Design of Sub-mW RF CMOS Low-Noise Amplifiers

by

DEREK HO

B. A. Sc., The University of British Columbia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

March 2007

© Derek Ho, 2007
Abstract

The quest for low power, low cost, and highly integrated transceivers has gained substantial momentum due to the explosion of wireless applications such as personal area networks and wireless sensor networks. This dissertation presents a comprehensive study and a design methodology for power-efficient CMOS radio-frequency (RF) low-noise amplifiers (LNAs).

To demonstrate the design methodology, a sub-mW fully integrated narrow-band source degenerated cascode RF LNA is designed and simulated in a standard 90nm CMOS process to operate in the 2.4GHz band. The LNA achieves a voltage gain of 22.7dB, noise figure (NF) of 2.8dB, 3rd-order intercept point (IIP3) of +5.14dBm, and 1dB compression point (P1dB) of −10dBm, while consuming 943μW from a 1V supply.

The main contributions of this work include: i) the introduction of a design methodology for power-efficient sub-mW source degenerated LNAs; ii) the collection of design graphs to facilitate the exploration of tradeoffs between LNA performance and power consumption; and iii) the use of an alternative analysis to find the dependency of gain, noise, and linearity on biasing conditions.
# Table of Contents

ABSTRACT .................................................................................................................. ii

TABLE OF CONTENTS ............................................................................................... iii

LIST OF TABLES ........................................................................................................... v

LIST OF FIGURES ......................................................................................................... vi

ACKNOWLEDGMENTS .................................................................................................. vii

CHAPTER 1. INTRODUCTION ....................................................................................... 1
  MOTIVATION .............................................................................................................. 1
  RESEARCH OBJECTIVE ............................................................................................ 2
  THESIS ORGANIZATION ......................................................................................... 3

CHAPTER 2. BACKGROUND ......................................................................................... 5
  TWO-PORT NETWORKS .............................................................................................. 5
  IMPEDANCE MATCHING .......................................................................................... 6
  SCATTERING PARAMETERS .................................................................................... 7
  LINEARITY ................................................................................................................ 9
    Harmonic Distortion ............................................................................................... 9
    Intermodulation .................................................................................................... 11
    1 dB Compression Point ....................................................................................... 13
    3rd Order Intercept Point ..................................................................................... 13
  STABILITY ................................................................................................................ 14
  NOISE IN TWO-PORT SYSTEMS ............................................................................. 15

CHAPTER 3. CMOS LNA FUNDAMENTALS ................................................................. 19
  NOISE SOURCES IN CMOS .................................................................................... 19
    Thermal Noise ....................................................................................................... 19
    Induced Gate Noise .............................................................................................. 20
    Distributed Gate Noise ........................................................................................ 22
  INTRINSIC MOSFET TWO-PORT NOISE PARAMETERS ........................................ 23
  INDUCTIVE SOURCE DEGENERATION ................................................................ 24
    Input Impedance Match ....................................................................................... 25
    Circuit Transconductance .................................................................................... 26
  CIRCUIT TOPOLOGIES ........................................................................................... 27
  PERFORMANCE OF THE CASCODE AMPLIFIER .................................................. 29
    Voltage and Power Gain ...................................................................................... 30
    Noise ..................................................................................................................... 30
    Linearity ................................................................................................................. 31
  INDUCTOR DESIGN .................................................................................................. 33
    Physical Dimensions ............................................................................................ 33
    Inductor Figures of Merit ..................................................................................... 34
    Inductor Modeling ................................................................................................. 36

CHAPTER 4. DESIGNING FOR POWER-EFFICIENT OPERATION ................................ 37
  DEVICE BIASING ...................................................................................................... 37
    Terminal I-V Characteristics ................................................................................ 38
    Gain and Transconductance ............................................................................... 40
  TRANSISTOR SIZING ............................................................................................... 44
  STEP-BY-STEP LNA DESIGN METHODOLOGY .................................................... 46
List of Tables

Table 1 Summary of LNA component values ................................................................. 59
Table 2 Summary of LNA performance ........................................................................ 61
Table 3 Comparison of CMOS low-power LNAs ......................................................... 62
List of Figures

Chapter 2

Figure 2.1 Two-port network diagram.................................................. 5
Figure 2.2 A system with a source driving a load.................................. 6
Figure 2.3 S parameter representation of a two-port network....................... 7
Figure 2.5 Frequency locations of distortion terms.................................. 12
Figure 2.4 Noise modeling. (a) Noisy two-port, (b) Input-referred noise model.... 15

