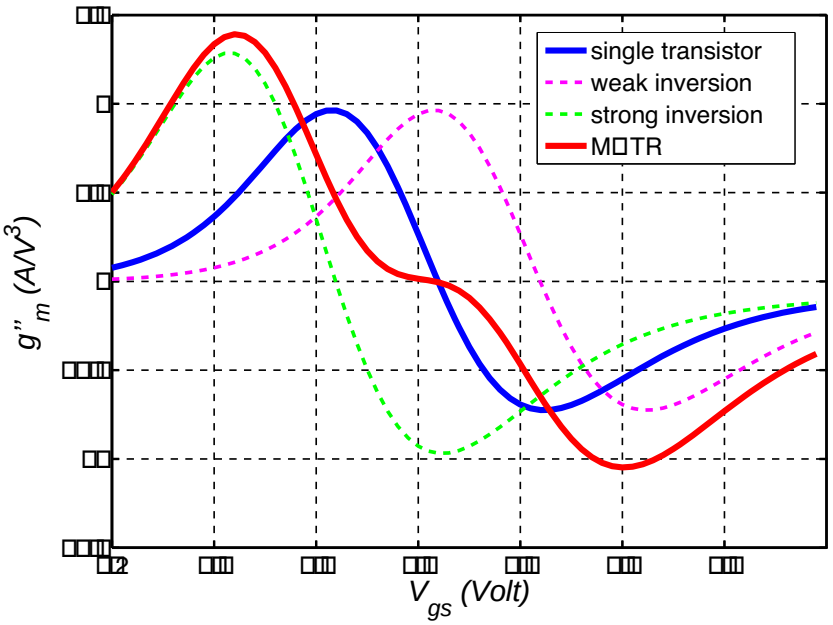
Although nano-scale transistors also show strong nonlinearity from the output drain conductance, the impacts can be mitigated by lowering the resistance connected to the output and by resorting to a current-mode style of design [64]. The g_m'' sweet spot, therefore, is an unique property that can be taken advantage of for LNA linearization. However, the steep slope of the zero cross-over makes this scheme sensitive to process and bias variations. Moreover, the second derivative of g_m is usually ill modeled because device model extraction in today's practice only checks the matching of I_{ds} and g_m between the model and measured data, but neglects their higher-order derivatives. This leaves room for more error in the g_m''-based linearization designs.

4.1.2 Multi-Gated Transistor (MGTR)

Besides the sweet spot, g_m'' curve also shows a peak and a valley in regions of opposite sign. This property can be leveraged to reduce the bias sensitivity of the previous scheme. As shown in Fig. 4.2, the new scheme consists of a composite transistor of two which are biased at the peak and the valley by adding different offset voltages. The transistor of positive peak is in weak inversion and the another in strong inversion. The offset voltages move the g_m'' curve in both directions and align the peak with the valley. The size between the two transistors is matched as well so that their g_m'' magnitude equals the point of interest [65, 66]. As such, a g_m'' curve with much flatter zero-crossing is synthesized. It has a broader range to accommodate bias variation while keeping g_m'' still small. Because of the cancelation, $MGTR$ linearization is less affected by the linearity of the individual transistor.



(a) g_m and its derivatives



(b) IIP_3

Figure 4.2: Illustration of $MGTR$ scheme

4.2 Distortion cancelation design

Having reviewed the $MGTR$ technique, we now turn to the LNA linearization in Fig. 3.2. Assume transistors M_1 through M_4 are the only nonlinear devices of consideration. Cascode common-gate transistor M_5 provides the current buffering and ideally does not induce additional distortion. Since third-order distortion is concerned, it is assumed that for each transistor there are three current sources between their drain and source nodes, represented by $g_m v_{gs}$, $g'_m v_{gs}^2$ and $g''_m v_{gs}^3$, to model the linear and nonlinear currents caused by each devices. For the common-gate transistor M_1, this modeling strategy is similar to the way its drain noise current is modeled. Therefore, M_1 distortion current, both second and third order, is expected to cancel in the same vein as noise cancelation. On the other hand, M_3/M_4 could dominate the distortion since they are in the second-stage of the amplifier. As a quick observation from Fig. 3.2, M_3 and M_4 can be configured in a $MGTR$ style so that their intrinsic third-order distortion cancels. As such, the proposed common-gate common-source cascade has the potential to achieve very low overall IM_3..

4.2.1 Volterra series analysis

To verify the above observation, the schematic of Fig. 4.3 for the common-gate stage is used to perform the distortion analysis. C_1 and C_x account for the parasitic capacitance associated with the nodes. To capture high frequency effect, Volterra series analysis is performed. Denote V_1 and V_x with their 1^{st}, 2^{nd}, and 3^{rd}-order Volterra kernels

$$\begin{aligned} V_x &= A_1(s_1) \circ V_s + A_2(s_1, s_2) \circ V_s^2 + A_3(s_1, s_2, s_3) \circ V_s^3 \\ V_1 &= B_1(s_1) \circ V_s + B_2(s_1, s_2) \circ V_s^2 + B_3(s_1, s_2, s_3) \circ V_s^3 \end{aligned} \tag{4.1}$$

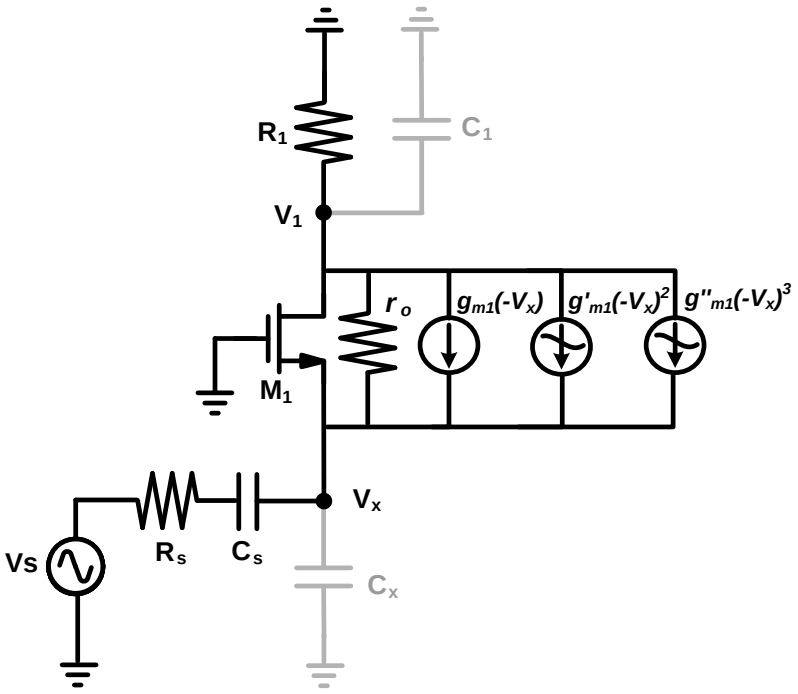


Figure 4.3: Common-gate schematic for distortion analysis

On the first order, the following KCL equations are established.

$$\begin{aligned} \frac{A_1(s)}{Z_x(s)} + \frac{A_1(s)-1}{Z_s} + \frac{A_1(s)-B_1(s)}{r_o} &= -g_{m1}A_1(s) \\ g_{m1}A_1(s) &= \frac{B_1(s)}{Z_1(s)} + \frac{B_1(s)-A_1(s)}{r_{o1}} \end{aligned} \tag{4.2}$$

where $Z_s(s)$=$R_s + \frac{1}{sC_s}$, $Z_x(s)$=$\frac{1}{sC_x}$ and $Z_1(s)$=$R_1 \parallel \frac{1}{sC_1}$. By solving Eq.4.2, the first-order Volterra kernel $A_1(s)$ and $B_1(s)$ are derived as

$$\begin{aligned} A_1(s) &= \frac{Z_1(s)+r_{o1}}{H(s)} \\ B_1(s) &= \frac{Z_1(s)\times(1+g_{m1}r_{o1})}{Z_1(s)+r_{o1}}A_1(s) \end{aligned} \tag{4.3}$$

where

$$H(s) = Z_s(s)(1+g_{m1}r_{o1}) + (Z_1(s)+r_{o1})\left(1+\frac{Z_s(s)}{Z_x(s)}\right) \tag{4.4}$$

For the second-order terms, the following KCL equations are used.

$$\begin{aligned} \frac{A_2(s)}{Z_x(s)} + \frac{A_2(s)}{Z_s} + \frac{A_2(s) - B_2(s)}{r_o} &= -\left(g_{m1}A_2(s) + g'_{m1}A_1(s)^2\right) \\ g_{m1}A_2(s) + g'_m A_1(s)^2 &= \frac{B_2(s)}{Z_1(s)} \end{aligned} \tag{4.5}$$

$A_2(s_1, s_2)$ and $B_2(s_1, s_2)$ are found as

$$\begin{aligned} A_2(s_1, s_2) &= \frac{\frac{1}{2}g'_{m1}r_{o1}Z_s(s_1+s_2)A_1(s_1)A_1(s_2)}{H(s_1+s_2)} \\ B_2(s_1, s_2) &= \frac{-Z_1(s_1+s_2)}{Z_x(s_1+s_2) \parallel Z_s(s_1+s_2)}A_2(s_1, s_2) \end{aligned} \tag{4.6}$$

Likewise, $A_3(s_1, s_2, s_3)$ and $B_3(s_1, s_2, s_3)$ are

$$\begin{aligned} A_3(s_1, s_2, s_3) &= \frac{-Z_s r_{o1}\left(-g'_{m1}\overline{A_1(s_1)A_2(s_2, s_3)} + \frac{1}{6}g''_{m1}A_1(s_1)A_1(s_2)A_1(s_3)\right)}{H(s_1+s_2+s_3)} \\ B_3(s_1, s_2, s_3) &= \frac{-Z_1(s_1+s_2+s_3)}{Z_x(s_1+s_2+s_3) \parallel Z_s(s_1+s_2+s_3)}A_3(s_1, s_2, s_3) \end{aligned} \tag{4.7}$$

where $\overline{A_1(s_1)A_2(s_2, s_3)}$ represents the 2^{nd}-order interaction Volterra operator by [67]

$$\overline{A_1(s_1)A_2(s_2, s_3)} = \frac{1}{3}\left(A_1(s_1)A_2(s_2, s_3) + A_1(s_2)A_2(s_3, s_1) + A_1(s_3)A_2(s_1, s_2)\right) \tag{4.8}$$

The harmonic voltages at V_1, V_x are then amplified through transistors M_3 and M_4 by

$$V_{out} = \left(g_{m3}V_1 + g_{m4}V_x + \frac{g'_{m3}}{2}V_1^2 + \frac{g'_{m4}}{2}V_x^2 + \frac{g''_{m3}}{6}V_1^3 + \frac{g''_{m4}}{6}V_x^3\right) \times Z_L \tag{4.9}$$

Plugging in the above Volterra kernel expression, the fundamental and 3^{rd}-order V_{out} expressions are found.

$$\begin{aligned} V_{out,fund} &= \left(\left(A_1(s) \circ V_s\right) \times g_{m4} + \left(B_1(s) \circ V_s\right) \times g_{m3}\right) \times Z_L \\ V_{out,3^{rd}} &= \Bigg(\left(\left(A_3(s_1, s_2, s_3) \circ V_s^3\right) \times g_{m4} + \left(B_3(s_1, s_2, s_3) \circ V_s^3\right) \times g_{m3}\right) \\ &\quad + \left(\left(A_1(s) \circ V_s\right)^3 \times \frac{g''_{m4}}{6} + \left(B_1(s) \circ V_s\right)^3 \times \frac{g''_{m3}}{6}\right) \\ &\quad + \left(\left(\overline{A_1(s_1)A_2(s_2, s_3)} \circ V_s^3\right) \times g'_{m4} + \left(\overline{B_1(s_1)B_2(s_2, s_3)} \circ V_s^3\right) \times g'_{m3}\right)\Bigg) \times Z_L \end{aligned} \tag{4.10}$$

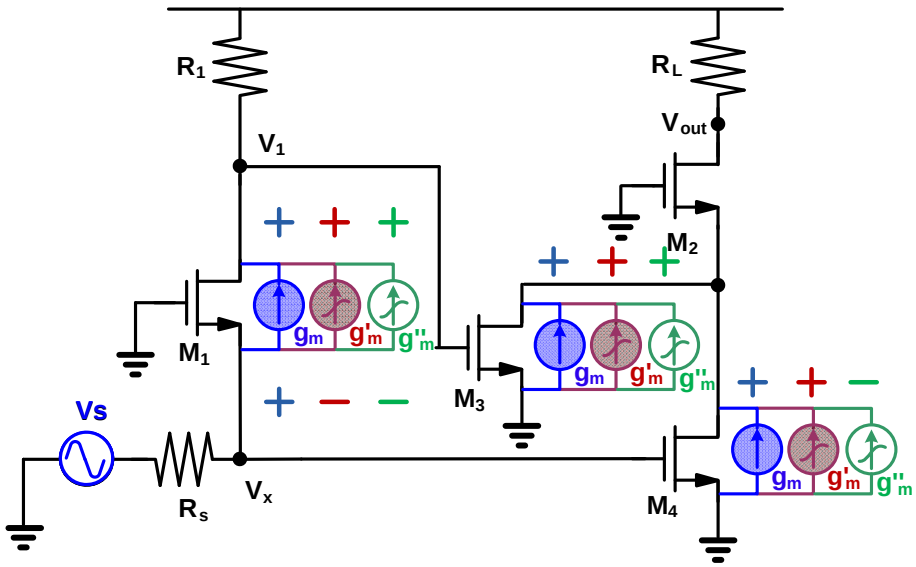


Figure 4.4: V_{out} IIP_3 versus g'_m of the common-gate transistor

In order to deconvolve Eq. 4.10, assume input impedance is matched so the ratios of $A_1(s)/B_1(s)$,$A_2(s_1,s_2)/B_2(s_1,s_2)$, $A_3(s_1,s_2,s_3)/B_3(s_1,s_2,s_3)$ reduce to R_s/R_1, $-R_s/R_1$ and $-R_s/R_1$ respectively at low frequencies. Fig. 4.4 shows the visualization of Eq. 4.10 by using 3 different current sources between drain and source to represent different order of harmonic current generated by g_m, g'_m and g''_m of individual transistors. A pair of plus and minus (or plus and plus) signs marks the polarity between different order of voltage at V_1 and V_x, and polarity of g_m (and its derivatives) of M_3 and M_4, respectively. The first $V_{out,IM3}$ is caused by the intrinsic IM_3 of M_1 magnified by g_m of common-source transistors. In Fig. 4.4, the (plus,minus) polarity between third-order V_1 and V_x and (plus, plus) polarity between M_3 and M_4's g_m implies a subtraction at V_{out}. This mechanism is exactly the same as the M_1 noise cancelation. Therefore M_1 noise cancelation also provides intrinsic IM_3 cancelation.

The second $V_{out,IM3}$ is viewed as the linear voltage at V_x and V_1 amplified by g''_m of M_3 and M_4. By employing $MGTR$ to M_3 and M_4, the polarity of their g''_m is chosen different to let cancelation at V_{out} take place. From noise cancelation design, the first $V_{out,IM3}$ cancelation is controlled by the ratio of g_{m3} and g_{m4}. With $MGTR$ employed to cancel the second $V_{out,IM3}$, the g''_m of M_3 and M_4 also needs to be matched. In other words, these two cancelations already constrain the selection of size and bias for M_3 and M_4.

However, there is a remaining $V_{out,3^{rd}}$ term that is not canceled by the aforementioned two schemes. It is caused by the linear and second-order nonlinear voltage at V_x and V_1 mixed and then multiplied with the second-order nonlinear coefficient, g'_m, of M_3 and M_4. This is regarded as the so-called second-order interaction that is commonly encountered in the cascade and feedback systems [47]. Although polarity of $A_2(s_1, s_2)$ and $B_2(s_1, s_2)$ is opposite and cancelation at V_{out} is also possibel, due to the lack of control of their values, the residual IM_3 after subtraction can still be substantial.

To examine how significantly the residual IM_3 may impact the overall IIP_3 of V_{out}, a circuit of Fig. 3.2 is designed using $0.13\mu m$ CMOS technology. After the device bias and size are chosen such that it provides 50 Ω input impedance match, and noise and intrinsic distortion cancelations, the values of g_m and its derivatives for the transistors are extracted from the *Spectre* simulation using foundry's BSIM4 model. The transistor M_1 in the original design has a g'_{m1} of 42 mA/V^2. Then a calculation on IIP_3 based on Eq. 4.10 by arbitrarily assigning different M_1 g'_m value is performed. A drastic change of IIP_3 is seen in Fig. 4.5. This confirms that the remaining term in Eq. 4.10 is limiting the highest achievable IIP_3. The peak in Fig. 4.5 is slightly off the center point. Referring to eq. 4.10,

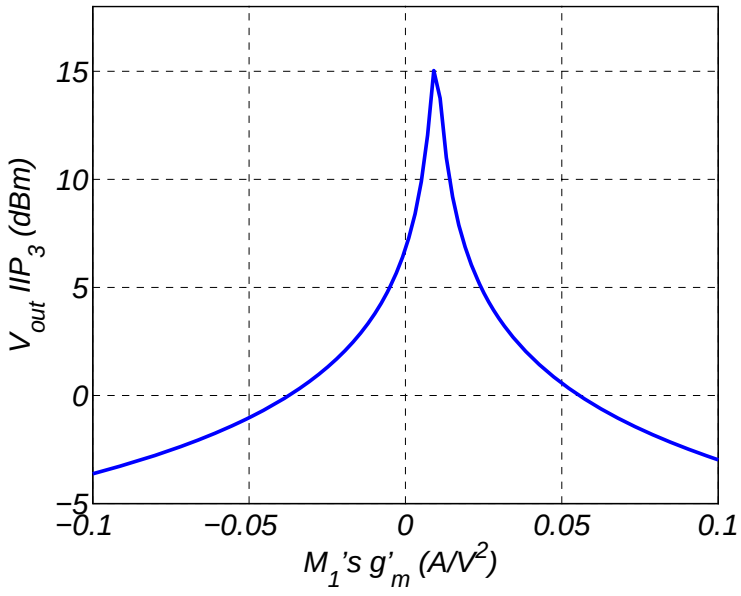


Figure 4.5: V_{out} IIP_3 versus g'_m of the common-gate transistor

this is simply because a nonzero δ_{opt} is chosen and a small g'_{m1} is used to counteract the residual IM_3 by M_1 itself.

4.2.2 Removal of second-order interaction

In order to eliminate this residual $V_{out,3^{rd}}$ due to second-order interaction, a NMOS-PMOS complementary pair is used. Fig. 4.6 shows the modified common-gate stage implemented with a P/NMOS pair. In general, PMOS is a dual to the NMOS device and shares the same g_m properties. With the AC coupling capacitor C_{12} connecting the drain of both transistors, the P/NMOS pair forms a composite transistor, each of which has the same common-gate configuration as in Fig. 3.2. Therefore the previous noise and gain analysis still applies except device parameters of the composite schematic are used instead.

For example, R_1 in Fig. 3.2 is replaced with $R_1 \parallel R_2$ in Fig. 4.6 and g_{m1} replaced with $g_{m1} \parallel g_{m2}$. As for the distortion current generated by the transistor pair, the result is

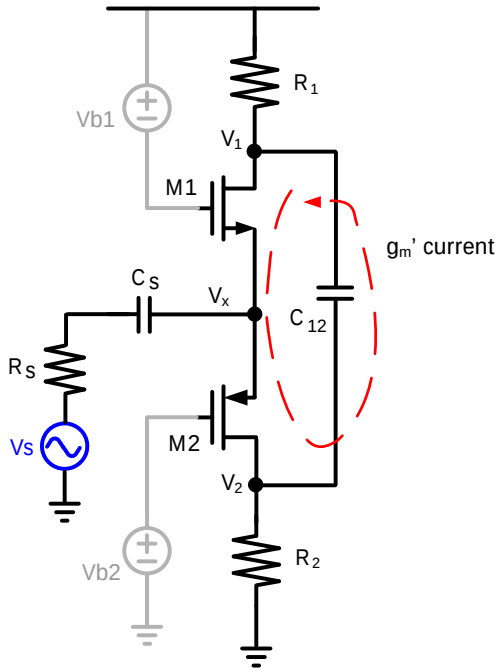


Figure 4.6: Circuit schematic to remove IM_3 second-order interaction

vastly different depending on the order of the harmonics of consideration. If we track the direction of the current flow, the current generated by g''_m of both transistors is departing from each other, similar to that by g_m as well as drain current noise. With C_{12}, g''_m currents are combined and flow into resistor R_1 in parallel with R_2. The current generated by g'_m of both transistors, on the other hand, will flow in the same direction. This is because g'_m is an even-order nonlinear coefficient acting upon v_{gs}^2. The even-order power of v_{gs} inherently dissolves the sign difference between P/NMOS's v_{gs} and results in the same current flow di-

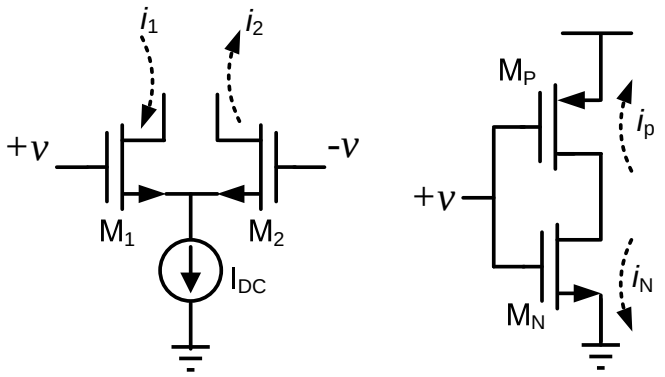


Figure 4.7: Analogy of IM_2 removal with fully-differential circuit

rection [1]. An analogy to this feature is illustrated in more details in the next schematic. An ideal differential pair takes $i_1 - i_2$ as the output and is known to be even-order-distortion-free. With a single-ended input applied to an inverter pair, both PMOS and NMOS are driven in a way similar to the differential-pair transistors. Again the output is taken in the form of $i_P - i_N$, therefore the second-order distortion in the inverter-pair output should cancel as well, assuming matched device parameters. In this way, C_{12} in Fig. 4.6 provides a path for NMOS's g'_m current to loop back through PMOS, or vice versa. In a matched case, i.e. identical g'_m of both transistors, the g'_m current will remain circulating within the NMOS-PMOS-C_{12} loop. The g'_m current is confined away from any of the resistors and result in zero second-order nonlinear voltage at V_x and V_1. Referring back to Eq. 4.10, with zero $A_2(s_1, s_2)$ and $B_2(s_1, s_2)$, the third $V_{out,3^{rd}}$ is eliminated even with finite g'_m of the common-source transistors.