Chapter 3

Figure 3.1 MOSFET small-signal model with thermal noise......................... 19
Figure 3.2 Induced gate noise model. (a) Frequency dependent, (b) Frequency independent... 21
Figure 3.3 Inductive source degeneration............................................. 25
Figure 3.4 Narrow-band topologies. (a) Common-source, (b) Cascode.................. 27
Figure 3.5 VIP3 of a 40nm nFET vs. $V_{GS}$ for different $V_{DS}$ ($V_{TH}$ = 0.23V) [22]............................... 32
Figure 3.6 The layout of a square spiral inductor.................................... 33
Figure 3.7 Pi-model of inductor....................................................... 36

Chapter 4

Figure 4.1 $I_D$ vs. $V_{GS}$ for a 90nm nFET ($V_{DS}$ = 1V)........................................... 38
Figure 4.2 $I_D$ vs. $V_{DS}$ for a 90nm nFET ($W/L = 20 \mu m/0.1 \mu m$, $V_{DS}$ = 1V).................. 39
Figure 4.3 $f_T$ vs. $V_{GS}$ for a 90nm nFET ($V_{DS}$ = 1V)................................. 40
Figure 4.4 $g_m$ vs. $V_{GS}$ for a 90nm nFET ($V_{DS}$ = 1V)............................. 41
Figure 4.5 $g_m/I_D$ vs. $V_{GS}$ for a 90nm nFET ($V_{DS}$ = 1V).............................. 42
Figure 4.6 $r_o$ ($=1/g_{ds}$) vs. $V_{DS}$ for a 90nm nFET ($V_{DS}$ = 1V).................... 43
Figure 4.7 Intrinsic gain $g_m r_o$ ($= g_m/g_{ds}$) vs. $V_{GS}$ for a 90nm nFET ($V_{DS}$ = 1V)............. 44
Figure 4.8 NF vs. $W$ of cascode and common-source amplifiers.................... 45
Figure 4.9 Cascode amplifier. (a) Schematic, (b) Simplified small-signal model for input matching analysis................................. 46

Chapter 5

Figure 5.1 Schematic of a cascode amplifier.......................................... 52
Figure 5.2 $I_D$ of a cascode LNA. (a) $I_D$ vs. $V_{GS}$ ($W_1=W_2=25 \mu m$), (b) $I_D$ vs. $W$ ($V_{GS}$=0.4V)............. 54
Figure 5.3 Voltage gain of a cascode LNA. (a) $A_v$ vs. $V_{GS}$, (b) $A_v$ vs. $W$................................. 55
Figure 5.4 NF of a cascode LNA. (a) NF vs. $V_{GS}$ and (b) NF vs. $W$....................... 56
Figure 5.5 Noise contribution of the signal source and LNA components............ 57
Figure 5.6 $IIP3$ vs. $V_{GS}$ of the cascode LNA.......................................... 58
Figure 5.7 $P_{1dB}$ vs. $V_{GS}$ of the cascode LNA......................................... 58
Figure 5.8 $\pi$-model of a 5nH spiral inductor........................................... 59
Figure 5.9 Gain and NF of the proposed LNA........................................... 60
Figure 5.10 $S_{11}$ of the proposed LNA.................................................. 61
Figure 5.11 Graphical illustration comparing recently published LNAs and this work... 62
Acknowledgments

This work\textsuperscript{1} could not have been completed without the help of many people and I would like to take this opportunity to express my appreciation to them.

I would like to start by thanking my graduate advisor, Professor Shahriar Mirabbasi for his guidance and for always making decisions based on what is best for my education. This has led to not only a productive but also an enjoyable experience. I would like to thank Dr. Andre Ivanov for supervising my undergraduate research in the System-on-Chip lab, where I cultivated my early interest in VLSI circuits. I would also like to thank Professor Resve Saleh and Professor Steve Wilton for reading this dissertation and serving as committee members in my thesis defense.

I could not imagine where I would be without the help of Roberto Rosales and Karim Allidina. They deserve special recognition for having the patience to explain to me any question that I threw at them.