[1] A similar and independent work based on this property can be found in [68]

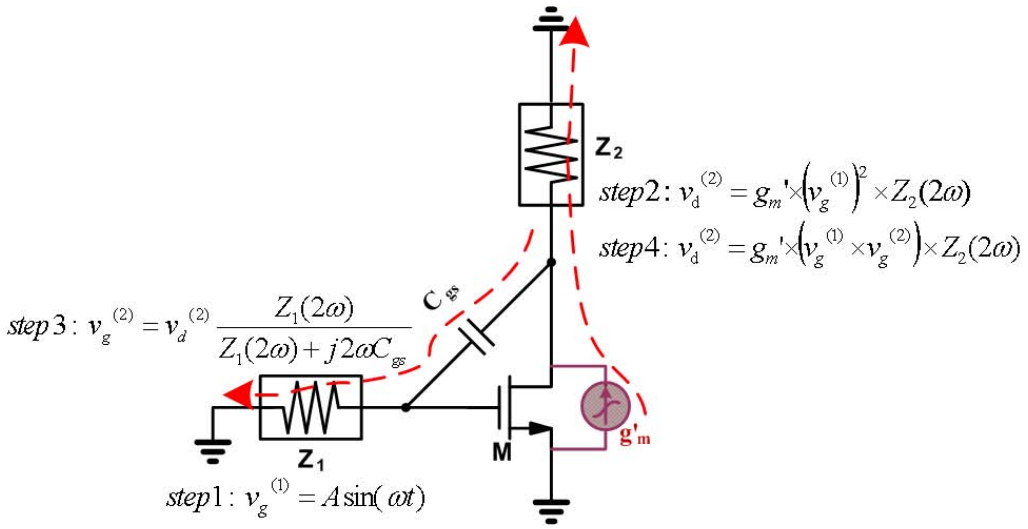


Figure 4.8:

4.2.3 Other nonideal cancelation

Our discussion thus far assumes that the common-source transistors contribute to output distortion by "feedforward" multiplying V_1 and V_x voltages with their g_m and derivatives, i.e. unilateral operation is assumed. In practice, even with perfectly linear voltage at V_x and V_1, the finite g'_m of the common-source transistors produces second-order nonlinear current, $g'_m \times v_{gs}^2$, which reappears at the gate through overlap capacitor C_{gd}. The finite g'_m further mixes the feedback second-order nonlinear voltage at the gate with the input linear voltage and results in third-order nonlinear voltage at output. Such $V_{out,3^{rd}}$ generation mechanism is illustrated in steps in Fig. 4.8. This second-order interaction is similar to the previous one as denoted in Eq. 4.10 except that it arises from the local feedback through finite g'_m and C_{gd}. Taking this into account, IIP_3 of the common-source stage is modified by an offset term [69]

$$IIP_3(2s_1 - s_2) \propto \frac{1}{\varepsilon((s_1 - s_2), (s_1 + s_2))} \tag{4.11}$$

$$\varepsilon((s_1 - s_2), (s_1 + s_2)) = g''_m - \frac{2g'^2_m}{3}(2K(s_1 - s_2) + K(s_1 + s_2)) \tag{4.12}$$

where $K(s)$ is the network function depending on the impedance at the source and drain of the transistor. Engineering the impedance at source and drain nulls out this additional $IM3$ [69, 70]. For example, the cascode transistor M_5 in Fig. 4.4 provides a low impedance at all frequencies and help alleviates this residual effect.

4.3 Noise Consideration of $MGTR$

The noise analysis presented in chapter 3 considers only the drain current thermal noise of transistors in the saturation region. As $MGTR$ scheme utilizes at least one transistor in weak inversion region, e.g. M_3 in Fig. 3.2, its noise characteristics differ from that of transistors in saturation, e.g. M_4. Moreover, the induced-gate noise, the noise current by the random fluctuation of potentials along the channel across the oxide, is not included either. These two effects are likely to alter the design considerations and deserves more investigation.

4.3.1 Drain current thermal noise revisit

As shown in Appendix A, an universal expression of the power spectral density of drain current thermal noise at all levels of channel inversion is

$$\overline{i^2_{nd}} = 4kt\frac{\mu}{L^2}Q_{inv} \tag{4.13}$$

where μ is the carrier mobility, L is the channel length and Q_{inv} is the total inversion layer charge [3]. Plugging in Q_{inv} at the saturation region,

$$Q_{inv} = \frac{2}{3} WLC_{ox}(V_{gs} - V_{th}) \tag{4.14}$$

, we derive output noise as

$$\overline{i_{nd}^2} = 4kt(\frac{2}{3})g_m \tag{4.15}$$

which matches the well-known expression for drain current thermal noise [34]. On the other hand, Q_{inv} in weak inversion is shown to be [71]

$$Q_{inv} = \frac{L^2}{2\mu\frac{KT}{q}} I_{ds} \tag{4.16}$$

With $g_m = \frac{q}{KT} I_{ds}$, this in turn gives

$$\overline{i_{nd}^2} = 2qI_{ds} = 4kt(\frac{1}{2})g_m \tag{4.17}$$

It is intriguing to discover that although thermal noise is assumed in the above analysis, Eq. 4.17 appears as if a different type of noise, shot noise, is assumed. Shot noise is the noise associated with DC current flow across a PN junction and is typically used for modeling BJT noise current. The behavior resemblance between MOSFET in weak inversion and BJT leads to the same noise expression. Comparing Eq. 4.15 and Eq. 4.17 and given the same g_m, the drain current thermal noise in weak inversion is smaller than in strong inversion. Since our noise optimization is based on the noise expression in saturation region and have already neglected M_3 due to its low value, the analysis result of Eq. 3.28 still holds.

[3] neglecting high-order effects such as mobility degradation, velocity saturation and short-channel effects

4.3.2 Induced-gate noise

The drain current thermal noise originates from the random movement of the carrier charges as they move along the channel. The random fluctuations of the potential in the channel are coupled to the gate terminal through the oxide capacitance, and a minute noise current is induced at the gate terminal, called "induced-gate" noise [72]. A quick inspection suggests that with small g_m, more thermal noise voltage is produced in the channel and in turn induces more charges on the other side of the oxide capacitor. Since the induced current flows through the capacitor, its magnitude is frequency dependent. These are verified in Eq. A.15 and Eq. A.1, the induced-gate noise in strong and weak inversion, which reconcile on the following expression.

$$\overline{i_{ng}^2} \propto \frac{\omega C_{gs}}{g_m} \tag{4.18}$$

Because of the inverse proportion to transconductance g_m, gate-induced noise is more pronounced if transistors are biased in weak inversion where g_m is low. The schematic of Fig. 3.2 is redrawn in Fig. 4.9 with added notation for induced-gate noise. M_1's gate noise is disregarded because of its common-gate configuration shunting the gate noise to the ground. The g_m ratio between M_3 and M_4 is roughly confined by

$$\frac{g_{m4}}{g_{m3}} \approx \frac{R_1}{R_s} \tag{4.19}$$

and the value is in the range of 2-5 in practice. It is thus more preferable to allocate M_3 as $MGTR$'s subthreshold transistor. The reasons are two fold. First, the severe noise figure degradation by M_3's induced-gate noise is suppressed by the voltage gain ratio $\frac{R_1}{R_s}$. Secondly, M_4's induced-gate noise is in shunt with the R_s noise and directly impacts overall

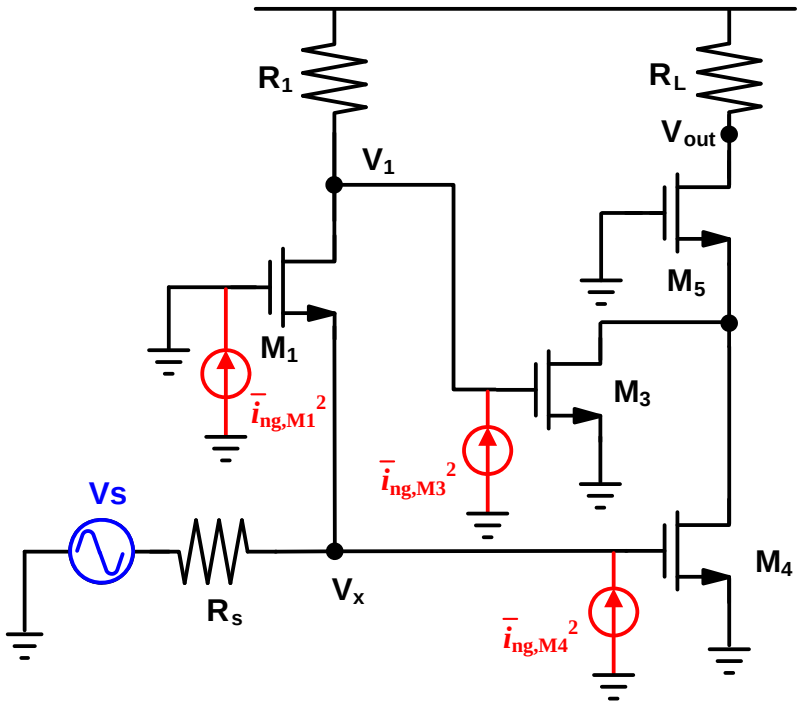


Figure 4.9: LNA schematic with induced-gate noise

noise figure. Furthermore, only some of M_4's gate noise is related to its drain noise current. The uncorrelated gate noise degrades the overall LNA noise factor by

$$\begin{aligned} EF_{M_{4,ng,uncorrelated}} &= \frac{\overline{i^2_{ng,M_4}}}{\overline{i^2_{n,R_s}}} \\ &= \frac{\frac{4kt\omega^2 C_{gs}^2}{g_{m4}} \frac{16}{135} \left(1-|c|^2\right)}{4kt\frac{1}{R_s}} \\ &= g_{m4} R_s \frac{16}{135} \left(\frac{\omega}{\omega_0}\right)^2 \left(1-|c|^2\right) \\ &\approx \frac{2}{3} g_{m4} R_s \frac{1}{5} \left(\frac{\omega}{\omega_0}\right)^2 \left(1-|c|^2\right) \end{aligned} \tag{4.20}$$

where ω_0 is the unity-gain frequency of transistor M_4. c is the correlation factor defined in Eq. A.17. $\frac{2}{3}$ can be viewed as the noise coefficient, γ, of the long-channel transistor. A factor of $\frac{1}{5}$ accounts for the alternative way of modeling $\overline{v_{ng}^2}$ with a series gate resistance of

$\frac{1}{5g_m}$. On the other hand, the correlated M_4 gate noise flows through resistor R_1 and adds to V_{out} noise after M_3 amplification. Because the phase of induced-gate noise is 90 degree apart from the drain current, the root-mean-square output noise power is the $r.m.s$ sum of both currents. This modify Eq. 3.16 with a factor of

$$EF_{M_{4,ng+nd,correlated}} = EF_{M_{4,nd}} \times \left(1 + \frac{8}{45}\left(\frac{\omega}{\omega_0}\right)^2 \left(\frac{g_{m4}R_s}{(1+\delta)(2+\varepsilon_{rr})}\right)^2 |c|^2\right) \tag{4.21}$$

For typical values of g_{m4}, δ and ε_{rr}, Eq. 4.21 and Eq. 4.20 simplify to

$$\begin{aligned} EF_{M_{4,ng,uncorrelated}} &\approx 0.4\left(\frac{\omega}{\omega_0}\right)^2 \\ EF_{M_{4,ng+nd,correlated}} &\approx EF_{M_{4,nd}} \times \left(1 + 0.84\left(\frac{\omega}{\omega_0}\right)^2\right) \end{aligned} \tag{4.22}$$

It is clear that the effect of induced-gate noise needs to be considered only if the LNA operates at high frequencies.

Chapter 5

A 0.8-2.1 GHz Prototype

A broadband 0.8-2.1 GHz LNA implemented in 0.13μm CMOS technology is used to verify the design of noise and distortion cancelation outlined in the previous chapters. The target frequency range is for multi-mode multi-band cellular applications. The LNA is built on standard-V_t transistors without linearity enhancement such as thick oxide devices. The only RF process option used is Metal-Insulator-Metal capacitors due to their high linearity and high capacitance density.

5.1 Implementation

Fig. 5.1 shows the complete schematic of the implemented LNA core circuit. It consists of a common-gate P/NMOS pair as the input stage, followed by two common-gate NMOS transistors. The input P/N pair reuses the same DC bias current to save power consumption. AC-coupling capacitor, C_{12}, shunts the output of the common-gate transistors to achieve ac equivalence for both nodes. To allow for flexible bias tuning during prototype

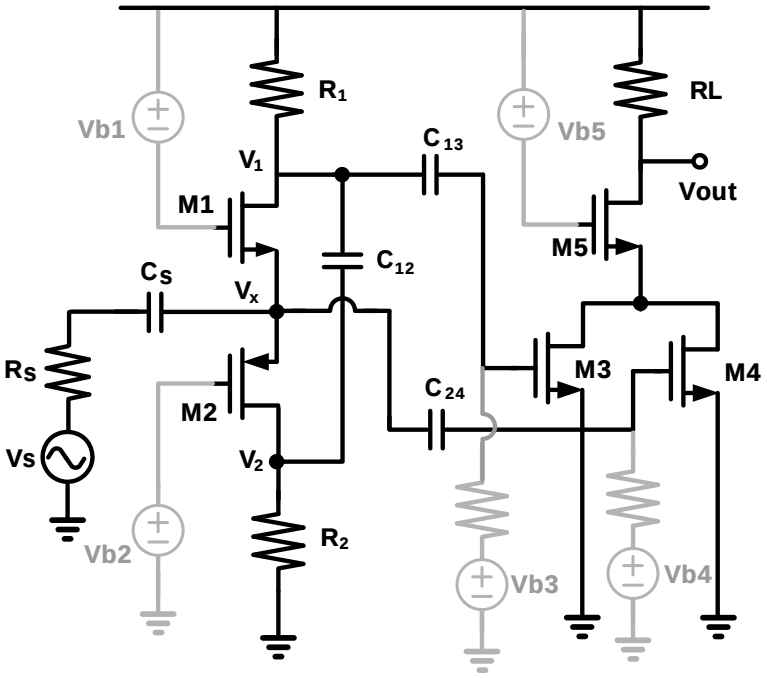


Figure 5.1: Complete LNA schematic

characterization, each transistor is biased independently through an external voltage source connected through on-chip diode-connected bias branch. AC coupling between stages is implemented with the aid of capacitors C_{13}, C_{24} and C_s. Low-value silicide poly resistor is chosen for the resistors used in the schematic. Since a very high IIP_3 is pursued, adding an explicit output buffer for 50 Ω measurement setup is avoided. Otherwise the measured linearity will be colored by the intrinsic distortion of the buffer circuit itself. Therefore a very low load resistance of 55 Ω is adopted to facilitate the linearity measurement. This adversely reduces amplifier's voltage gain and increased R_L noise contribution. In the fully integrated front-end, the LNA output is routed directly to the input of the mixer so no 50 Ω matching is required and R_L value can be made higher. On the other hand, low

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

Table 5.1: Device size

resistor value helps "desensitize" the distortion caused by transistor's drain conductance, an important underlying assumption of our distortion cancelation analysis. By attaching the probe onto the output pad, the effective output impedance is halved, thus measured S_{21}, or power gain, is 6 dB lower than on-chip voltage gain. Since R_L is already sufficiently small, in the simulation the IIP_3 with and without attaching the probe does not show measurable difference.

Transistors M_1 to M_5 are implemented by assigning different number of fingers to the unit-width device, which has a dimension of $6.52\mu m/0.12\mu m$. Double-sided gate contacts are used to minimize the parasitic resistance of the poly gate material. Care is also spent to the placement of proper RC filtering and routing of the bias lines so that the gate bias is connected to a clean DC voltage. Table 5.1 summarizes the device size. The resistance value is slightly lower than what has been used in the previous analysis. This is because transistor's intrinsic gain product, $g_m r_o$, is low in the deep submicron technology and R_1, R_2 need to be reduced to maintain the same input matching. Fig. 5.2 shows the Spectre simulation of individual noise factor by sweeping Vb_4 after device size and other biases are chosen. This plot is similar to Fig. 3.5 except that g_{m3} is held constant. M_1

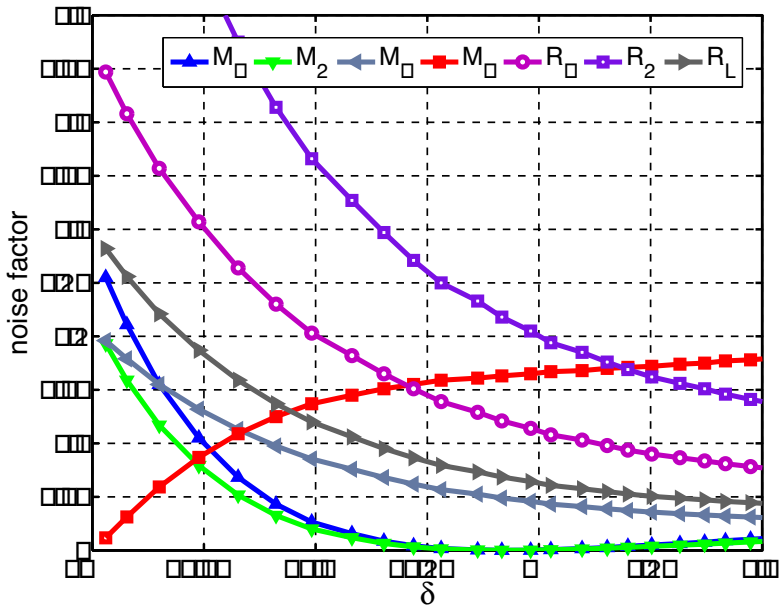


Figure 5.2: Noise factor by Spectre simulation

and M_2 noise is canceled at $\delta = 0$ as expected. The smaller value of R_1 and R_2 raised the non-canceling F_{R1} and F_{R2} even higher than F_{M4} in the region of interest. It also causes higher F_{RL} and F_{M3} due to reduced common-gate gain. This drives the selection of a larger δ than predicted in chapter 2. For a design with 2 dB LNA noise figure, a snapshot of the noise breakdown is shown in Fig. 5.2.

The amplifier's stability is examined by the K factor, defined in [73] as

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |S_{12}S_{21} - S_{11}S_{22}|^2}{2|S_{12}S_{21}|^2} \tag{5.1}$$

The simulated K value including the bias network is larger than one for frequencies up to 10 GHz.

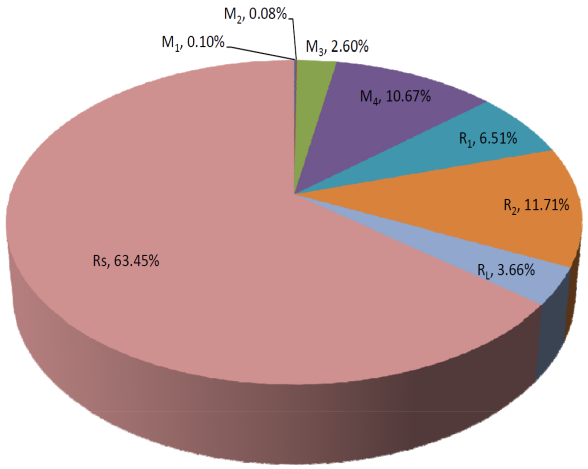


Figure 5.3: Output noise breakdown

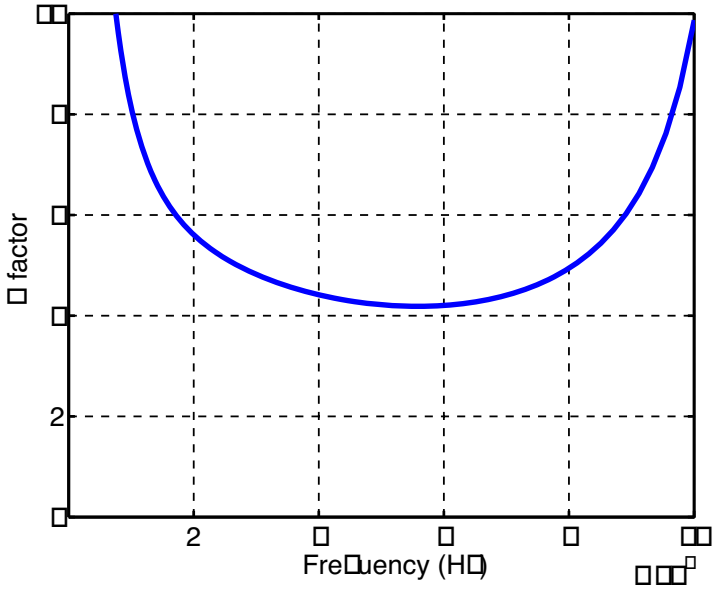


Figure 5.4: Simulated LNA stability factor

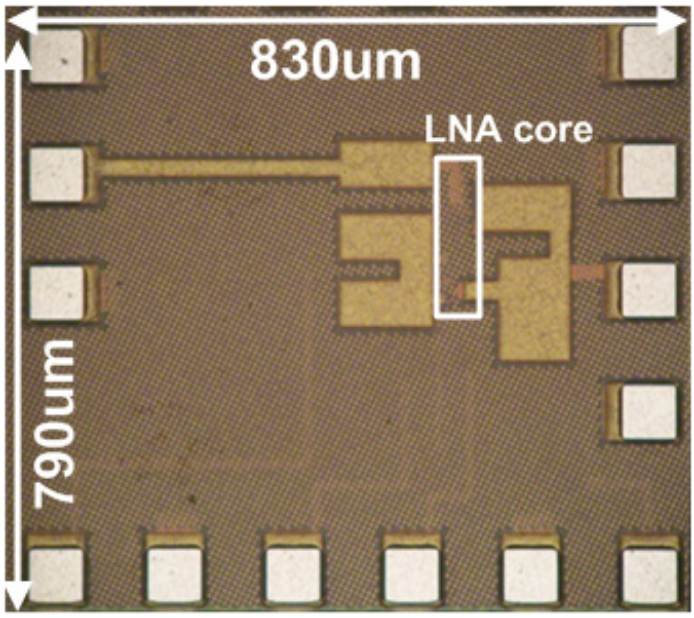


Figure 5.5: Chip microphotograph

The chip microphotograph is shown in Fig. 5.5. It occupies an area of $830 \times 790 \mu m^2$. The die size is mainly limited by the surrounding pads for testing purpose. The area of core circuit is only $320 \times 310 \mu m^2$ including the MIM capacitors.

5.2 Measurement Results

The chip measurements is carried out by on-wafer probing using Cascade Microtech 12000 probe system. The input and output impedance matching is characterized by S-parameters measurements with Anritsu's 37397C Vector Network Analyzer. Fig. 5.6 shows better than $-8.5\ dB$ of S_{11} and S_{22} is achieved and matches with post-layout simulation. S_{11} impedance matching can be improved by including the bond wire inductor, which typically has a self-inductance of 1-2 nH, to tune out the parasitic capacitance. The measured low

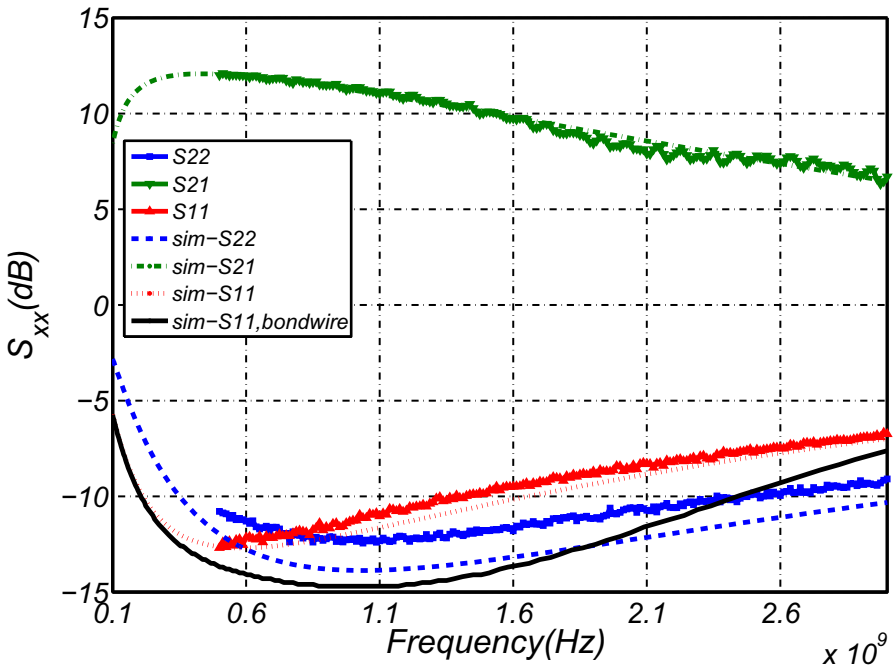


Figure 5.6: Measured S parameters

frequency power gain, S_{21}, is 11 dB, which amounts to 17 dB voltage gain on the chip. The 3 dB bandwidth is 2.1 GHz.

The noise figure measurement is performed with Agilent's N8974A Noise Figure Analyzer. In addition, the output of the LNA is connected to a discrete broadband amplifier before it reaches the input of the test equipment. The purpose of this modified setup is to bring up the signal and noise level of the DUT and to minimize the error introduced by the measurement setup [74]. The gain and noise figure of the discrete amplifier, as well as loss of the cables in use, are measured separately before it is hooked up to the probing system. The LNA's own noise figure is then extracted from the measured noise figure by applying

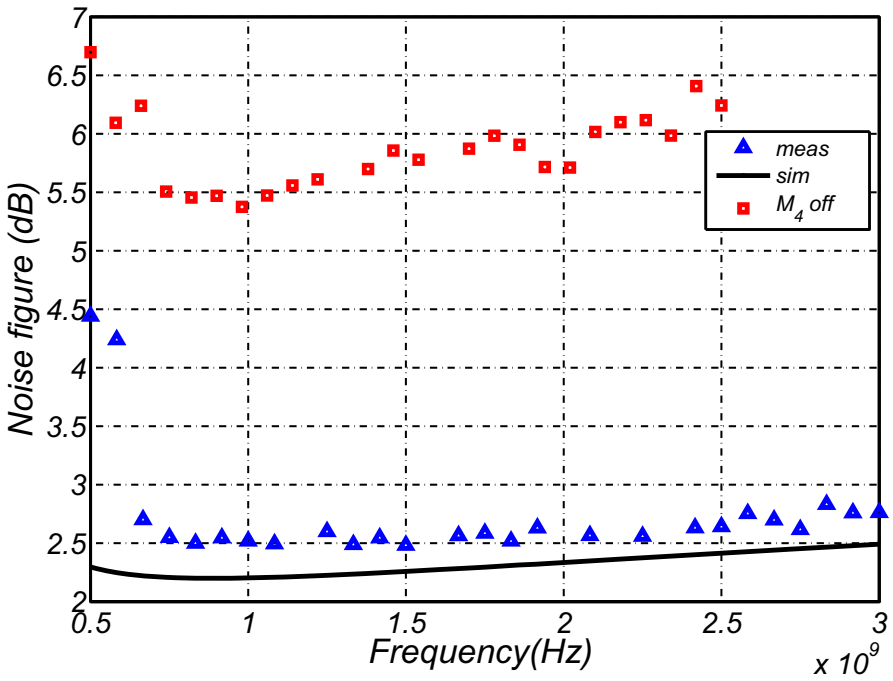


Figure 5.7: Measured noise figure

the calibration data into the Friis equation. The results are shown in Fig. 5.7. Compared to the case without noise cancelation, i.e. M_4 off, as large as 3 dB improvement in noise figure is achieved.

The IIP_3 of the LNA is evaluated by exciting 2 tones at ω_1 and ω_2, and by reading the spectrum output at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ with Agilent's E4440A spectrum analyzer. To begin with, a WCDMA compliant blocker test is used to determine the value of ω_1 and ω_2. Two sinusoidal tones located at 1.76 GHz and 1.95 GHz model a strong AM blocker and a transmitter on-chip leakage. Through amplifier's third-order nonlinearity, the combination of $2\omega_2 - \omega_1$ results in an IM_3 product at 2.14 GHz, which falls onto the desired receiver band. The same frequency setting is reused for testing at 900 MHz band. Fig. 5.8 shows

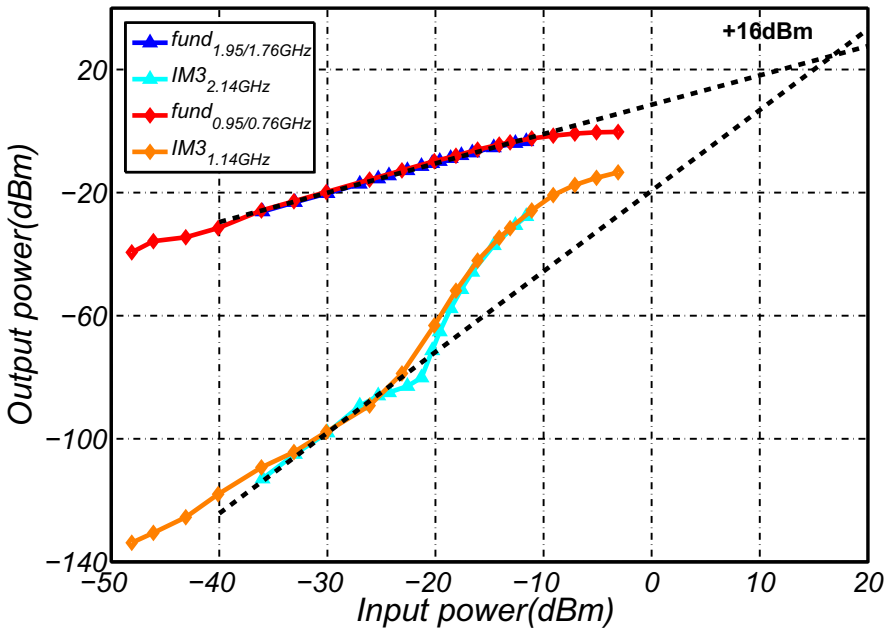


Figure 5.8: Measured IIP_3

the results. The IM_3 in both bands are very similar to each other, which confirms the cancelation scheme is indeed effective over wide frequency range. The extrapolated IIP_3 of both bands achieves as high as +16 dBm. Moreover, the IM_3 cancelation scheme holds for the blocker power as large as −20 dBm. This greatly relaxes the isolation requirement imposed on the duplexer ahead of the LNA and power amplifier. On the other hand, the measured P_{1dB} is −12 dBm. The difference between measured IIP_3 and P_{1dB} is much larger than 9.6 dB. This is because the theoretic calculation of P_{1dB} considers third-order coefficient a_3 only. While the distortion cancelation strives to cancel a_3, higher-order coefficients such as a_5 and a_7 are not taken into account, which also plays a significant role

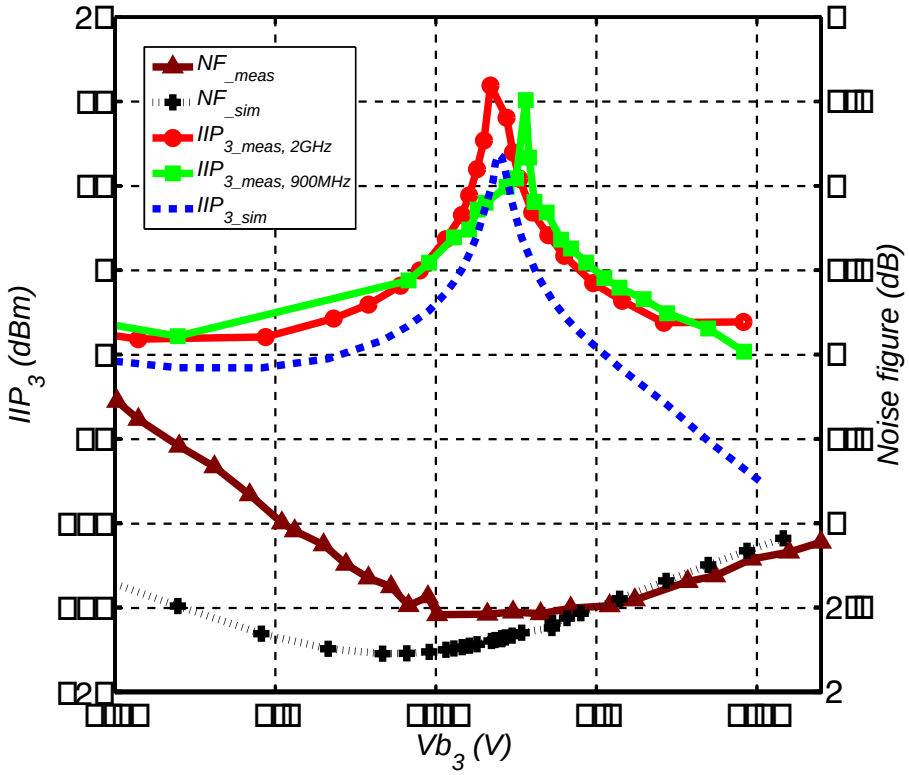


Figure 5.9: Measured and simulated IIP_3 and NF sensitivity

in the large signal regime.

It is important to examine the bias sensitivity of the distortion and noise cancelation schemes. One way to emulate the effects of matching and threshold variation is to vary one of the bias point, say, Vb_3. Fig. 5.9 shows results of such measurement and compares with the post-layout simulation. The aligned IIP_3 peak and NF valley confirm simultaneous noise and distortion cancelation. Reasonable agreement is also observed between the simulation and the measurement for both noise figure and IIP_3. Due to the limited accuracy of higher-order g_m derivatives in the device model, the simulated IIP_3 does not track the measured data exactly. On the other hand, discrepancy of NF in the lower Vb_3 region

is noticeable. As Vb_3 decreases and biases transistor M_3 into deep subthreshold region, this drives the NF contour to the bottom in Fig. 3.3, where rapid gradient of the contour curves makes NF more susceptible to parameter variations as well as modeling inaccuracy. Allowing a bias window of 50 mV for worst-case process and bias variations, performance of IIP_3 greater than +5 dBm and NF around 2.5 dB can still be maintained.

The parameter variation of the common-gate stage has less sensible impacts on the overall IM_3 cancelation as only one term, g'_m, in Eq. 4.10 is affected. This is confirmed by the broader bell shape of the curves in Fig. 5.10, where bias voltage of common-gate transistors is varied. The noise figure under the same bias sweep was shown in Fig. 5.11. The noise figure remains almost constant regardless of M_1 and M_2's bias condition. Again this clearly proves the effectiveness of noise cancelation.

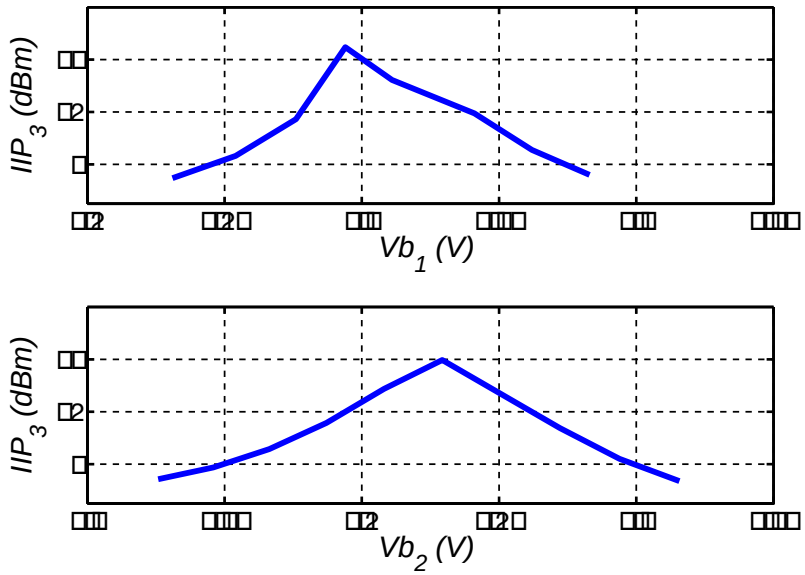


Figure 5.10: Measured IIP_3 sensitivity of Vb_1 and Vb_2

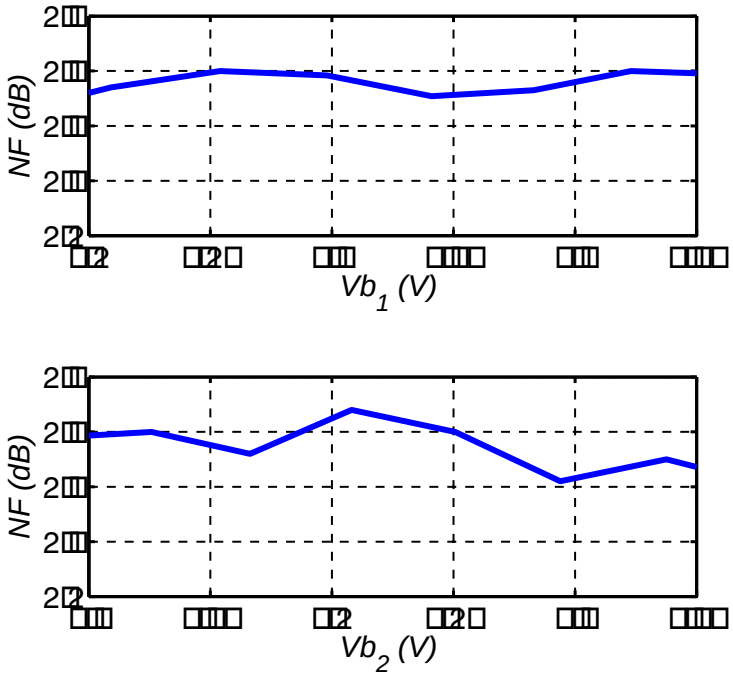


Figure 5.11: Measured NF sensitivity of Vb_1 and Vb_2

5.3 IIP_3 two-tone-spacing dependence

The measured IIP_3 is found to be dependent on the frequency spacing between the injected two tones, i.e. $\omega_1 - \omega_2$. This is exemplified in Fig. 5.12 where one tone was fixed at 1.95 GHz and 0.95 GHz, respectively, while the frequency of the other tone varies. This 2-tone frequency dependence originates from the insertion of capacitor C_{12}. Although C_{12} helps direct g'_m nonlinear current to circulate in the loop of M_1-M_2-C_{12}, this is only true for IM_2 current at high frequencies, i.e. $\omega_1 + \omega_2$. With input 2 tones at ω_1 and ω_2, an IM_2 current at $\omega_1 - \omega_2$ also forms which sees high impedance presented by C_{12}. This lower-band IM_2 current is blocked and has to flow instead through resistors R_1 and

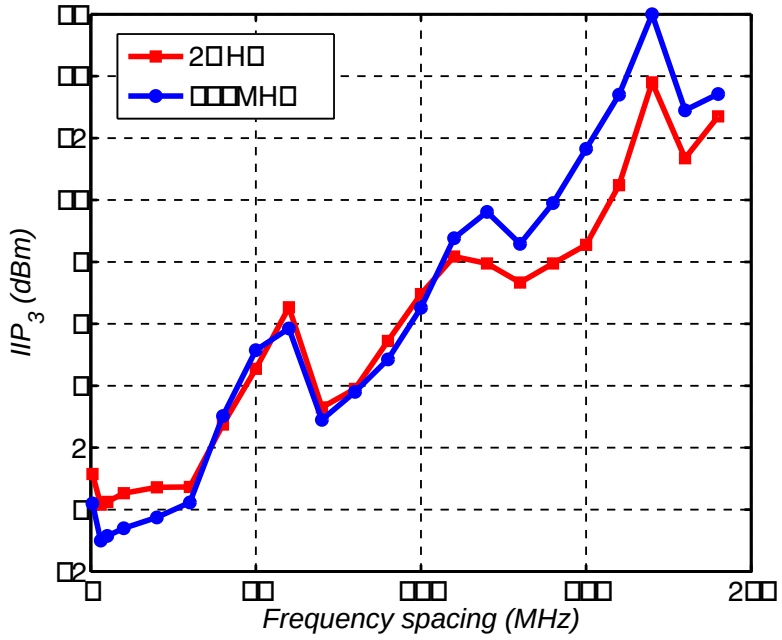


Figure 5.12: Measured IIP_3 dependency on the two-tone spacing

R_2, even if g'_{m1} and g'_{m2} are matched. The low-frequency IM_2 current thereby establishes non-negligible second-order nonlinear voltages at V_1 and V_x that leads to finite third-order output distortion through M_3 and M_4's g'_m. In Fig. 5.13, a *Spectre* simulation of the second-order nonlinear voltage at V_1 with different 2-tone frequency spacing confirms this dependence. With practical capacitor value of, say, 15 pF, the LNA IIP_3 still shows strong dependence on the tone spacing in the ten's of MHz range. If 15 pF is replaced with an infinite capacitance, V_1 at $\omega_1 - \omega_2$, and output IIP_3, stay constant. This phenomenon is analyzed in Appendix B with a detailed Volterra series analysis taking into account the role of C_{12}. According to Eq. B.20 and Eq. B.21, the second-order Volterra series kernel, $A_2(s_1, s_2)$ and $B_2(s_1, s_2)$ consists of a term $\triangle A_2(s_1, s_2)$ and $\triangle B_2(s_1, s_2)$, respectively, which

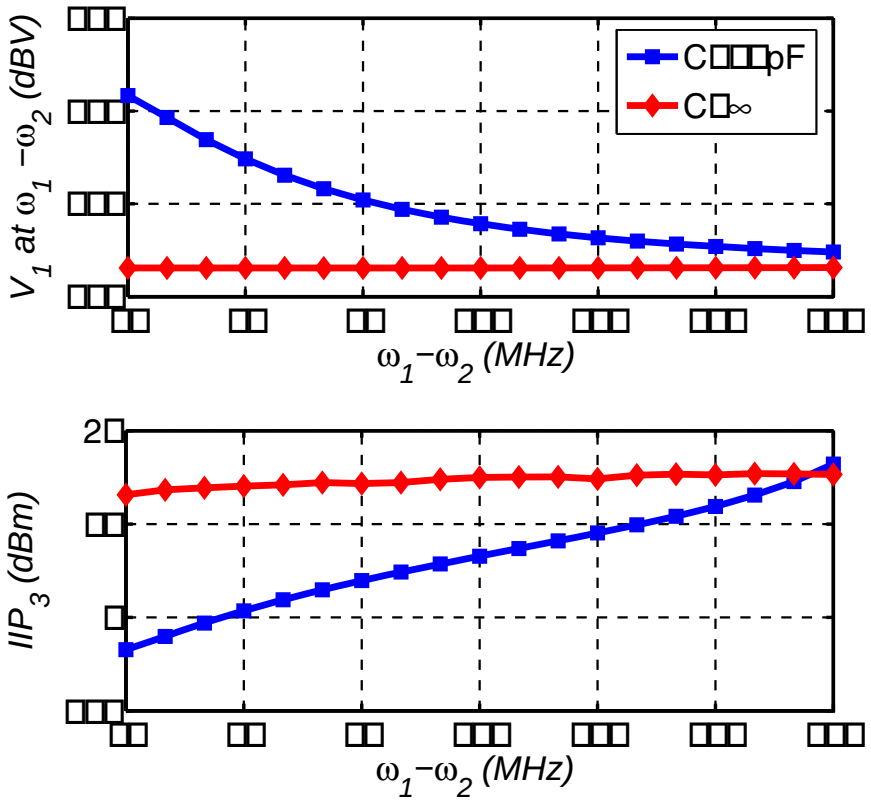


Figure 5.13: Effect of 2^{nd}-order distortion on LNA IIP_3

are associated with $C_{12}(s_1, s_2)$ and, more importantly, independent of $g'_{mn} - g'_{mp}$. In other words, even if g'_{mn} is matched with g'_{mp}, there is still a residual A_2 and B_2 if $C_{12}(s_1, s_2)$ is not zero. To mitigate this frequency dependence, the coupling capacitor has to maximize so that a sufficiently low impedance at low IM_2 frequency is realized. This introduces more unwanted parasitic capacitance associated with C_{12} and reduces the LNA 3 dB bandwidth. The selection of C_{12} thus depends on the frequency range and the blocker profile of the LNA. Fortunately in today's typical multi-radio scenario, it is the out-of-band blockers which set the linearity constrains. They are usually located on the spectrum a few tens or hundreds

of MHz away from each other and the desired receiving channel. The required capacitor size is greatly reduceds.