I probably would have gone insane without the companions at the SOC lab to listen to all my problems and give me advice (even though some advice was insane too). You know who you are: Usman Ahmed, Shirley Au, Nathalie Chan, Melody Chang, Scott Chin,

\textsuperscript{1} This research is funded by Natural Science and Engineering Research Council of Canada (NSERC). CAD tools are provided by Canadian Microelectronics Corporation (CMC) Microsystems.
David Chiu, Rod Foist, Amit Kedia, Sohaib Majzoub, Xiongfei Meng, Dipanjan Sengupta, and Howard Yang.

I am sure I would be a different person without the influence of Wilson Fung and Jennifer Li. Being friends for more than a decade, Wilson and I grew up together. His commitment to doing good work has become my role model. Having a technical discussion with him is nothing less than inspiring. Jennifer and I have gone through both ups and downs. Her love is indispensable to my general well-being.

Last but not least, I would like to thank my sister for putting up with me for as long as she has existed. I'd also like to express my deepest gratitude to my parents for their unconditional love. They have made tremendous sacrifices to give me the best upbringing a child can possibly receive.

Derek Ho

Vancouver, BC
Chapter 1

Introduction

Motivation

The design of low-power wireless transceivers has gained substantial significance due to the explosion of wireless applications such as personal area networks and wireless sensor networks. These applications demand for small, low-cost, and low-power wireless transceivers which require a high level of integration with a minimal amount of off-chip components. The first active block in most wireless receivers is the low-noise amplifier (LNA). The LNA needs to amplify the signal without adding a large amount of noise and distortion while consuming minimal power.

RF circuits have traditionally been implemented in compound semiconductor technologies such as Gallium Arsenide (GaAs) or Silicon Germanium (SiGe). Since the CMOS technology is employed for the digital transceiver back-end, it is attractive to implement the RF front-end also in CMOS, with the goal to integrate all parts of the receiver on a single chip to reduce cost and time to market. In the recent past, numerous CMOS RF circuits have been presented and have demonstrated good performance [1]-[4].
The design of a power-efficient LNA in CMOS is particularly challenging due to a number of reasons:

1) The performance such as gain and bandwidth of a metal-oxide-semiconductor field-effect transistor (MOSFET) is poor when biased at the small drain current necessary for power-efficient operation.

2) The choice of circuit topologies is limited due to the reduction of MOSFET output resistance and supply voltage as CMOS technology scales.

3) On-chip passive elements have poor quality factor, limited range of values, and large area consumption.

4) Conventional power-constrained noise optimization techniques [5] often lead to subthreshold operation when power consumption is restricted to 1mW or below.

Research Objective

The objective of this thesis is to develop a narrow-band RF CMOS LNA design technique suitable for the prevalent inductively degenerated LNA architecture. This work aims to achieve the following targets:

1) Devise a methodology that leads to a power-efficient LNA design.

2) Explore the tradeoffs between LNA performance and power consumption.

3) Find a circuit topology capable of low-voltage low-power operation.

4) Demonstrate a high performance design with on-chip low-\(Q\) passive components in a deep submicron technology.
In devising the design methodology, a major goal is to review, coordinate, and exploit several device-level properties proposed in the recent literature:

1) Device transit frequency $f_T$ and unity power gain frequency $f_{MAX}$ depend strongly on drain current density, which is mainly controlled by the gate-source voltage [6].

2) The MOSFET has a large transconductance per unit drain current $g_{m}/I_D$ in weak inversion and progressively degrades towards strong inversion [7].

3) The MOSFET minimum noise figure improves as gate length decreases [6], [8].

4) MOSFET linearity improves as drain current density increases with a significant peaking in the moderate inversion region [9], [10].

As part of the overall objective, a 2.4GHz LNA is designed and simulated in a 90nm CMOS technology with sub-mW power consumption using the proposed design methodology. The simulation results demonstrate the applicability of the methodology to the design of narrow-band RF front-ends needed for many low-power applications.

**Thesis Organization**

In this thesis, an LNA design methodology is devised and documented. Chapter 2 presents background information about two-port networks that is fundamental to the design of LNAs and other RF circuits. Chapter 3 focuses on the design of CMOS LNAs. Chapter 4 discusses the design of transistor biasing conditions for power-efficient amplifiers. This chapter concludes with a step-by-step design procedure.
details the application of the proposed methodology to the design of a sub-mW LNA. Simulation results are presented at the end of this chapter. Finally, Chapter 6 draws some conclusions and suggests future works.
Chapter 2

Background

To simplify the design and analysis of analog circuits, it is useful to abstract circuit blocks into two-port networks. This chapter begins with a discussion of parameters that are used to characterize two-port networks. Then, performance measures of the two-port network such as gain, noise, linearity, and stability that are important in the design of LNAs are presented.