5.4 Discussions

We have demonstrated the measured results of a state-of-the-art low noise figure and exceptionally high IIP_3 for a 0.8 to 2.1 GHz noise and distortion canceling LNA. In order to extend the LNA bandwidth to accommodate upper GHz frequencies, the parasitic capacitance associated with AC-coupling capacitors has to be minimized. Since capacitor C_1 and C_2 are to isolate the DC bias between the stages, they can be removed in order to cut the parasitic capacitances. For example, attaching a bleeding current source at the gate of M_3 and M_4 allows to change M_3 and M_4 bias without affecting the previous stage. A more elaborate approach is by taking advantage of the triple-well technology typically available in nano-scale CMOS technology. Connecting the body bias to an independent DC voltage, transistor's g_m'' can be altered by varying the body bias. To the first order, the bulk potential changes the threshold voltage V_{TH} by

$$V_{TH} = V_{TH0} + \gamma \left(\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F} \right) \tag{5.2}$$

V_{TH0} is the threshold voltage with $V_{SB} = 0$, γ is the body effect factor, and ϕ_F is the Fermi level. Shown in Fig. 5.14, a V_{SB} of $-0.5\ V$ shifts g_m'' curve by 120 mV. Considering the bias difference between typical MGTR transistors is less than 200 mV, this amount of g_m'' shift is capable of aligning M_4's g_m'' valley with M_3's peak in Fig. 5.15 using the bias voltages established by M_2's V_{DS} and R_2 only.

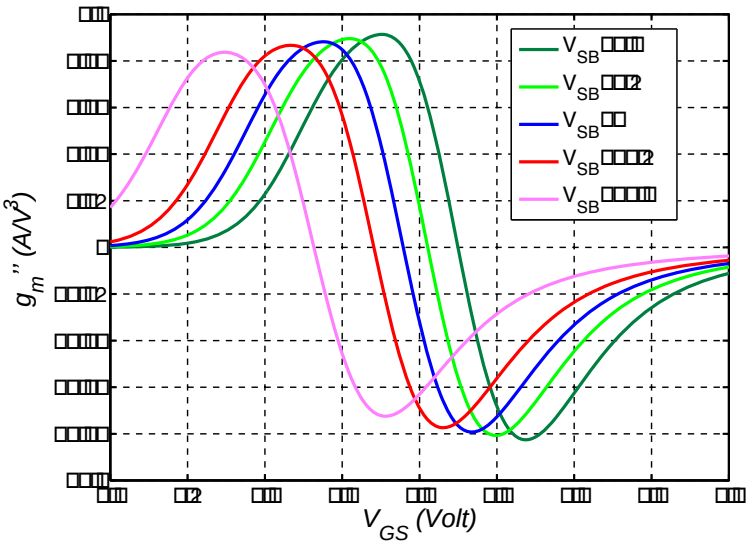


Figure 5.14: g_m'' by changing body bias

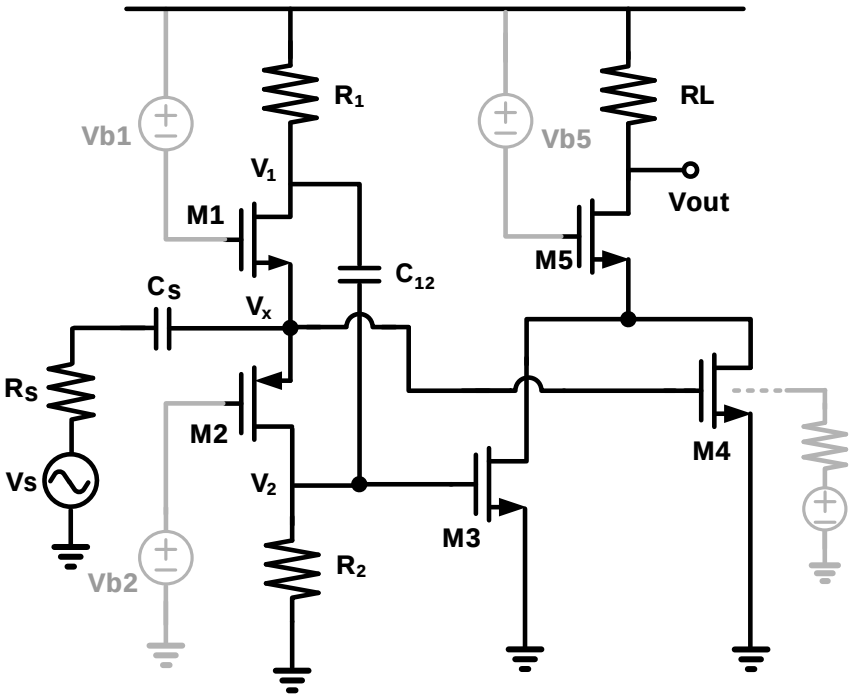


Figure 5.15: Modified LNA schematic with C_{12} only

Chapter 6

Large-signal Distortion Characterization

Taylor series has been widely adopted for distortion characterization. It facilities hand calculation and leads to many insightful expressions such as IIP_3 as defined in chapter 2. Theoretically a Taylor series with infinite number of coefficients is able to predict the output distortion without lost of information. In practice, the higher-order coefficients are truncated and only the first few are adopted. A consequence due to exclusion of higher-order terms (HOT) is the reduced accuracy, especially of concern when the signal magnitude becomes large. Mathematically speaking, function approximation by Taylor series hold valid as long as signal magnitude is within the radius of convergence. Although CAD programs are effective in simulating large-signal distortions, they often fail to provide insights toward circuit's nonlinear response. We therefore look for an alternative approach which can offer the design guideline more transparently than CAD algorithms and at the same maintain

more accuracy than conventional Taylor series.

6.1 Taylor Series Recap

To illustrate the limitation of conventional Taylor series approach, a Spectre simulation based on sweet spot and $MGTR$ schemes is conducted. The biases for both schemes are selected, by referring back to Fig. 5.8, so that their equivalent a_3 is the same and close to zero. Fig. 6.1 shows the output HD_3 as input magnitude increases. From Eq. 2.23,

$$s_{out}(3\omega_1) = \frac{a_3}{4} \times s_{in}(\omega_1)^3 \tag{6.1}$$

, which corresponds to the straight line in Fig. 6.1 with slope of 3 in dB-dB scale. Simulations of both schemes follow the slope of 3 well until v_{in} exceeds a certain level. The discrepancy at large signal is attribute to the higher-order coefficients not accounted for in Eq. 6.1. However, even in the small signal regime, there exhibits another discrepancy between sweet-spot and $MGTR$ at the same signal swing. Notice that this discrepancy is not foreseen by Eq. 6.1.

6.2 Signal-locus-based characterization

According to Eq. 2.21, successive derivatives, or approximations, of different orders are used to provide an estimate of the output. The coefficients for output approximation are fixed because the derivatives refer to the same static point. Static approximation is valid as long as circuit nonlinearity does not change along the course of input signal. This is usually true for small signal magnitude. As signal magnitude increases, the signal travels an appre-

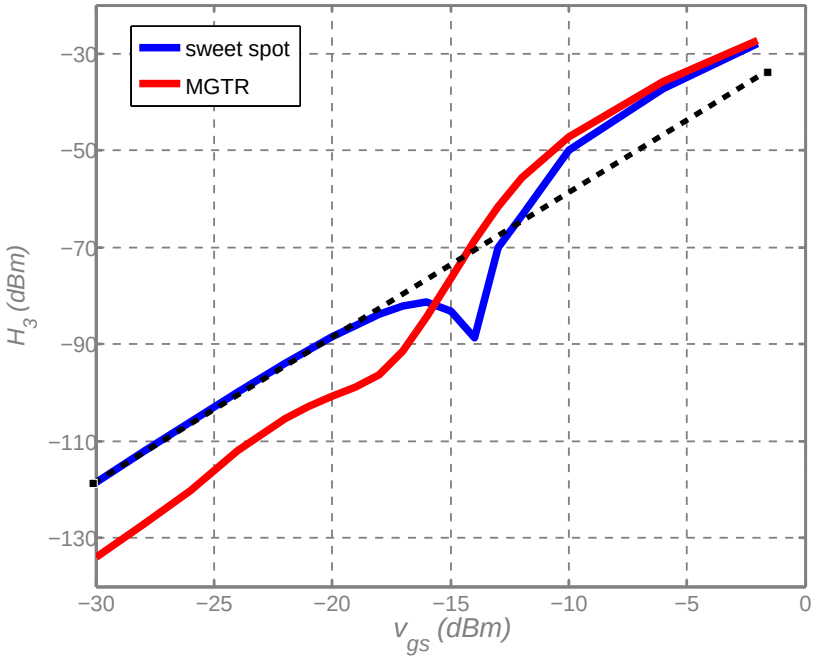


Figure 6.1: HD_3 by sweet spot, MGTR and Taylor series

ciable distance on the transfer curve and experiences changes of circuit nonlinearity. With reference point fixed, the approximation error accumulates up to the point where Taylor series representation becomes inaccurate. Including $HOTs$ in the polynomial representation improves the accuracy because the otherwise accumulated residual error continues to be estimated by the $HOTs$. However adding $HOTs$ makes hand calculation cumbersome. Taking a_3 as an example, the presence of coefficients higher than order of 3, say a_5 and a_7, causes changes of a_3 values across a wider input range. Alternatively speaking, the shape of lower-order coefficients is a reflection of the $HOTs$. Therefore even the value of the initial a_3 is equal, the different shape of a_3 of sweet-spot and $MGTR$ in Fig. 4.2 results in different third harmonic distortion.

To overcome the above limitation, the fixed reference point is discarded and replaced with continuous update along the course of signal swing [75]. As signal moves along the transfer curve, the instantaneous derivative values are used to provide the output estimation for next step. As long as the distance between two adjacent steps are sufficiently small, the output approximation remains accurate even with first-order linear projection only. This is referred to as signal-locus-based characterization because it always takes on the most up-to-date coefficient value.

6.2.1 Model derivation

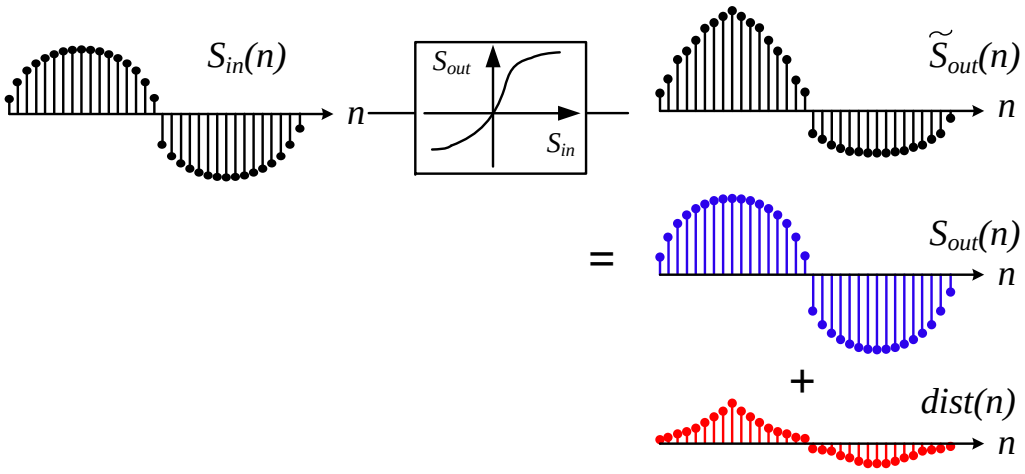


Figure 6.2: Representation of output distortion

Fig. 6.2 illustrates the setup for deriving a quantitative expression of the proposed approach. Assume a memoryless nonlinear system driven by a sinusoidal input with

amplitude A sampled by N times per cycle,

$$\begin{aligned} s_{in}(n) &= A \times \sin(\frac{2\pi n}{N}), n = 0, 1, ..., N-1. \\ d(n) &= s_{in}(n) - s_{in}(n-1) \end{aligned} \tag{6.2}$$

$d(n)$ is the input step between two successive sampling instants. Due to system nonlinearity,

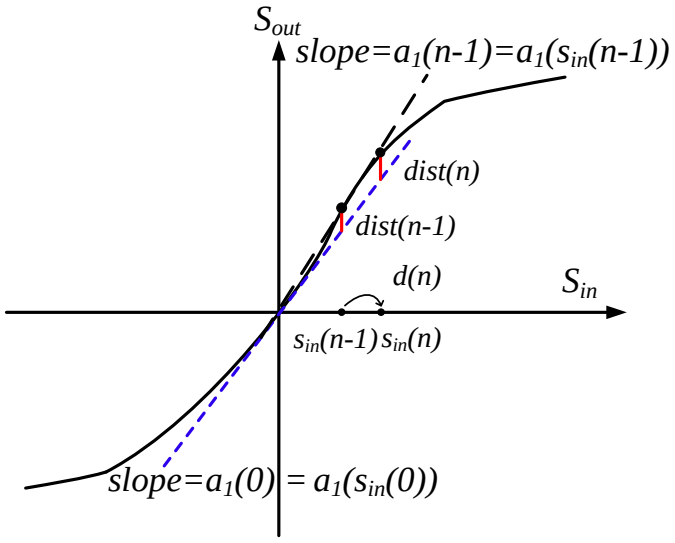


Figure 6.3: Close-in on the output estimation

the output sequence, $\tilde{s}_{out}(n)$, is distorted from the linear amplification of initial input, shown as

$$\tilde{s}_{out}(n) = s_{out}(n) + dist(n) \tag{6.3}$$

Referring to Fig. 6.2 and if the system is linear, the output $s_{out}(n)$ is just an extension of $s_{in}(n)$ on the straight projection along $a_1(0)$ line. It is clear that $a_1(n-1)$ better estimates $\tilde{s}_{out}(n)$ than $a_1(0)$ at instant n-1. By writing

$$s_{out}(n) = s_{out}(0) + a_1(0) \times \sum_{i=1}^{n} d(i) \tag{6.4}$$

and

$$\begin{aligned}\tilde{s}_{out}(n) &= \tilde{s}_{out}(n-1) + a_1(n-1) \times d(n) \\ &= \tilde{s}_{out}(n-2) + a_1(n-2) \times dx(n-1) + a_1(n-1) \times d(n) \\ &= \tilde{s}_{out}(0) + \sum_{i=1}^{n} a_1(i-1) \times d(i)\end{aligned} \tag{6.5}$$

distortion $dist(n)$ is expressed as

$$dist(n) = \sum_{i=2}^{n} (a_1(i-1) - a_1(0)) \times d(i) \tag{6.6}$$

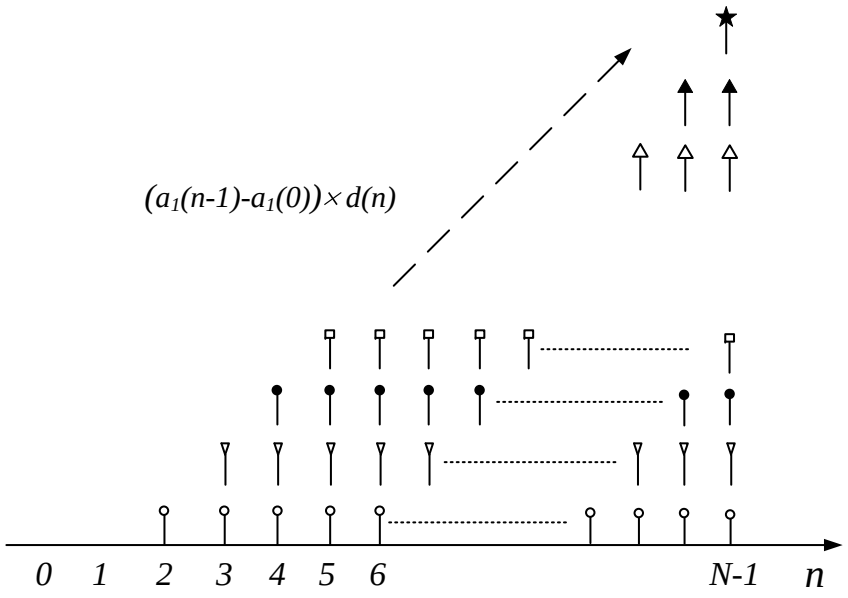


Figure 6.4: Composition of $dist(n)$

Fig. 6.4 illustrates the composition of the distortion for one full cycle. It consists of the sum of pulse sequences which have shrinking width and delayed occurrence. The pulse height is $(a_1(i-1) - a_1(0)) \times d(i)$. Borrowing the shorthand expression for a pulse

sequence from digital signal processing theory, the k^{th} output harmonic, H_k, is found as

$$H_k = \sum_{n=2}^{N-1} \left| h(n) \times e^{-j \cdot \frac{2\pi}{N} \cdot k \cdot \frac{n}{2}} \times \frac{\sin(\frac{2\pi}{N} \cdot k \cdot \frac{n+1}{2})}{\sin(\frac{2\pi}{N} \cdot k \cdot \frac{1}{2})} \right| \tag{6.7}$$

, where $h(n) = (a_1(n-1) - a_1(0)) \times d(n)$. The exponential, sinc, and $h(n)$ term account for the delayed occurrence, pulse width, and height respectively. Then define the third-order output harmonic distortion, HD_3 as

$$\begin{aligned} HD_3 &= \frac{H_1}{H_3} \\ &= \frac{\sum_{n=2}^{N-1} \left| h(n) \times e^{-j \cdot \frac{2\pi}{N} \cdot 3 \cdot \frac{n}{2}} \times \frac{\sin(\frac{2\pi}{N} \cdot 3 \cdot \frac{n+1}{2})}{\sin(\frac{2\pi}{N} \cdot 3 \cdot \frac{1}{2})} \right|}{\sum_{n=2}^{N-1} \left| h(n) \times e^{-j \cdot \frac{2\pi}{N} \cdot 1 \cdot \frac{n}{2}} \times \frac{\sin(\frac{2\pi}{N} \cdot 1 \cdot \frac{n+1}{2})}{\sin(\frac{2\pi}{N} \cdot 1 \cdot \frac{1}{2})} \right|} \end{aligned} \tag{6.8}$$

6.2.2 a_3-only expression

$dist(n)$ in Eq. 6.6 is specified by first-order coefficient, a_1. It is of our interest to rewrite it in terms of a_3 as our linearization schemes utilize a_3 properties. Apply power series approximation with

$$\begin{aligned} a_1(n) &= a_1(n-1) + a_2(n-1) \times d(n-1) \\ a_2(n-1) &= a_2(n-2) + a_3(n-2) \times d(n-2) \end{aligned} \tag{6.9}$$

After backward recursive replacement of Eq. 6.9 into Eq. 6.6, we arrive at an expression in terms of a_3,

$$dist(n) = a_2(0) \times \sum_{i=1}^{n-1} \left(d(i) \cdot \sum_{j=i+1}^{n} d(j) \right) + \sum_{i=1}^{n-2} \left(a_3(i-1) \cdot d(i-1) \cdot \sum_{j=i+1}^{n-1} \left(d(j) \cdot \sum_{k=j+1}^{n} d(k) \right) \right) \tag{6.10}$$

and $a_2(0)$, which is the second-order nonlinear coefficient at the initial point. The role of $a_2(0)$ is examined by applying sinusoidal input $s_{in}(n) = A\sin(\frac{2\pi n}{N})$ and the term associated with $a_2(0)$ in Eq. 6.10 becomes

$$dist(n)\Big|_{a_2(0)} = \frac{1}{4}A^2\big(1 - \cos(2\pi \cdot \frac{2n}{N})\big) \tag{6.11}$$

It is a distortion component with two times the input signal frequency and is neglected in the subsequent discussions on third-order distortion.

6.2.3 Verification

Eq. 6.8 is verified by a behavioral simulation in Matlab. An empirical MOSFET drain current function valid from subthreshold to strong inversion region is used [76, 77]:

$$I_{ds} = \frac{K}{2} \times \frac{f^2(V_{gs})}{1 + \theta \cdot f(V_{gs}) + \vartheta \cdot f^2(V_{gs})}$$
$$f(V_{gs}) = 2\eta\phi_t ln(1 + e^{(V_{gs}-V_{th})/2\eta\phi_t}) \tag{6.12}$$

where the values of K, θ, ϑ, $2\eta\phi_t$ and V_{th} are extracted from 0.13 μm technology. They are 61.87 m, 6.868, 114.8 m, 80.1 m and 0.478 respectively. First, expressions of g_m and its derivatives are derivated based on Eq. 6.12. Their values are computed by plugging in the parameter values with the DC bias set from 0.5 to 0.6 volt and the input amplitude ranging from 5 mV to 500 mV. HD_3 is then directly calculated using Eq. 6.8. For comparison, the same sinusoid is fed to Eq. 6.12 to get the simulated drain current waveform. The spectral results of the waveform are found by taking its FFT. Fig. 6.5 compare both results. The peak of the curves corresponds to the sweet spot where the third-order nonlinear coefficient crosses the zero value. It starts deviating from the initial sweet spot as input magnitude

increases. The shift is to compensate the impacts of $HOTs$. An excellent agreement is achieved even for input amplitude as large as 500 mV.

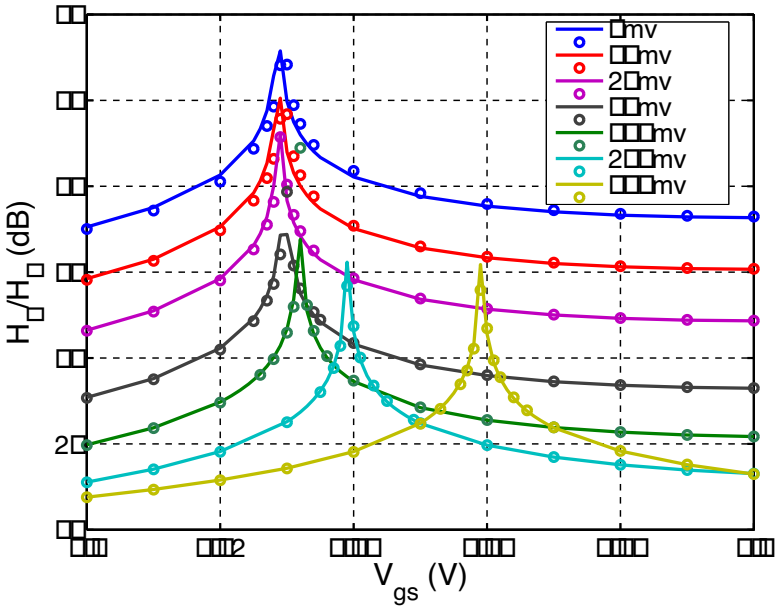


Figure 6.5: HD_3 by Eq. 6.8 and FFT of $\tilde{S}_{out}(n)$

6.2.4 H_3 Decomposition

To further appreciate the implications of Eq. 6.7, assume a nonlinear system with a fixed third-order nonlinearity, i.e. $a_3(n)$ = a nonzero constant and $a_1(n)$ and $a_2(n)$ as specified in Eq. 6.9. Separating Eq. 6.7 in real and imaginary parts, Fig. 6.6 shows the individual terms making up the summation. H_3 is the summation result of the figure in the third row over the full cycle. Examining the odd and even symmetry of these curves with respect to 0.5 T, H_3 is contributed solely by the imaginary part of Eq. 6.7 while the real part cancels out after integrating over one cycle. With the same setup but this time assumes a

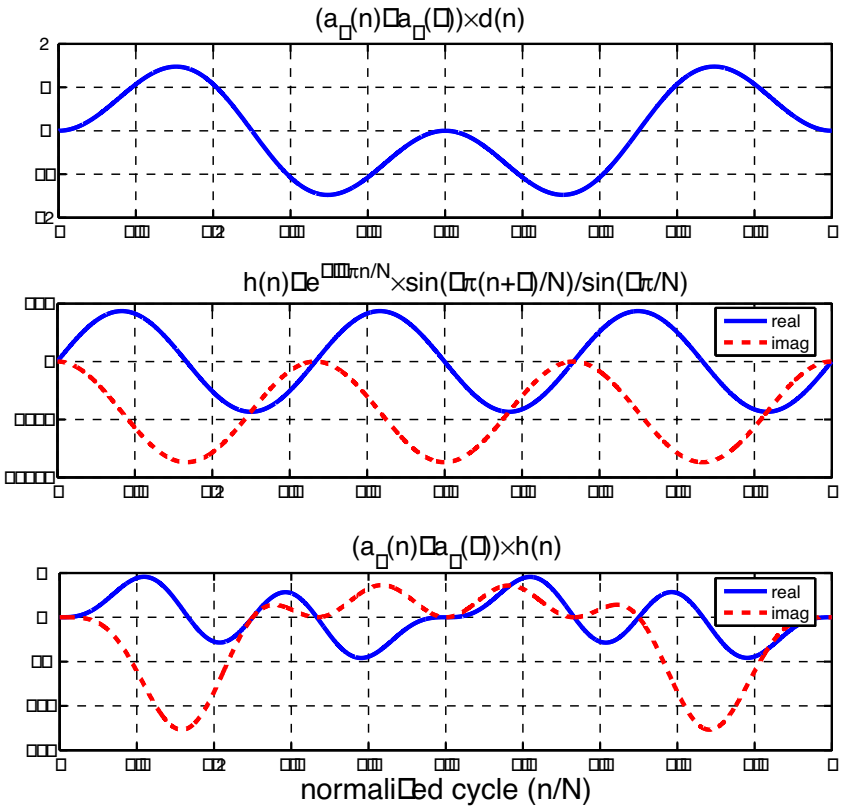


Figure 6.6: H_3 decomposition with $a_3(n)$=constant

fixed fourth-order nonlinearity, $a_4(n) = constant$. No third-order distortion is expected, as proven by Fig. 6.7 by the odd symmetry w.r.t. 0.5 T for the imaginary part, and separate odd symmetry w.r.t. 0.25 T and 0.75 T for the real part. Fig. 6.6 and Fig. 6.7 reveal that the third-order harmonic is determined by the imaginary value of the H_3 summation. Furthermore, the symmetry property of these curves is related only to that in the first row, once value of K, the order of nonlinearity is chosen. For instance, $a_3(n) = constant$ results in the even symmetry w.r.t. 0.5 T for $(a_1(n) - a_1(0)) \times d(n)$ in Fig. 6.6, which later translates into nonzero H_3. These graphic results can be generalized to conclude that

third-order harmonic exists if and only if there is even symmetry of $a_3(n)$ w.r.t. the center point, i.e any *HOTs* with even-order higher than a_3.

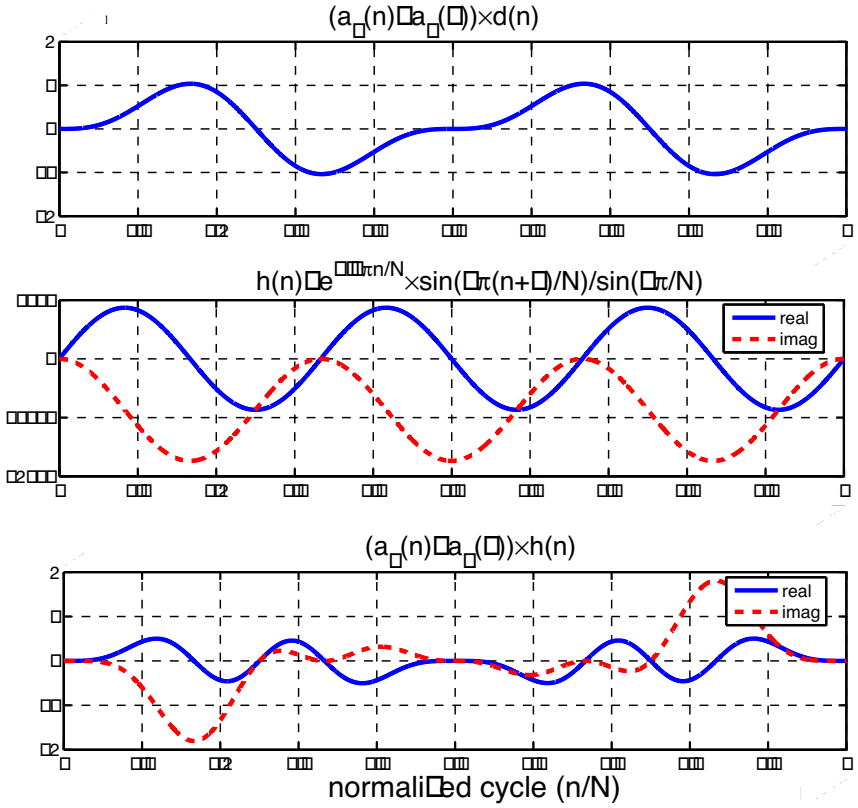


Figure 6.7: H_3 decomposition with $a_4(n)$=constant

6.3 Implication to amplifier linearization

In this section, we will apply the new approach to examine existing amplifier linearization schemes, and use it to explain the discrepancies in Fig. 6.1.

6.3.1 Sweet-spot v.s. $MGTR$

Ignore $a_2(0)$ and rewrite Eq. 6.10 with

$$dist(n) = \sum_{i=1}^{n-2} \left(\Big((a_3(i-1) - a_3(0)) + a_3(0) \Big) \cdot d(i-1) \cdot \sum_{j=i+1}^{n-1} \big(d(j) \cdot \sum_{k=j+1}^{n} d(k) \big) \right) \quad (6.13)$$

where $a_3(0)$ is the initial a_3, and $a_3(i-1) - a_3(0)$ accounts for the dynamic change of $a_3(n)$. Although $a_3(0)$ for sweet-spot and $MGTR$ is chosen the same, the shape of the curve, i.e. $a_3(i-1) - a_3(0)$, differs. The advantages of $MGTR$ are two fold. The flatter a_3 zero crossing reduces the bias sensitivity of $MGTR$'s $a_3(0)$. More importantly, the distortion caused by $a_3(n) - a_3(0)$ dynamics is greatly suppressed due to its smaller value. The smaller $a_3(n) - a(0)$ of $MGTR$ is attribute to $a_3(n)$ of transistors in weak inversion (WI) and strong inversion (SI) moving in opposite direction with the same locus $d(n)$, shown by expressing Eq. 6.13 with individual $a_3(n)$ sharing the same $d(n)$

$$\begin{aligned} dist(n) = \sum_{i=1}^{n-2} \Bigg(\Big(& a_{3,WI}(0) + a_{3,SI}(0) + \\ & a_{3,WI}(i-1) - a_{3,WI}(0) + a_{3,SI}(i-1) - a_{3,SI}(0) \Big) \cdot \\ & d(i-1) \cdot \sum_{j=i+1}^{n-1} \big(d(j) \cdot \sum_{k=j+1}^{n} d(k) \big) \Bigg) \end{aligned} \quad (6.14)$$

6.3.2 Split-$MGTR$

Size matching plays a critical role for a proper $MGTR$ linearization. It allows both $a_3(0)$ and the dynamic $a_3(n) - a_3(0)$ be minimized. In practice, size mismatch is inevitable due to process variations. Moreover the size ratio of two transistors can not be made arbitrarily fine because of limited lithography accuracy. Like sweet-spot, $MGTR$ is also susceptible to the inferior modeling accuracy of g_m derivatives. These practical factors

create more room of discrepancy in the fabricated circuits. Once the silicon is fabricated, the suboptimal size ratio can not be adjusted. Transistor bias being the only tuning knob, it is difficult to achieve the same level of linearization as expected by simulation.

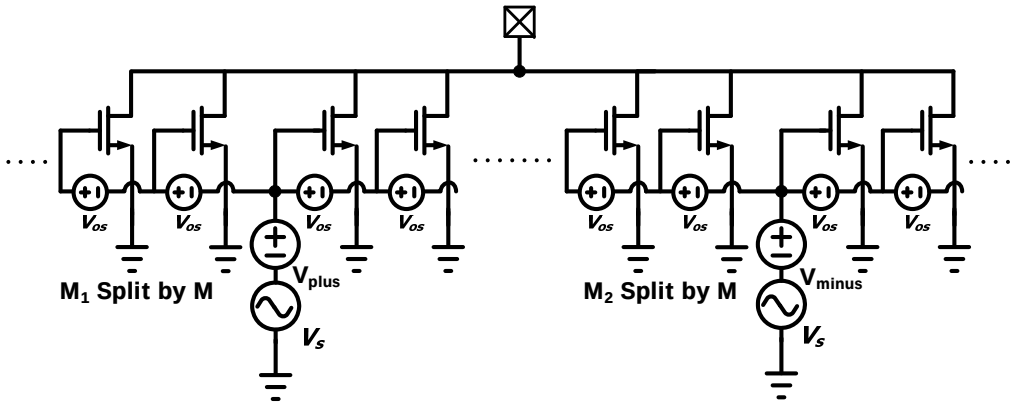


Figure 6.8: Split-Multi Gated Transistor (split-$MGTR$)

A modified scheme called "split-$MGTR$" overcomes the practical limitation by splitting the two $MGTR$ transistors into several, say, N, unit transistors. The bias of these unit transistors is equally separated by a small offset from the main $MGTR$ biases. Fig. 6.8 shows the split-$MGTR$ circuit configuration. Referring to g_m'' plane in Fig. 6.9, the large dot in red and medium dot in green represents the $g_m''(0)$ for sweet-spot and $MGTR$, respectively, while a handful of smaller dots refers to those of split-$MGTR$. Unlike conventional $MGTR$, cancelation of $g_m''(n) - g_m''(0)$ in split-$MGTR$ applies to transistors in the same region of g_m'' polarity. Since the unit transistors in the same region have identical size and layout, the likelihood of size mismatch is minimized. Furthermore, by manipulating the offset voltage as a post-fabrication knob, the same level of linearization can be achieved even with modeling error and after the transistor is fabricated.

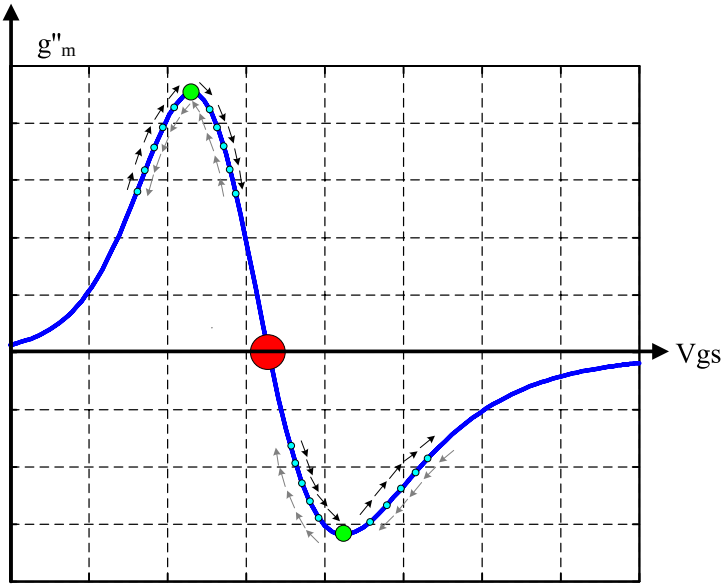
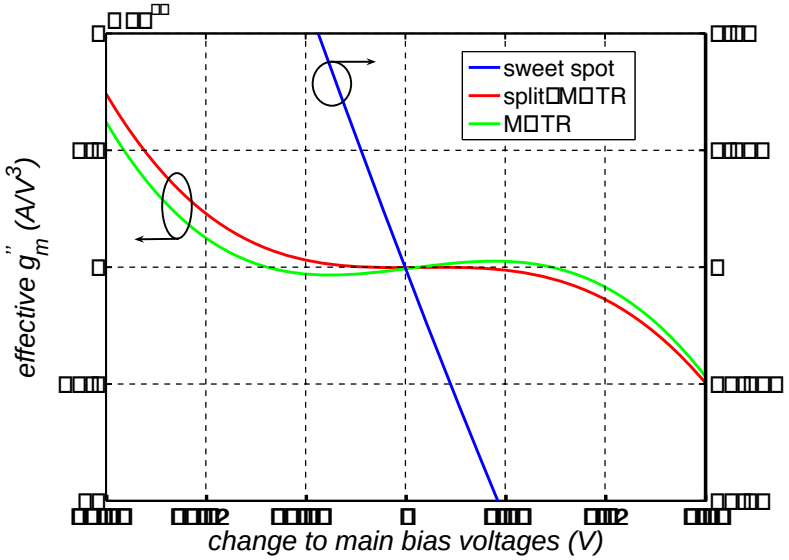
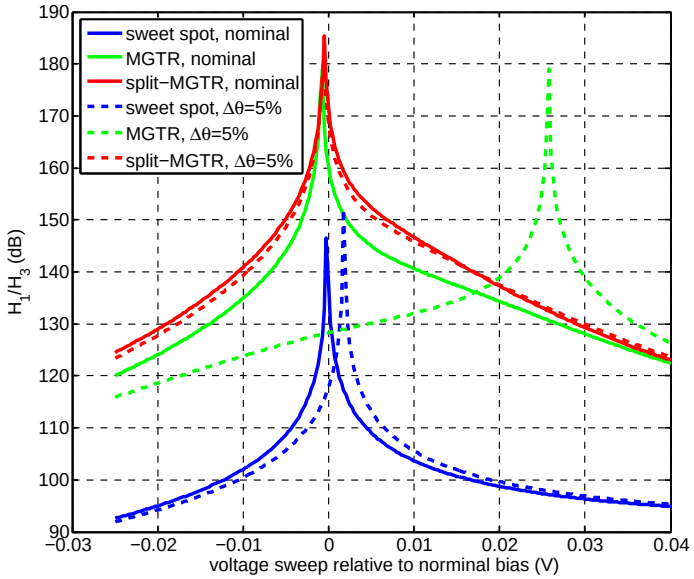


Figure 6.9: Illustration of split-$MGTR$ on g_m'' plane

Fig. 6.10(a) compares the three linearization schemes using the same behavioral transistor model as before. By limiting the accuracy of size ratio to 0.1 and bias resolution of 1 mV, split-$MGTR$ achieves even flatter g_m'' than conventional $MGTR$. Accounting for 5 % variation in θ, split-$MGTR$ in Fig. 6.10(b) maintains almost the same performance with the same master bias and by adjusting the offset voltages. On the other hand, conventional $MGTR$ has to adjust its master bias by as large as 25 mV, which may cause adverse effects due to significant bias shift. In this particular example, sweet-spot shows less variation than $MGTR$ because θ in Eq. 6.13 represent the mobility degradation and manifests only at substantial gate bias beyond the point of sweet spot.



(a) g_m'' with nominal design



(b) HD_3 with $\Delta\theta$=5%

Figure 6.10: Comparison of three linearization schemes

6.4 Summary

Conventional distortion characterization using truncated Taylor series is simple for hand calculation but loses accuracy when signal magnitude increases. We therefore present a dynamic Taylor series approach by keeping the coefficient values most up-to-date when signal moves on the transfer curve. The new method provides a different way of understanding circuit nonlinear response by focusing on the shape of its coefficient characteristic curve. A modified linearization scheme is verified with the behavioral model and shows the potential of better linearity and robustness.

Chapter 7

Prototype II: 2-Tone-Indepndent Broadband Linearization

The IIP_3 2-tone frequency dependence observed in the first prototype LNA could pose a concern for applications where interference signals are close to each other in the spectrum. In this chapter, a revised solution will be investigated to completely eliminate this issue.

7.1 Common-gate stage revisit

First, remove capacitor C_{12} in the schematic of Fig. 7.1, and re-examine the flow of harmonic currents in the circuit. For simplicity, assume the second stage performs an ideal current summation by the common-source transistors, M_3, M_4 and M_5 connecting to V_1, V_x and V_2 separately. Starting with the first order, i.e. linear current by g_m, and

referring to the voltage at V_x as 1, the relative first-order voltages at nodes V_1 and V_2 are

$$\begin{aligned} V_1^{(1)} &= \frac{R_1}{R_a} \\ V_2^{(1)} &= \frac{R_2}{R_b} \end{aligned} \tag{7.1}$$

, where R_a and R_b are the resistance looking into the source of M_1 and M_2, respectively.

$$\begin{aligned} R_a &= \frac{1}{g_{m1}} + \frac{R_1}{g_{m1} r_{o1}} \\ R_b &= \frac{1}{g_{m2}} + \frac{R_2}{g_{m2} r_{o2}} \end{aligned} \tag{7.2}$$

As for the second-order harmonic voltages, since g'_m currents of M_1 and M_2 flow in the same direction, there is no current through R_s if magnitude of both g'_m is matched. As such, the second-order harmonic voltage at V_x is zero and those at V_1 and V_2 have a ratio of $\frac{-R_1}{R_2}$[1]. At the third order, M_1's g''_m current causes the relative magnitude of the third-order harmonic voltages among V_1, V_x and V_2 as

$$\begin{aligned} V_{1,M_1}^{(3)} &= R_1 \\ V_{x,M_1}^{(3)} &= -(R_s \| R_b) \\ V_{2,M_1}^{(3)} &= -(R_s \| R_b) \times \frac{R_2}{R_b} \end{aligned} \tag{7.3}$$

Likewise, consider M_2's g''_m and the current flow from bottom up, the relative voltage values become

$$\begin{aligned} V_{2,M_2}^{(3)} &= R_2 \\ V_{x,M_2}^{(3)} &= -(R_s \| R_a) \\ V_{1,M_2}^{(3)} &= -(R_s \| R_a) \times \frac{R_1}{R_a} \end{aligned} \tag{7.4}$$

Harmonic voltages at V_1, V_2 and V_x are then amplified by g_m and its derivatives of common-

[1] Also true if r_o nonlinearities are considered

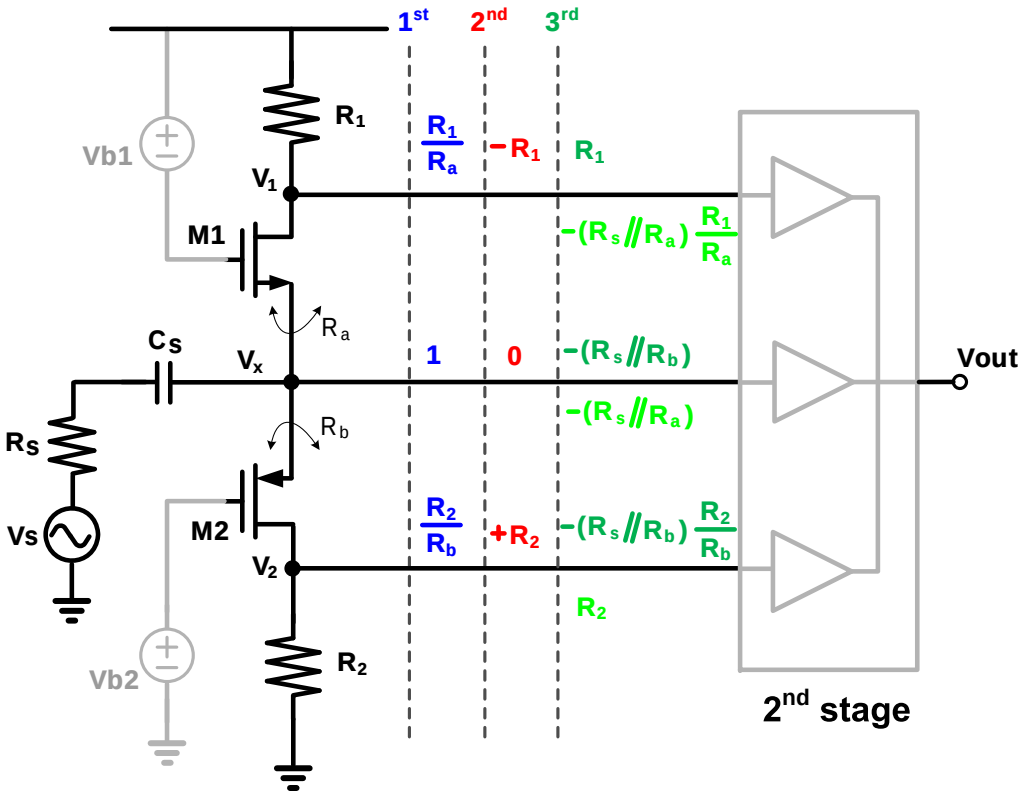


Figure 7.1: Illustration of relative harmonic voltages at the common-gate stage

source transistors. Results are summed at V_{out}. It is already known from Eq. 4.10 that co-existence of first- and second-order harmonic voltages at V_1, V_2 and V_x leads to third-order nonlinear output through second-order interaction with common-source transistors' g'_m. The previous LNA implementation makes the second harmonic voltages at V_1, V_2 and V_x all become zero by confining common-gate stage's g'_m current circulating within M_1, M_2 and C_{12}. In Fig. 7.1, only V_x is zero under matched condition. However, the opposite polarity between the second harmonic voltages at V_1 and V_2 in Fig. 7.1 and the same polarity of g'_m in the common-source transistors make the polarity of V_{out} distortion by second-order interaction remain opposite. The polarity difference can be exploited to

realize full cancelation of the third harmonic distortion, as will be explained next.

7.2 Noise Analysis

Consider V_{out} noise due to various noise sources,

$$\begin{aligned}
\bar{V}^2_{out,R_1} &= 4KTR_1 \times g^2_{m3} \times R^2_L \times \Delta f \\
\bar{V}^2_{out,R_2} &= 4KTR_2 \times g^2_{m5} \times R^2_L \times \Delta f \\
\bar{v}^2_{out,M_3} &= 4KT\frac{\gamma}{\alpha} \times g_{m3} \times R^2_L \times \Delta f \\
\bar{v}^2_{out,M_4} &= 4KT\frac{\gamma}{\alpha} \times g_{m4} \times R^2_L \times \Delta f \\
\bar{v}^2_{out,M_5} &= 4KT\frac{\gamma}{\alpha} \times g_{m5} \times R^2_L \times \Delta f \\
\bar{v}^2_{out,R_s} &= \frac{4KT}{R_s} \times (R_s \parallel R_{in})^2 \times (g_{m4} + g_{m3} \times \frac{R_1}{R_a} + g_{m5} \times \frac{R_2}{R_b})^2 \times R^2_L \times \Delta f
\end{aligned} \tag{7.5}$$

, where R_{in} is defined as

$$R_{in} = R_a \parallel R_b \tag{7.6}$$

It can also be shown that V_{out} noise due to common-gate transistors are

$$\begin{aligned}
\bar{v}^2_{out,M_1} &= 4KT\frac{\gamma}{\alpha} \times \frac{g_{m1}}{(1 + g_{m1}(R_s \parallel R_b))^2} \times \left(g_{m3}R_1 - (R_s \parallel R_b)(g_{m4} + g_{m5} \times \frac{R_2}{R_b})\right)^2 \times R^2_L \times \Delta f \\
\bar{v}^2_{out,M_2} &= 4KT\frac{\gamma}{\alpha} \times \frac{g_{m2}}{(1 + g_{m2}(R_s \parallel R_a))^2} \times \left(g_{m5}R_2 - (R_s \parallel R_a)(g_{m4} + g_{m3} \times \frac{R_1}{R_a})\right)^2 \times R^2_L \times \Delta f
\end{aligned} \tag{7.7}$$

The subtraction in the main bracket in Eq. 7.7 indicates M_1 and M_2 noise cancelation. A slightly different cancelation mechanism exists between this circuit and that of Fig. 3.2. Taking M_1 noise for example, noise voltages at V_x and V_2 have the same polarity but opposite to that at V_1. As such, the term, $g_{m5} \times \frac{R_2}{R_b}$, appears in addition to g_{m4} to assist

the cancelation. By making zero value of the capital bracket in Eq. 7.7,

$$g_{m4} = (g_{m5}R_2 - \frac{R_sR_1}{R_s+R_a}g_{m3})\frac{R_s+R_a}{R_sR_a} \tag{7.8}$$

Note that g_{m3} and g_{m5}, R_1 and R_2, and R_a and R_b are interchangeable due to circuit duality. Both expressions result in the same g_{m4}. Therefore the following ratio has to be satisfied and constitutes the first criteria for full cancelation of M_1 and M_2 noise

$$\frac{g_{m3}}{g_{m5}} = \frac{R_2}{R_1} \tag{7.9}$$

7.3 Distortion Analysis

Consider the distortion caused by the intrinsic third-order nonlinearity of the common-gate transistors. As previously pointed out, their g''_m current is modeled in the same way their noise current is modeled, therefore, Eq. 7.9 also serves as the criteria for canceling M_1 and M_2's intrinsic g''_m distortion.

7.3.1 Second-order interaction

Referring to Fig. 7.1, the complete removal of third harmonic output due to interaction between the common-source and common-gate stages occurs if

$$V_1^{(1)} \times V_1^{(2)} \times g'_{m3} + V_2^{(1)} \times V_2^{(2)} \times g'_{m5} = 0 \tag{7.10}$$

is satisfied. This calls for

$$\begin{aligned} \frac{R_1}{R_a}(-R_1) \times g'_{m3} + \frac{R_2}{R_b}(R_2) \times g'_{m5} = 0 \\ \Rightarrow \frac{g'_{m3}}{g'_{m5}} = \frac{R_a}{R_b}\left(\frac{R_2}{R_1}\right)^2 \end{aligned} \tag{7.11}$$

as the second cancelation criteria. To simplify the design, assume M_3 and M_5 are implemented by unit transistors with the same bias voltage. This transforms the required g_m and g'_m ratio into size ratio, or the number of unit devices, between M_3 and M_5. By equating Eq. 7.9 and Eq. 7.11,

$$\begin{aligned} \frac{g_{m3}}{g_{m5}} = \frac{\text{size of } M_3}{\text{size of } M_5} = \frac{R_2}{R_1} \quad \text{and} \\ \frac{R_1}{R_2} = \frac{R_a}{R_b} \end{aligned} \tag{7.12}$$

Substitute Eq. 7.12 into Eq. 7.7, we get

$$\begin{aligned} \bar{v}^2_{out,M_1} &= 4KT\frac{\gamma}{\alpha} \times \frac{g_{m1}R_b^2}{(R_s + R_b + g_{m1}R_sR_b)^2} \times (g_{m3}R_1 - g_{m4}R_s)^2 \times R_L^2 \times \Delta f \\ \bar{v}^2_{out,M_2} &= 4KT\frac{\gamma}{\alpha} \times \frac{g_{m2}R_a^2}{(R_s + R_a + g_{m2}R_sR_a)^2} \times (g_{m3}R_1 - g_{m4}R_s)^2 \times R_L^2 \times \Delta f \end{aligned} \tag{7.13}$$

and the third cancelation criteria for M_4,

$$\frac{g_{m3}}{g_{m4}} = \frac{R_s}{R_1} \tag{7.14}$$

7.3.2 MGTR

Up to this point, only the size ratio between M_3 and M_5, and R_1 and R_2 are decided. The remaining third harmonic output yet to be canceled is the intrinsic distortion caused by common-source transistors' g''_m. Not too surprisingly, this distortion is canceled by applying $MGTR$ to M_3, M_4 and M_5 with the last cancelation criteria

$$\begin{aligned} (\frac{R_1}{R_a})^3 \times g''_{m3} + (\frac{R_2}{R_b})^3 \times g''_{m5} + g''_{m4} = 0 \\ \Rightarrow \frac{g''_{m3}}{g''_{m4}} = -(\frac{R_a}{R_1})^3 \times (\frac{R_2}{R_1}) \end{aligned} \tag{7.15}$$

which completes device bias and size values for M_3, M_5 and M_4.

7.4 Noise/Distortion cancelation co-design

From Eq. 7.13, the same residual cancelation factor is defined.

$$\delta = \frac{g_{m4} \times R_s}{g_{m3} \times R_1} - 1 \tag{7.16}$$

Also, use ε_{rr} in Eq. 3.12 to account for input impedance mismatch, V_{out} noise is rewritten as

$$\begin{aligned}
\bar{v}^2_{out,R_s} &= \frac{4KT}{R_s} \times g^2_{m3} \times R^2_1 \times (\frac{1+\varepsilon_{rr}}{2+\varepsilon_{rr}}) \times (1+\delta+\frac{1}{1+\varepsilon_{rr}}) \times R^2_L \times \Delta f \\
\bar{v}^2_{out,M_1} &= 4KT\frac{\gamma}{\alpha} \times g_{m1} \times g^2_{m3} \times (\frac{R_1}{R_s})^2 \times (\frac{1}{g_{m1}} \parallel R_s \parallel R_b)^2 \times \delta^2 \times R^2_L \times \Delta f \\
\bar{v}^2_{out,M_2} &= 4KT\frac{\gamma}{\alpha} \times g_{m2} \times g^2_{m3} \times (\frac{R_1}{R_s})^2 \times (\frac{1}{g_{m2}} \parallel R_s \parallel R_a)^2 \times \delta^2 \times R^2_L \times \Delta f
\end{aligned} \tag{7.17}$$

The noise factors of individual devices are

$$F_{R_1} = \frac{R_s}{R_1} \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \tag{7.18}$$

$$F_{R_2} = \frac{R_s}{R_2} \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \tag{7.19}$$

$$F_{M_3} = \frac{\frac{\gamma}{\alpha}}{g_{m3}R_1} \times \frac{R_s}{R_1} \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \tag{7.20}$$

$$F_{M_5} = \frac{\frac{\gamma}{\alpha}}{g_{m3}R_1} \times \frac{R_s}{R_2} \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \tag{7.21}$$

$$F_{M_4} = \frac{\frac{\gamma}{\alpha}}{g_{m3}R_1} \times (1+\delta) \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \tag{7.22}$$

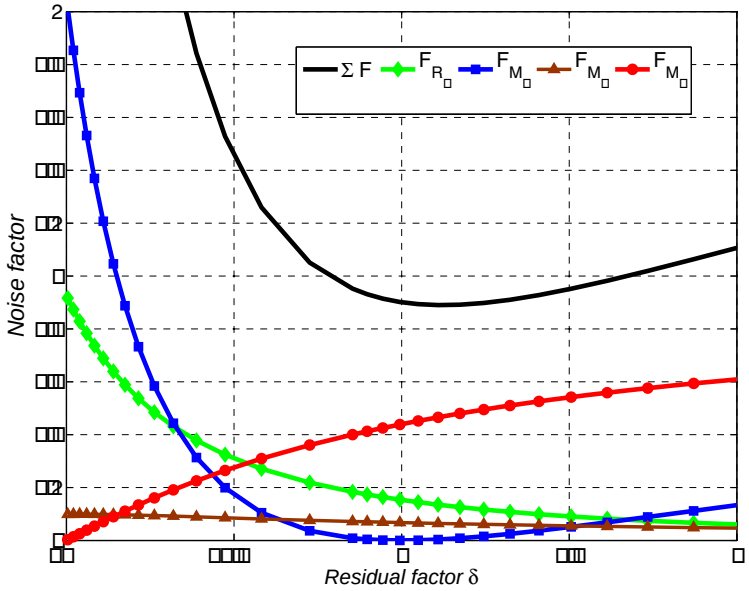


Figure 7.2: Noise factor of individual devices

$$F_{M_1} = \frac{\frac{\gamma}{\alpha}}{R_s} \times \frac{g_{m1}}{(\frac{1}{g_{m1}} \parallel R_s \parallel R_b)^2} \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \times \delta^2 \times (1+\delta) \tag{7.23}$$

$$F_{M_2} = \frac{\frac{\gamma}{\alpha}}{R_s} \times \frac{g_{m2}}{(\frac{1}{g_{m2}} \parallel R_s \parallel R_a)^2} \times \left(\frac{2+\varepsilon_{rr}}{1+(1+\varepsilon_{rr})(1+\delta)}\right)^2 \times \delta^2 \times (1+\delta) \tag{7.24}$$

The above expressions have a striking resemblance with those in chapter 3 due to circuitry similarities. The only exception is the additional term of $1+\delta$ associated with F_{M_1} and F_{M_2}. This additional term accounts for the aided cancelation by the noise current flowing through the counterpart common-gate transistor. From noise optimization in chapter 3, this aided M_1 and M_2 cancelation allows us to put more emphasis on minimizing M_4 noise, instead of M_1 and M_2, by allocating larger g_{m4}. As a result, δ_{opt} grows and shifts to the

right of the value predicted by Eq. 3.28 [2]. This is seen in Fig. 7.2 where, the valley of F aligns very close to $\delta = 0$. This is a favorable property since $\delta = 0$ is the underlying assumption behind complete distortion cancelation described in section 7.3.

7.5 Implementation

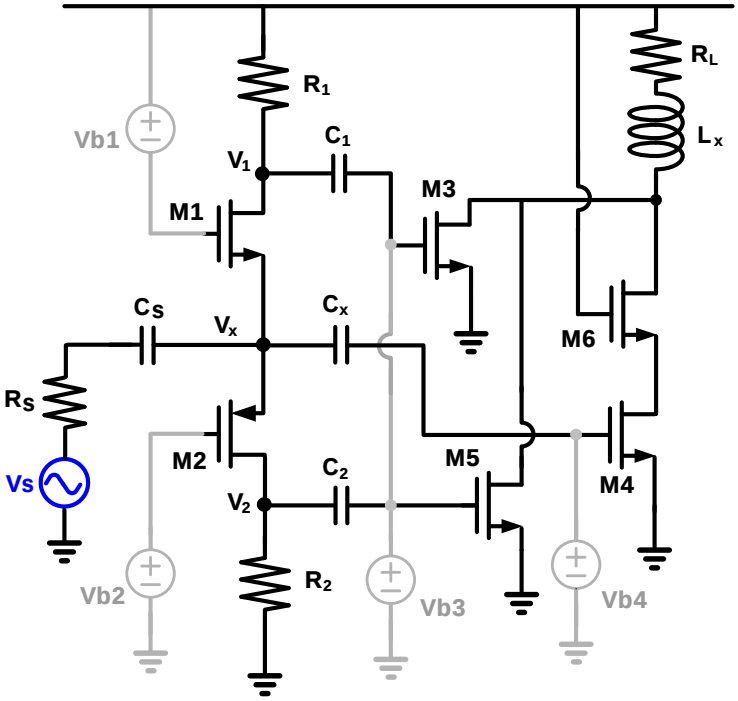


Figure 7.3: Measured IIP_3

Fig. 7.3 shows the full schematic of the modified LNA. AC coupling capacitors C_1, C_2 and C_x are used to isolate the DC bias and do not contribute to IIP_3 frequency dependence. For compact implementation, these capacitors can be removed using techniques

[2] Another way of looking at this δ_{opt} shift is by applying δ_{opt} of Eq. 3.28 to the above expressions, most noise factors remain unchanged except F_{M_1} and F_{M_2} become smaller for a negative δ. δ_{opt} moves to the right to account for the lower F_{opt}.

mentioned in section 5.4. The parasitic capacitance introduced by these capacitors limits the LNA bandwidth. A shunt-peaking inductor is used to compensate the bandwidth shrinkage.

Special attention is paid to the way the harmonic currents are summed in the common-source stage. Referring to Fig. 7.1, the second-order harmonic voltage at V_x is designed to be perfectly null so that the third harmonic distortion by M_4's g'_m is zero. On the other hand, the second-order harmonic voltages at V_1 and V_2 are measurable. They could leak to V_x through M_4's C_{gd} capacitor and compromise the design of zero second-order harmonic voltage. M_6 is therefore placed between M_4 and M_3 and M_5 to provide sufficient isolation. Since M_3 and M_5 are the transistors in weak inversion and run at low current level, their output conductance is still high compared to the load resistor R_L, and does not affect the current summation.

The circuit is designed in 65 nm CMOS process with the device values summarized in Table 7.5. It utilizes low-V_t transistors in a low-power 65 nm technology. Fig. 7.4 compares the simulated noise figure with and without the cancelation by M_4. With supply voltage scaling from 1.5 V in the previous implementation to 1.2 V, resistor value, R_1 and R_2, decreases and increases the lowest achievable noise figure.

M_1	$(0.5\mu m/0.06\mu m)\times 40$	R_1	400Ω
M_2	$(0.5\mu m/0.06\mu m)\times 92$	R_2	310Ω
$M_{3,5}$	$(0.5\mu m/0.06\mu m)\times 10$	R_L	60Ω
M_4	$(6\mu m/0.06\mu m)\times 20$	$C_{x,1,2}$	0.5 pF
M_6	$(6\mu m/0.06\mu m)\times 48$	L_x	1.8 nH

Table 7.1: Device size

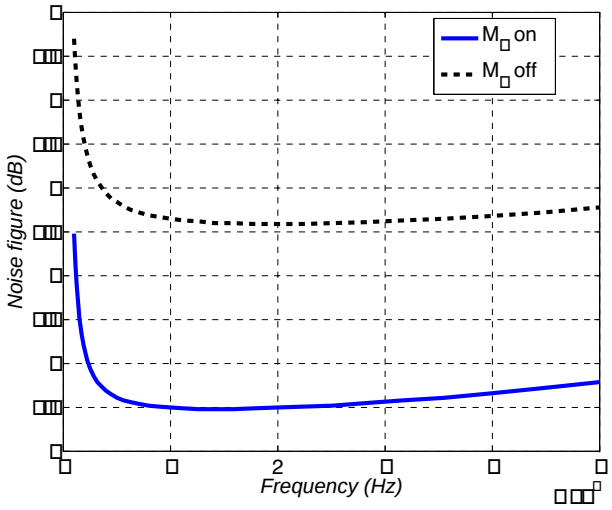


Figure 7.4: LNA noise figure with and without M_4

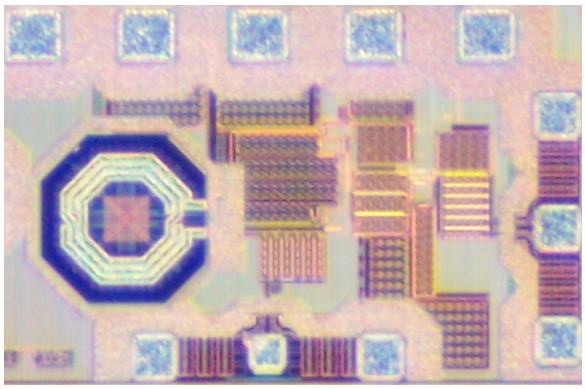


Figure 7.5: Chip microphotograph

7.6 Measurement

The chip microphotograph is shown in Fig. 7.5. The die area including the pads is $750\mu m \times 300\mu m$. Excluding the pads, most of the die area is taken up by the passive

devices, e.g. MIM capacitors and a two-turn symmetric inductor from foundry's inductor library. The same measurement setup in chapter 5 is repeated to conduct the following measurements.

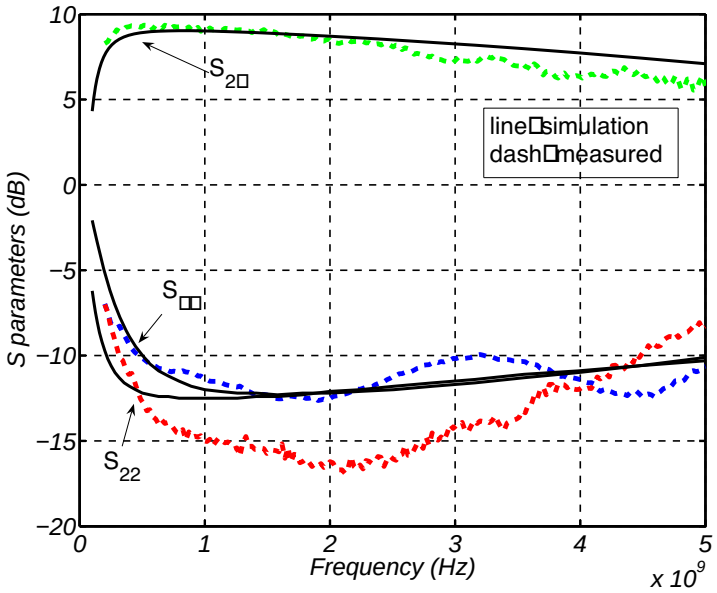


Figure 7.6: Measured S parameters

Fig. 7.6 shows the measured S parameters. Both input and output impedance matching are maintained up to 5 GHz., with a peak S_{21} of 9 dB and 3 dB bandwidth about 5 GHz as well. Fig. 7.7 shows the measured noise figure. It matches with the post-layout simulation at low frequencies but the discrepancy rises up at higher frequencies. This is due to the extra capacitance associated with the intermediate nodes resulting from process variation of the MIM capacitors.

LNA linearity is tested by applying 2 sinusoidal tones of 1 GHz and 1.1 GHz, and measuring the output at 1 GHz and 1.2 GHz respectively. Fig. 7.8 shows the results.

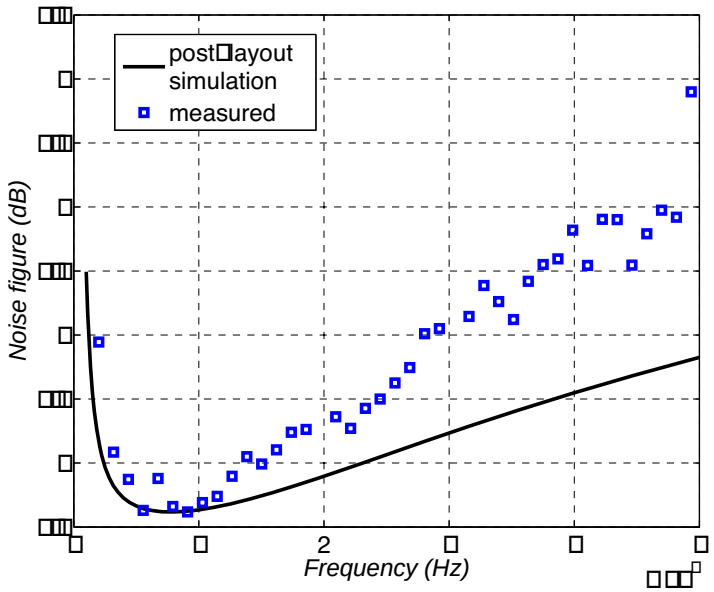


Figure 7.7: Measured noise figure

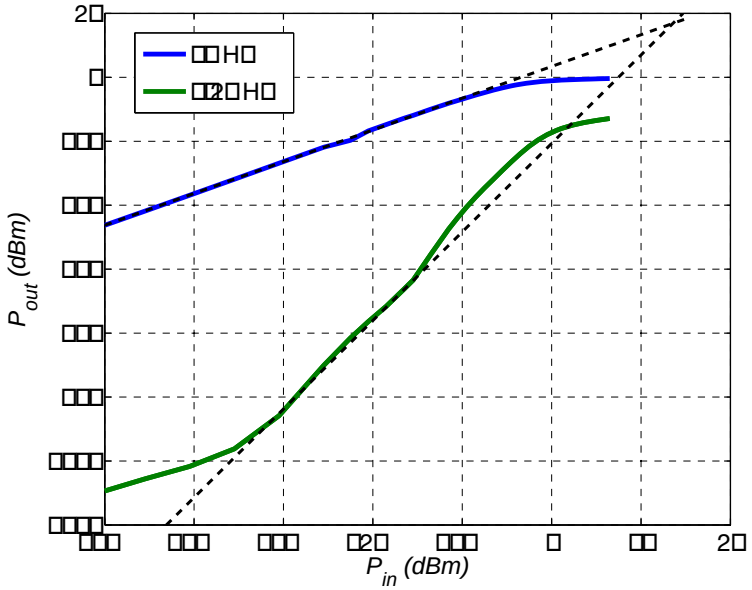


Figure 7.8: Measured IIP_3 with input two tones at 1 GHz and 1.1 GHz

To investigate the linearity with different frequency setting, a wide range of two-tone frequencies are applied. Fig. 7.9 shows the results with center frequency from 1 GHz to 4 GHz and tone spacing ranging from 100 KHz to 500 MHz. The bias voltages are

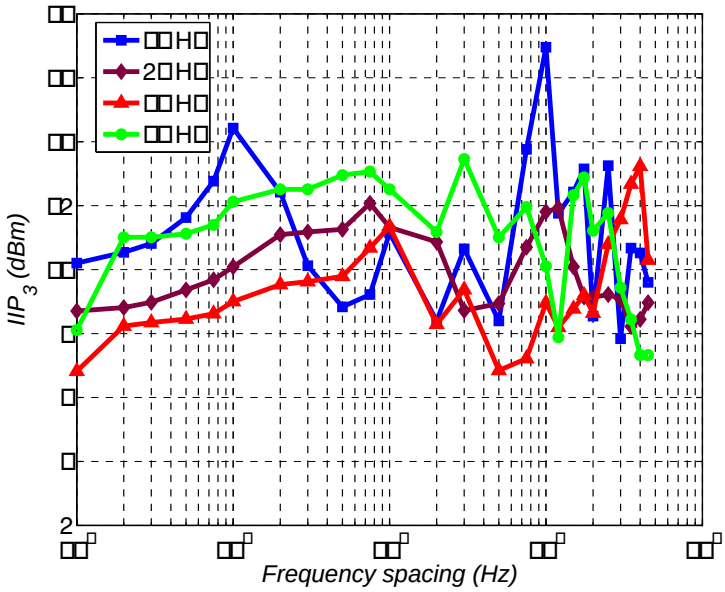


Figure 7.9: Measured IIP_3 with different frequency setting

kept the same while these frequencies are varied. The IIP_3 maintains a constantly high value between +8 and +12 dBm[3]. The residual variations are attribute to the remaining frequency dependency such as feedback through C_{gd}. The measured 1 dB compression point, P_{1dB}, is -9 dBm. Fig. 7.10 compares the bias sensitivity of both cancelation schemes. The center frequency of the measurements is at 1.5 GHz with tone spacing of 100 MHz. The measured noise figure stays insensitive to bias change, and there is a bias window of 20 mV for IIP_3 of +10 dBm. Finally Fig. 7.10 shows the IIP_3 sensitivity of V_{b1} and V_{b2}.

[3]calculated from $IIP_3 = P_{in} + 0.5 \times IM_3$ (all in dB) instead of graphic extraction

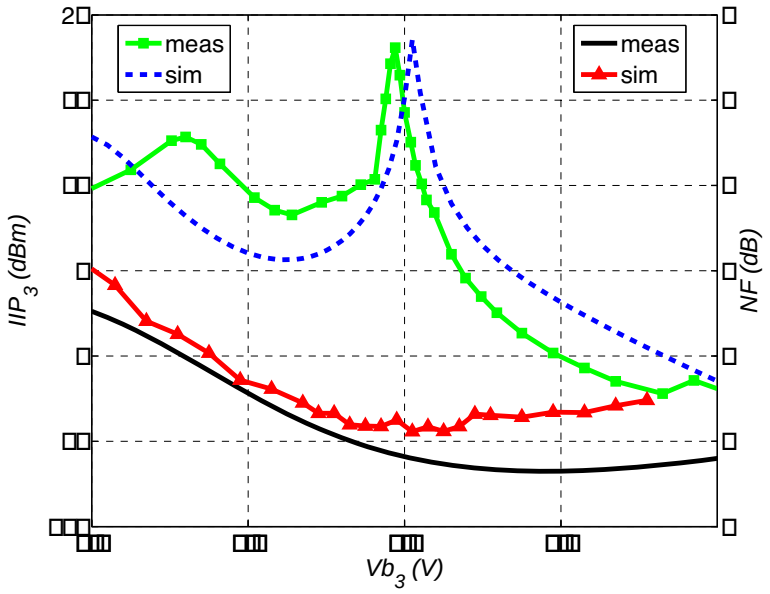


Figure 7.10: Bias sensitivity of the measured IIP_3 and noise figure

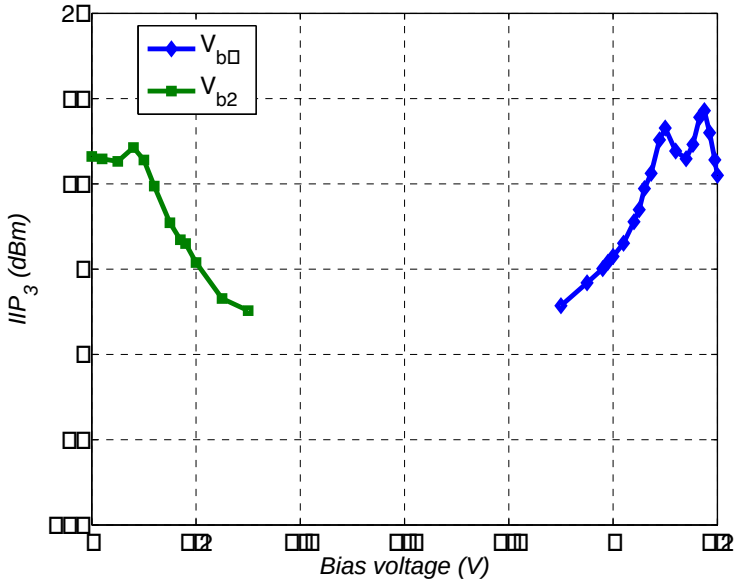


Figure 7.11: IIP_3

7.7 Summary

The performances of two LNA prototypes are summarized in Table 7.7 and compared with other state-of-the-art works. Both circuits achieve similar voltage gain of 16 dB while the noise figure of the 65 nm implementation degrades by roughly 1 dB. This is the direct impact of supply voltage scaling which reduces the resistor values. On the other hand, distortion cancelation relies mainly on the matching between different transistor pairs rather than the linearity of a single transistor. Therefore IIP_3 of the LNA remains the same.

	Technology (μm)	BW	S_{21}	NF	IIP_3	power (mW)
[5]	0.25	$0.002-1.6$	13.7[4]	$2-2.4$	0	35
[78]	0.18	$1.2-11.9$	9.7	$4.5-5$	-6.2	20
[79][1]	0.13	$0.4-5$	19	$3-4.2$	$+1$	11.7
[32]	0.13	$3.1-10.6$	15.1	2.5	-5.1	9
[80]	0.13	$2-5.2$	16	$4.7-5.7$	-6	38
[81]	0.13	$1-7$	17	2.4	-4.1	25
[82] [1]	0.065	$0.2-5.2$	15.6	<3.5	>0	21
[83]	0.09	$.5-8.2$	25	$1.9-2.6$	-4	41.8[3]
[53]	0.09	$0-6.5$	16.5	2.7	-4.3	9.7
LNA1	0.13	$0.8-2.1$	>8.5	2.6	$+16$	17.4
LNA2	0.065	$0.8-5$	16	$3.5-5.5$	$8-12$	17.4

[1] single-ended to differential

[2] differential LNA

[3] V_{dd}=1.8 V

[4] voltage gain

Table 7.2: Performance comparison of broadband LNA.

Chapter 8

Conclusions

8.1 Thesis Summary

In summary of this research, we proposed, analyzed and implemented the noise and distortion canceling schemes in broadband LNA designs which aim to tackle the classical design tradeoff among low noise, high linearity and broadband impedance matching.

In chapter 1, we surveyed the current trend towards highly integrated mobile terminals. We identified that in both conventional homodyne and recent discrete-time receiver architectures, a single broadband low noise amplifier will remain as an indispensable part. The unprecedented high linearity requirement, however, imposes a great challenge to the broadband LNA design, but little emphasis on this issue has been found in the prior research.

In chapter 2 we briefly went over a concise but in-depth review on major LNA performance metrics. The definition of each metric was reviewed, and its implications to

receiver link design were demonstrated with typical circuits and system examples.

In chapter 3, we presented detailed analysis on the design of noise cancelation in a broadband common-gate common-source casecade LNA. For the first time in the literature, a formulation for the noise-power optimization of the noise canceling scheme is derived. A sub-2 dB noise figure with practical circuit parameters is predicted achievable with noise cancelation scheme.

In chapter 4 we focused on the design of third-order output distortion cancelation in the proposed LNA. Different output IM_3 mechanisms were identified with the aid of a Volterra series analysis. Complete removal of output IM_3 was achieved by combining noise cancelation, $MGTR$, and a P/NMOS pair to eliminate the remaining second-order interaction at the common-gate stage.

Measurement results of a 0.8-2.1 GHz prototype in 0.13 μm CMOS were presented in chapter 5. The measured IIP_3 was 2-tone frequency dependent, due to the large capacitive impedance presented to the low-frequency IM_2 current. This issue was solved in chapter 7 in a second implementation in 65 nm CMOS by removing the capacitor and by modifying the design of harmonic current summation. It achieved an equally high IIP_3 and was free of 2-tone frequency dependence. Due to reduced supply voltage, the new noise figure increases by 1 dB to 3.5 dB.

Aside from circuits design at transistor level, chapter 5 looked into accurate distortion characterization when signal magnitude becomes large. Conventional Taylor series is known to be valid for small signals only and does not account for large-signal distortion accurately. A dynamic Taylor series approach was proposed instead. The new approach

explains the discrepancy between sweet-spot and $MGTR$ linearization, and suggests a modified linearization with superior performance and robustness.

8.2 Suggestions for Future Research

Based on the results of the two noise and distortion canceling LNA prototypes, the following directions is worth pursuing for future research. First, noise cancelation has been widely, however exclusively, applied to broadband common-gate or shunt-shunt feedback amplifiers. We already analyzed that a very low noise figure is achievable with cancelation, but at cost of substantial power consumption. In the common-gate common-source cascade example, the increased bias current is mainly dissipated to reduce the noise figure of the main common-source transistor. If the noise of the common-source transistor is also canceled, its bias current can be much reduced for the same noise figure performance. Therefore the circuit of dual noise cancelation loops in Fig. 3.7 becomes very attractive in terms of its noise-power tradeoff. A detailed analysis on the performance of this circuit will lead to a power-efficient low noise design.

On transistor linearization, as shown in Fig. 5.14, adjusting g''_m characteristics through body bias allows for flexible DC bias without use of bleeding current sources or ac coupling capacitors. This is a promising technique to further extend the bandwidth of the LNA.

The great advantage of noise and distortion cancelation schemes is that they do not rely on the performances of a single transistor, which usually becomes worse as technology scales. On the other hand, this means the distortion design is highly dependent

on the matching between two electronics parameters. It therefore is critical to reduce the parameter mismatch due to bias/process variations. In this regard, distortion cancelation is more sensitive than noise cancelation as more cancelation mechanisms are involved. Unfortunately, most conventional bias circuitries, i.e. PTAT (Proportional to Absolute Temperature), constant g_m [34], is not able to provide the g'_m and g''_m conditions needed for the proposed schemes. On the other hand, by assigning the LNA itself to detect its output distortion, and then feeding the information back to adjust its individual bias voltages, the output distortion level magnitude can be used to drive the feedback loop to automatically tune the circuitry, and is much favored. This built-in self-test capability involves on-chip distortion sensing and the associated mixed-mode signal processing. These techniques have been demonstrated in a recent work on spectral sensing of onchip supply noise [84]. Application of these techniques to the LNA design deserve further investigation towards a viable implementation.

Bibliography

[1] [Online]. Available: http://www.instat.com

[2] M. Gustafsson *et al.*, "A Low Noise Figure 1.2-V CMOS GPS Receiver Integrated as a Part of a Multimode Receiver," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 7, pp. 1492–1500, July 2007.

[3] T. Sowlati *et al.*, "Single-chip Multiband WCDMA/HSDPA/HSUPA/EGPRS Transceiver with Diversity Receiver and 3G DigRF Interface without SAW Filters in Transmitter / 3G Receiver Paths," in *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, Feb. 2009, pp. 116–117,117a.

[4] R. Gharpurey and S. Ayazian, "Feedforward Interference CanceLlation in Narrow-band Receivers," in *Design, Applications, Integration and Software, 2006 IEEE Dallas/CAS Workshop on*, Oct. 2006, pp. 67–70.

[5] F. Bruccoleri, E. Klumperink, and B. Nauta, "Noise Cancelling in Wideband CMOS LNAs," in *Solid-State Circuits Conference, 2002. Digest of Technical Papers. ISSCC. 2002 IEEE International*, vol. 1, 2002, pp. 406–407 vol.1.

[6] T. Zahariadis, "Migration toward 4G Wireless Communications," *Wireless Communications, IEEE*, vol. 11, no. 3, pp. 6–7, June 2004.

[7] [Online]. Available: http://www.3gpp.org/HSPA

[8] [Online]. Available: http://www.dvb-h.org/

[9] [Online]. Available: ftp://ftp.3gpp.org/Inbox/2008_web_files/LTA_Paper.pdf

[10] [Online]. Available: http://grouper.ieee.org/groups/802/16/

[11] [Online]. Available: http://www.fwdconcepts.com

[12] [Online]. Available: http://www.idc.com/

[13] S. D'Amico *et al.*, "A CMOS 5 nv/*surd*Hz 74-dB-Gain-Range 82-dB-DR Multistandard Baseband Chain for Bluetooth, UMTS, and WLAN," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 7, pp. 1534–1541, July 2008.

[14] V. Giannini, J. Craninckx, S. D'Amico, and A. Baschirotto, "Flexible Baseband Analog Circuits for Software-Defined Radio Front-ends," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 7, pp. 1501–1512, July 2007.

[15] A. Abidi, "Evolution of a Software-Defined Radio Receiver's RF Front-end," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2006 IEEE*, June 2006, pp. 4 pp.–.

[16] K. Muhammad *et al.*, "The First Fully Integrated Quad-band GSM/GPRS Receiver in a 90-nm Digital CMOS Process," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 8, pp. 1772–1783, Aug. 2006.

[17] R. Staszewski, K. Muhammad, and D. Leipold, "Digital Signal Processing for RF at 45-nm CMOS and Beyond," Sept. 2006, pp. 517–522.

[18] W. Y. Ali-Ahmad, "Rf System Issues Related to CDMA Receiver Specifications," *RF Design*, September 1999.

[19] C.-C. Nguyen, "Rf MEMS in Wireless Architectures," June 2005, pp. 416–420.

[20] S.-J. Park, K.-Y. Lee, and G. Rebeiz, "Low-loss 5.15-5.7 GHz RF MEMS Switchable Filter for Wireless LAN Applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 11, pp. 3931–3939, Nov. 2006.

[21] H. Darabi, "A Blocker Filtering Technique for Wireless Receivers," in *Solid-State Circuits Conference, 2007. Digest of Technical Papers. IEEE International*, Feb. 2007, pp. 84–588.

[22] A. Safarian, A. Shameli, A. Rofougaran, M. Rofougaran, and F. De Flaviis, "Integrated Blocker Filtering RF Front-ends," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE*, June 2007, pp. 13–16.

[23] D. Shaeffer and T. Lee, "A 1.5-V, 1.5-GHz CMOS Low Noise Amplifier," *Solid-State Circuits, IEEE Journal of*, vol. 32, no. 5, pp. 745–759, May 1997.

[24] T.-K. Nguyen, C.-H. Kim, G.-J. Ihm, M.-S. Yang, and S.-G. Lee, "CMOS Low-noise Amplifier Design Optimization Techniques," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, no. 5, pp. 1433–1442, May 2004.

[25] A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*. Cambridge, UK: Cambridge University Press, 2007.

[26] A. Bevilacqua and A. Niknejad, "An Ultrawideband CMOS Low-noise Amplifier for 3.1-10.6-GHz Wireless Receivers," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 12, pp. 2259–2268, Dec. 2004.

[27] D. Allstot, X. Li, and S. Shekhar, "Design Considerations for CMOS Low-noise Amplifiers," in *Radio Frequency Integrated Circuits (RFIC) Symposium, Digest of Papers. 2004 IEEE*, June 2004, pp. 97–100.

[28] S. Wang, A. Niknejad, and R. Brodersen, "A Sub-mw 960-MHz Ultra-wideband CMOS LNA," in *Radio Frequency integrated Circuits (RFIC) Symposium, Digest of Papers. 2005 IEEE*, June 2005, pp. 35–38.

[29] A. Liscidini, M. Brandolini, D. Sanzogni, and R. Castello, "A 0.13 μm CMOS Front-end, for DCS1800/UMTS/802.11b-g with Multiband Positive Feedback Low-noise Amplifier," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 4, pp. 981–989, April 2006.

[30] S. Shekhar, J. Walling, and D. Allstot, "Bandwidth Extension Techniques for CMOS Amplifiers," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 11, pp. 2424–2439, Nov. 2006.

[31] S. Bagga *et al.*, "A Band-reject IR-UWB LNA with 20 dB WLAN Suppression in 0.13 μm CMOS," Nov. 2008, pp. 449–452.

[32] M. Reiha and J. Long, "A 1.2 V Reactive-Feedback 3.110.6 GHz Low-noise Amplifier

in 0.13 *muhboxm* CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 5, pp. 1023–1033, May 2007.

[33] Razavi, *RF Microelectronics*, 1998.

[34] S. H. L. Paul R. Gray, Paul J. Hurst and robert G. Meyer, *Analysis and Design of Analog Integrated Circuits.* John Wiley & Sons, Inc.: Cambridge University Press, 2001.

[35] K. Han *et al.*, "Complete High-Frequency Thermal Noise Modeling of Short-Channel MOSFETs and Design of 5.2-GHz Low Noise Amplifier," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 3, pp. 726–735, March 2005.

[36] F. Gatta, E. Sacchi, F. Svelto, P. Vilmercati, and R. Castello, "A 2-dB Noise Figure 900-MHz Differential CMOS LNA," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 10, pp. 1444–1452, Oct 2001.

[37] P. Woerlee *et al.*, "Rf-CMOS Performance Trends," *Electron Devices, IEEE Transactions on*, vol. 48, no. 8, pp. 1776–1782, Aug 2001.

[38] C.-H. Chen and M. Deen, "Channel Noise Modeling of Deep Submicron MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 49, no. 8, pp. 1484–1487, Aug 2002.

[39] A. Scholten *et al.*, "Noise Modeling for RF CMOS Circuit Simulation," *Electron Devices, IEEE Transactions on*, vol. 50, no. 3, pp. 618–632, March 2003.

[40] S. Taylor and A. Waltho, "Multi-radio Platforms in Scaled CMOS Technology," in

Silicon Monolithic Integrated Circuits in RF Systems, 2007 Topical Meeting on, Jan. 2007, pp. 195–200.

[41] M. Brandolini, P. Rossi, D. Manstretta, and F. Svelto, "Toward Multistandard Mobile Terminals - Fully Integrated Receivers Requirements and Architectures," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, no. 3, pp. 1026–1038, March 2005.

[42] A. Bevilacqua and F. Svelto, "Non-Linear Spectral Analysis of Direct Conversion Wireless Receivers," in *Radio and Wireless Conference, RAWCON 2002. IEEE*, 2002, pp. 39–42.

[43] S. Mohammed and N. Yanduru, "Third-Order Nonlinear Distortion from Modulated Blockers in WCDMA Receivers," in *Circuits and Systems, MWSCAS '06. 49th IEEE International Midwest Symposium on*, vol. 2, Aug. 2006, pp. 473–476.

[44] F. Renna and S. Marsili, "A Gaussian Approximation of High-Order Distortion Spectrum in Broadband Amplifiers," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 55, no. 7, pp. 700–704, July 2008.

[45] V. Aparin and L. Larson, "Analysis and Reduction of Cross-Modulation Distortion in CDMA Receivers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 51, no. 5, pp. 1591–1602, May 2003.

[46] A. Springer, L. Maurer, and R. Weigel, "RF System Concepts for Highly Integrated RFICs for WCDMA Mobile Radio Terminals," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 50, no. 1, pp. 254–267, Jan 2002.

[47] P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*. Boston: Kluwer Academic Publishers, 1998.

[48] T. Soorapanth, , and T. H. Lee, "RF Linearity of Short-Channel MOSFETs," in *First International Workshop on Design of Mixed-Mode Integrated Circuits and Applications*, 1997, pp. 81–84.

[49] R. Bagheri *et al.*, "An 800-MHz6-GHz Software-Defined Wireless Receiver in 90-nm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 12, pp. 2860–2876, Dec. 2006.

[50] A. Liscidini, G. Martini, D. Mastantuono, and R. Castello, "Analysis and Design of Configurable LNAs in Feedback Common-Gate Topologies," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 55, no. 8, pp. 733–737, Aug. 2008.

[51] E. A. K. Federico Bruccoleri and B. Nauta, *Wideband Low Noise Amplifiers Exploiting Thermal Noise Cancellation*. Springer, 2001.

[52] W.-H. Chen, G. Liu, B. Zdravko, and A. Niknejad, "A Highly Linear Broadband CMOS LNA Employing Noise and Distortion Cancellation," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 5, pp. 1164–1176, May 2008.

[53] J. Borremans, P. Wambacq, C. Soens, Y. Rolain, and M. Kuijk, "Low-Area Active-Feedback Low-Noise Amplifier Design in Scaled Digital CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 11, pp. 2422–2433, Nov. 2008.

[54] M. Valkama, A. Ghadam, L. Anttila, and M. Renfors, "Advanced Digital Signal Pro-

cessing Techniques for Compensation of Nonlinear Distortion in Wideband Multicarrier Radio Receivers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 6, pp. 2356–2366, June 2006.

[55] E. Keehr and A. Hajimiri, "Equalization of IM3 Products in Wideband Direct-Conversion Receivers," in *Solid-State Circuits Conference, Digest of Technical Papers. 2008 IEEE International*, Feb. 2008, pp. 204–607.

[56] M. Faulkner, "Amplifier Linearization Using RF Feedback and Feedforward Techniques," *Vehicular Technology, IEEE Transactions on*, vol. 47, no. 1, pp. 209–215, Feb 1998.

[57] Y. Y. Woo, J. Yi, Y. Yang, and B. Kim, "SDR Transmitter Based on LINC Amplifier with Bias Control," in *Microwave Symposium Digest, 2003 IEEE MTT-S International*, vol. 3, June 2003, pp. 1703–1706 vol.3.

[58] Y. Ding and R. Harjani, "A +18 dBm IIP3 LNA in 0.35 μm CMOS," in *Solid-State Circuits Conference, Digest of Technical Papers. ISSCC. 2001 IEEE International*, 2001, pp. 162–163, 443.

[59] V. Aparin, G. Brown, and L. Larson, "Linearization of CMOS LNA's via Optimum Gate Biasing," in *Circuits and Systems, Proceedings of the 2004 International Symposium on*, vol. 4, May 2004, pp. IV–748–51 Vol.4.

[60] S. Ganesan, E. Sanchez-Sinencio, and J. Silva-Martinez, "A Highly Linear Low-Noise Amplifier," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 12, pp. 4079–4085, Dec. 2006.

[61] T. W. Kim, B. Kim, Y. Cho, B. Kim, and K. Lee, "A 13 dB IIP3 Improved Low-Power CMOS RF Programmable Gain Amplifier Using Differential Circuit Transconductance Linearization for Various Terrestrial Mobile D-TV Applications," in *VLSI Circuits, Digest of Technical Papers. 2005 Symposium on*, June 2005, pp. 344–347.

[62] R. Baki, T. Tsang, and M. El-Gamal, "Distortion in RF CMOS Short-Channel Low-Noise Amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 1, pp. 46–56, Jan. 2006.

[63] G. Niu, J. Pan, X. Wei, S. Taylor, and D. Sheridan, "Intermodulation Linearity Characteristics of CMOS Transistors in a 0.13 μm Process," in *Radio Frequency integrated Circuits (RFIC) Symposium, Digest of Papers. 2005 IEEE*, June 2005, pp. 65–68.

[64] S. Kang, B. Choi, and B. Kim, "Linearity Analysis of CMOS for RF Application," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 51, no. 3, pp. 972–977, Mar 2003.

[65] T. W. Kim, B. Kim, I. Nam, B. Ko, and K. Lee, "A Low-Power Highly Linear Cascoded Multiple-Gated Transistor CMOS RF Amplifier with 10 dB IP3 Improvement (Revised)," *Microwave and Wireless Components Letters, IEEE*, vol. 13, no. 9, pp. 420–422, Sept. 2003.

[66] V. Aparin and L. Larson, "Modified Derivative Superposition Method for Linearizing FET Low-Noise Amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, no. 2, pp. 571–581, Feb. 2005.

[67] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems.* New York: Wiley, 1980.

[68] I. Nam, B. Kim, and K. Lee, "CMOS RF Amplifier and Mixer Circuits Utilizing Complementary Characteristics of Parallel Combined NMOS and PMOS Devices," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, no. 5, pp. 1662–1671, May 2005.

[69] V. Aparin and L. Larson, "Linearization of Monolithic LNAs Using Low-Frequency Low-Impedance Input Termination," in *Solid-State Circuits Conference, ESSCIRC '03. Proceedings of the 29th European*, Sept. 2003, pp. 137–140.

[70] T. W. Kim, B. Kim, and K. Lee, "Highly Linear Receiver Front-End Adopting MOSFET Transconductance Linearization by Multiple Gated Transistors," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 1, pp. 223–229, Jan. 2004.

[71] Y. Tsividis, *Operation and Modeling of The MOS Transistor.* MacGraw-Hill, 1999.

[72] A. V. D. Ziel, *Noise in Solid State Devices and Circuits.* Wiley-Interscience, 1986.

[73] D. M. Pozar, *Microwave Engineering.* John Wiley & Sons, Inc., 1998.

[74] [Online]. Available: http://cp.literature.agilent.com/litweb/pdf/5952-3706E.pdf

[75] A. M. N. Wei-Hung Chen, "A Signal-Locus-Based Distortion Characterization Method and Its Application to Amplifier Linearization," *The 34th Government Microcircuit Applications and Critical Technology Conference*, March 2009.

[76] V. Aparin, *Linearization of CDMA Receiver Front-Ends*. Ph.D. thesis, UC San Diego, 2005.

[77] Y. Tsividis, K. Suyama, and K. Vavelidis, "Simple 'Reconciliation' MOSFET Model Valid in All Regions," *Electronics Letters*, vol. 31, no. 6, pp. 506–508, Mar 1995.

[78] C.-F. Liao and S.-I. Liu, "A Broadband Noise-Canceling CMOS LNA for 3.1-10.6-GHz UWB Receiver," in *Custom Integrated Circuits Conference, Proceedings of the IEEE 2005*, Sept. 2005, pp. 161–164.

[79] S. Chehrazi, A. Mirzaei, R. Bagheri, and A. Abidi, "A 6.5 GHz Wideband CMOS Low Noise Amplifier for Multi-Band Use," in *Custom Integrated Circuits Conference, Proceedings of the IEEE 2005*, Sept. 2005, pp. 801–804.

[80] R. Gharpurey, "A Broadband Low-Noise Front-End Amplifier for Ultra Wideband in 0.13-μm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1983–1986, Sept. 2005.

[81] R. Ramzan, S. Andersson, J. Dabrowski, and C. Svensson, "A 1.4V 25mw Inductorless Wideband LNA in 0.13μm CMOS," in *Solid-State Circuits Conference, Digest of Technical Papers. IEEE International*, Feb. 2007, pp. 424–613.

[82] S. Blaakmeer, E. Klumperink, D. Leenaerts, and B. Nauta, "Wideband Balun-LNA with Simultaneous Output Balancing, Noise-Canceling and Distortion-Canceling," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 6, pp. 1341–1350, June 2008.

[83] J.-H. Zhan and S. Taylor, "A 5GHz Resistive-Feedback CMOS LNA for Low-Cost

Multi-Standard Applications," in *Solid-State Circuits Conference, Digest of Technical Papers. IEEE 2006*, Feb. 2006, pp. 721–730.

[84] E. Alon and M. Horowitz, "Integrated Regulation for Energy-Efficient Digital Circuits," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 8, pp. 1795–1807, Aug. 2008.

[85] E. Wu and A. van der Ziel, "Induced-Gate Thermal Noise in High Electron Mobility Transistors," *Solid-State Electronics*, vol. 26, no. 7, pp. 639–642, July 1983.

Appendix A

MOSFET noise current

A.1 Drain current thermal noise

As shown in Fig. A.1, partition the MOSFET channel length into many small segments, dx. Assume the channel potential is V_0 at location x_0, and a noise voltage Δ_{v_x} is caused by the thermal noise agitation within d_x. Treat the channel length on both sides of dx as two back-to-back transistors operating at the same current level. According to [71], the noiseless drain current in the absence of velocity saturation and short-channel effects is

$$\begin{aligned} I_{ds} &= -\mu W Q'_{inv} \frac{dV(x)}{dx} \\ &= -\frac{\mu W}{L} \int_{V_S}^{V_D} Q'_{inv} dV(x) \end{aligned} \tag{A.1}$$

where $V(x)$ and Q'_{inv} are the channel potential and the inversion charge at x_0, respectively. V_S and V_D are the source and drain voltage with respect to the bulk. The local conductance

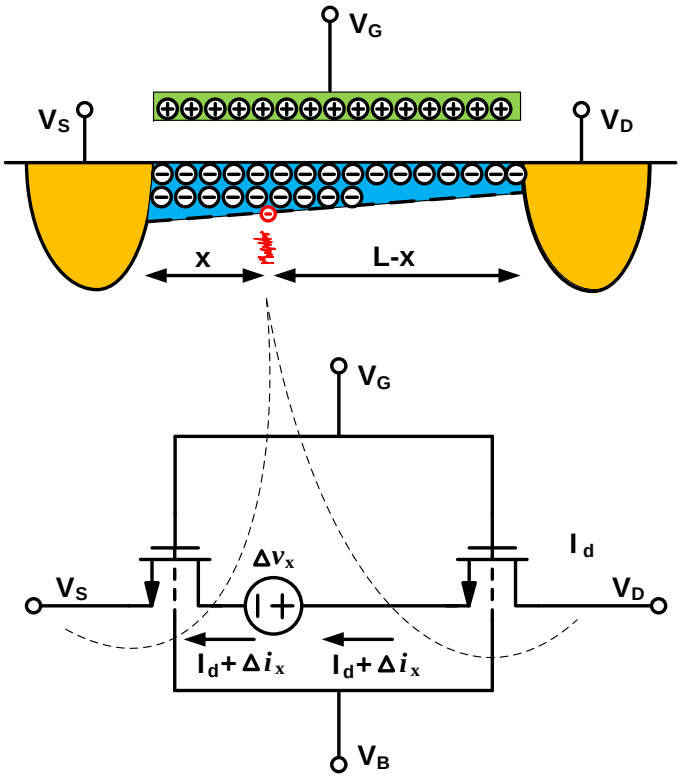


Figure A.1: transistor noise model

of segment dx is defined by

$$g(V(x)) = \mu W Q'_{inv} \tag{A.2}$$

The power spectral density of the thermal noise voltage due to dx is

$$\overline{\triangle v_x^2} = \frac{4ktdx}{g(V(x))} \triangle f \tag{A.3}$$

where $\triangle f$ is the observation bandwidth. The small noise voltage induces variations of channel current,$\triangle i_x$, at x_0. By maintaining

$$\begin{aligned} I_{ds} + \triangle i_x &= \frac{W}{x} \mu \int_{V_S}^{V_0} Q'_{inv} dV(x) \\ I_{ds} + \triangle i_x &= \frac{W}{L-x} \mu \int_{V_0+\triangle v_x}^{V_D} Q'_{inv} dV(x) \end{aligned} \tag{A.4}$$

for the back-to-back segmented transistors, the variation of channel current, $\triangle i_x$, due to small noise voltage, $\triangle v_x$, is obtained as

$$\triangle i_x = \frac{W}{L}\mu Q'_{inv}\triangle v_x \tag{A.5}$$

Furthermore, assume the noise of different channel segments is uncorrelated, the mean square value of the entire channel noise is equal to the sum of the individual mean square values of each segment. Integrating over along channel length L, the power spectral density of the overall noise becomes

$$\begin{aligned}\overline{i_{nd}^2} &= 4kt\frac{\mu}{L^2}\int_{x=0}^{x=L}(WQ'_{inv})dx\triangle f \\ &= 4kt\frac{\mu}{L^2}Q_{inv}\triangle f\end{aligned} \tag{A.6}$$

where Q_{inv} is the total inversion charge along the channel.

A.2 Induced-gate noise

Referring to the same schematic in Fig. A.1, consider the drain thermal noise $\triangle v_x$ induces a charge $\triangle q$ in the gate,

$$\triangle q = \int WC_{ox}\triangle v_x dx \tag{A.7}$$

The channel potential of the back-to-back transistor at $x > x_0$ is boosted by $\triangle q$, and induces extra charges on the associated capacitor plates. $\triangle q$ however, creates a potential imbalance between the capacitor plate of the two back-to-back transistors. The induced charges then redistribute until a potential equilibrium is reached. By first-order approximation of the

quasi-static equilibrium, the net charge induced by segment dx at x_0 is

$$\begin{aligned} \triangle q &= WC_{ox}\left(\int_0^L \triangle v_{x_0}\frac{x}{L}dx - \int_{x_0}^L \triangle v_{x_0}dx\right) \\ &= \frac{WC_{ox}g(V_0)\triangle v_{x_0}}{L}\left(\int_0^L \frac{xdx}{g(V)} - \int_{x_0}^L \frac{Ldx}{g(V)}\right) \end{aligned} \tag{A.8}$$

Notice that Eq. A.6 and Eq. A.8 are both valid for strong inversion and weak inversions.

A.2.1 strong inversion

Recall from Eq. A.1 and Eq. A.2 that

$$\begin{aligned} I_D &= g(V)\frac{dV}{dx} \\ &= \frac{1}{L}\int_0^{V_D} g(V)dV \end{aligned} \tag{A.9}$$

Referring to [71], the channel potential V as a function of x in a long-channel transistor is simplified as

$$x = L\frac{V_{GST}V - \frac{1}{2}V^2}{V_{GST}V_D - \frac{1}{2}V_D^2} \tag{A.10}$$

where $V_{GST} = V_{gs} - V_{th}$. For ideal square-law MOSFET, $g(V) = \mu C_{ox}W(V_{GST} - V)$ and $V_{GST} = V_D$. Plugging Eq. A.10 into Eq. A.8, we obtain

$$\triangle q = \frac{g(V_0)WC_{ox}}{I_D}(V_a - V_0)\triangle v_{x_0} \tag{A.11}$$

and

$$\triangle i_g = j\omega\triangle q = \frac{g(V_0)WC_{ox}}{I_D}(V_a - V_0)\triangle v_{x_0} \tag{A.12}$$

where

$$V_a = V_D\frac{\frac{1}{2}V_{GST} - \frac{1}{3}V_D}{V_{GST} - \frac{1}{2}V_D} \tag{A.13}$$

The overall induced gate current is found by substituting Eq. A.3 and Eq. A.9 into the integration of Eq. A.12 so that

$$\begin{aligned}\overline{i_{ng}^2} &= \frac{\omega^2 W^2 C_{ox}^2}{I_D^2} \int g(V_0)^2 (V_a - V_0)^2 \triangle v_{x_0}^2 \\ &= \frac{\omega^2 W^2 C_{ox}^2}{I_D^3} 4kt \int_0^{V_D} g(V_0)^2 (V_a - V_0)^2 dV_0 \triangle f\end{aligned} \tag{A.14}$$

Using $V_a = \frac{1}{12} V_D$ from Eq. A.13,

$$\overline{i_{ng}^2} = 4kt \frac{\omega^2 C_{gs}^2}{g_m} \frac{16}{135} \triangle f \tag{A.15}$$

By combining Eq. A.15 and Eq. A.6 with $Q_{inv} = \frac{2}{3} C_{ox} W L (V_{GS} - V_{th})$,

$$\begin{aligned}\overline{i_{nd}^2} &= 4kt \frac{2}{3} \mu C_{ox} \frac{W}{L} (V_{gs} - V_{th}) \quad \text{, and} \\ \overline{\triangle i_{nd} \triangle i_{ng}^*} &= 4ktj\omega C_{gs} \frac{1}{9} \triangle f\end{aligned} \tag{A.16}$$

, the correlation factor between drain thermal noise and induced-gate noise is found to be

$$c = \frac{\overline{\triangle i_{nd} \triangle i_{ng}^*}}{\sqrt{\overline{i_{nd}^2}\,\overline{i_{ng}^2}}} = j0.395 \tag{A.17}$$

A.2.2 weak inversion

In weak inversion, drain current is caused by the diffusion of the majority carrier in the substrate. According to [66, 85],

$$g(V) = g_0 e^{\frac{-V}{V_T}} \tag{A.18}$$

where $V_T = \frac{KT}{q}$, g_0 is the channel conductance per unit length at $V = 0$, i.e. the source terminal. From Eq. A.9

$$I_D = \frac{g_0 V_T}{L} (1 - e^{-\frac{V_D}{V_T}}) \tag{A.19}$$

For $V_D >> V_T$, Eq. A.19 simplifies to

$$I_D \approx \frac{g_0 V_T}{L} \tag{A.20}$$

Similar to Eq. A.10, the channel potential as a function of the position in the channel, or vice versa, is

$$x = \int_0^V \frac{g(u)du}{I_D} = \frac{g_0 V_T}{I_D}\left(1 - e^{-\frac{V}{V_D}}\right) \tag{A.21}$$

Substituting Eq. A.21 and Eq. A.18 into Eq. A.8 and rewriting it in the form of the gate -induced current,

$$\triangle i_{ng} \approx \frac{j\omega W C_{ox} g(V_0) \triangle v_{x0}}{I_D}\left(V_T - V_0\right) \tag{A.22}$$

By integrating Eq. A.22, the total gate noise current is

$$\begin{aligned} \overline{i_{ng}^2} &= \int_0^L \frac{(\overline{\triangle i_{ng}^2})}{dx} dx \\ &= 4kt\triangle f \frac{\omega^2 C_{ox}^2 W^2}{I_D^3} \int_0^{V_D} g^2(V)\left(V_T - V\right)^2 dV \\ &\approx KT\triangle f \omega^2 C_{ox}^2 W^2 L^2 \frac{V_T}{I_D} \end{aligned} \tag{A.23}$$

Replace $\frac{I_D}{V_T}$ with g_m, we get

$$\overline{i_{ng}^2} = KT\triangle f \frac{\omega^2 C_{gs}^2}{g_m} \tag{A.24}$$

Repeating Eq. A.16 with Eq. A.6 and Eq. A.24,

$$\begin{aligned} \overline{\triangle i_{nd} \triangle i_{ng}^*} &= 4kt \frac{j\omega C_{ox} W}{I_D^2 L} \int_0^{V_D} g^2(V_0)(V_T - V_0) dV_0 \triangle f \\ &\approx KT j\omega C_{ox} W L \triangle f \end{aligned} \tag{A.25}$$

With Eq. A.25 and in [66], the correlation factor for the transistor in weak inversion is shown to be

$$c = \frac{\overline{\triangle i_{nd} \triangle i_{ng}^*}}{\sqrt{\overline{i_{nd}^2}\,\overline{i_{ng}^2}}} = j0.707 \tag{A.26}$$

Appendix B

Complete Volterra series analysis

Referring to Fig. 5.1, the node voltage is expressed as

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

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

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

and a set of basic KCL equations is established,

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

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

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

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

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

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

Substituting i_{m1} and i_{m2} with their g_m polynomial, we obtain the following for the 1^{st}-order Volterra kernel.

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

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

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

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

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

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

$$H(s) = Z_s(s)\Big(1 + (g_{m1} + g_{m2})(r_{o1} \parallel r_{o2})\Big) + \Big((Z_1(s) \parallel Z_2(s)) + (r_{o1} \parallel r_{o2})\Big)\left(1 + \frac{Z_s(s)}{Z_x(s)}\right) \tag{B.14}$$

Next, consider the following equations for the 2^{nd}-order response.

$$\begin{aligned} &g_{m1}A_2(s_1, s_2) - \frac{g'_{m1}}{2}A_1(s_1)A_1(s_2) + \frac{A_2(s_1, s_2) - B_2(s_1, s_2)}{r_{o1}} + \\ &g_{m2}A_2(s_1, s_2) + \frac{g'_{m2}}{2}A_1(s_1)A_1(s_2) + \frac{A_2(s_1, s_2) - C_2(s_1, s_2)}{r_{o2}} + \frac{A_2(s_1, s_2)}{Z_x(s_1 + s_2)} = \frac{-A_2(s_1, s_2)}{Z_s(s_1 + s_2)} \end{aligned} \tag{B.15}$$

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

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

$A_2(s_1, s_2)$ and $B_2(s_1, s_2)$ are found as

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

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

where

$$\begin{aligned}\triangle A_2(s_1,s_2) =& \frac{1}{2}Z_{12}(s_1+s_2)A_1(s_1)A_1(s_2)\frac{Z_s(s_1+s_2)}{Z_1(s_1+s_2)+Z_2(s_1+s_2)}\times \\ & \Big((g'_{m1}-g'_{m2})(r_{o1}\parallel r_{o2})+\frac{g'_{m1}r_{o1}Z_2(s_1+s_2)-g'_{m2}r_{o2}Z_1(s_1+s_2)}{r_{o1}+r_{o2}}\Big)\end{aligned} \tag{B.20}$$

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

$$\begin{aligned}\triangle H(s_1,s_2) =& Z_{12}(s_1+s_2)\frac{Z_s(s_1,s_2)}{Z_1(s_1,s_2)+Z_2(s_1,s_2)}\frac{1}{r_{o1}+r_{o2}}\times \\ & \Bigg(\frac{(r_{o1}+Z_1(s_1+s_2))(r_{o2}+Z_2(s_1+s_2))}{Z_x(s_1+s_2)\parallel Z_s(s_1+s_2)}+ \\ & \Big((1+g_{m1}r_{o1})\big(r_{o2}+Z_2(s_1+s_2)\big)+(1+g_{m2}r_{o2})\big(r_{o1}+Z_1(s_1+s_2)\big)\Big)\Bigg)\end{aligned} \tag{B.22}$$

$C_2(s_1,s_2)$ is found by interchanging the element notation in $B_2(s_1,s_2)$ because of the circuit duality. Notice that $\triangle A_2(s_1,s_2)$, $\triangle B_2(s_1,s_2)$ and $\triangle H(s_1,s_2)$ drop out if $Z_{12}(s_1+s_2)$ is zero. The 2^{nd}-order distortion is canceled if g'_{m1} is matched to g'_{m2}.

Continue with the 3^{rd}-order response,

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

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

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

Assume again that $Z_{12}(s_1+s_2+s_3)$ is negligible then we arrive at

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

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