Two-Port Networks

A two-port network is shown in Fig. 2.1. There are usually two quantities, namely, voltage and current associated to each port. At low frequencies, two common representations that characterizes the network are the impedance matrix (Z parameters) and the admittance matrix (Y parameters) [11], [12].

![Two-Port Network Diagram](image)

Figure 2.1 Two-port network diagram.
The impedance and admittance matrices are defined by (2.1) and (2.2), respectively.

\[
\begin{bmatrix}
  v_1 \\
  v_2 
\end{bmatrix} =
\begin{bmatrix}
  Z_{11} & Z_{12} \\
  Z_{21} & Z_{22}
\end{bmatrix}
\begin{bmatrix}
  i_1 \\
  i_2
\end{bmatrix}
\]  

(2.1)

\[
\begin{bmatrix}
  i_1 \\
  i_2
\end{bmatrix} =
\begin{bmatrix}
  Y_{11} & Y_{12} \\
  Y_{21} & Y_{22}
\end{bmatrix}
\begin{bmatrix}
  v_1 \\
  v_2
\end{bmatrix}
\]  

(2.2)

Z parameters and Y parameters are particularly useful at low frequencies because they can be readily measured by applying either a test current or voltage to the input port and connecting the output port either as a short or open circuit. For example, from (2.1) \(v_1\) can be expressed as

\[v_1 = Z_{11}i_1 + Z_{12}i_2.\]  

(2.3)

If the output port is open circuited, \(i_2\) becomes zero, and \(Z_{11}\) can be calculated to be \(v_1/i_1\).

**Impedance Matching**

A system with a signal source driving a load is depicted in Fig. 2.2 where \(Z_S\) and \(Z_L\) are the source and load impedances, respectively.

\[Z_S = R_S + jX_S\]

\[Z_L = R_L + jX_L\]

\[V_s\]

Source  Load

Figure 2.2  A system with a source driving a load.
There are two types of impedance matching [13]. The first type of impedance matching concerns with minimizing signal reflection from the load back to the source. When \( Z_S = Z_L \), there is no reflection. This is important for the design of the receiver front-end as the frequency response of the antenna filter that precedes the LNA deviates from its normal operation if there are reflections from the LNA back to the filter. The second type of matching concerns with maximum power transfer from the source to the load. Hence it is often referred to as power matching. Power matching occurs when the load impedance is the complex conjugate of the source impedance. When the source and load impedances are real as in a typical 50Ω RF system, the conditions for power matching and impedance matching are equal.

**Scattering Parameters**

At RF and microwave frequencies, Z and Y parameters become very difficult to measure due to the need for broadband short and open circuits [14]. As a result, a different representation of the two-port network is needed at these frequencies. A popular representation is the scattering, or S, parameters. Instead of relying on ports being open and short circuited, S parameters have the advantage that they can be measured by matching the source and load impedances to a reference impedance \( Z_0 \). An S parameter representation of a two-port network is shown in Fig. 2.3.

![Figure 2.3 S parameter representation of a two-port network.](image-url)
The notion of the S parameter representation is to measure the normalized incident voltage wave $a_i$ entering the system at port $i$, as well as the corresponding reflected voltage wave $b_i$ leaving port $i$. The normalized incident and reflected voltage waves $a_i$ and $b_i$ are related to the terminal voltage and current at port $i$ by the following equations:

$$a_i = \frac{v_i + Z_o i_i}{2\sqrt{Z_o}}$$  \hspace{1cm} (2.4)

$$b_i = \frac{v_i - Z_o i_i}{2\sqrt{Z_o}}$$  \hspace{1cm} (2.5)

where $Z_o$ is assumed real as it is usually equal to 50Ω. The network of Fig. 2.3 can be express, in matrix form, as (2.6).

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$  \hspace{1cm} (2.6)

where $S_{11}$, $S_{12}$, $S_{21}$, $S_{22}$ are the scattering parameters. By expanding the scattering matrix, the following equations can be written:

$$S_{11} = \left. \frac{b_1}{a_1} \right|_{a_2=0}$$  \hspace{1cm} (2.7)

$$S_{12} = \left. \frac{b_1}{a_2} \right|_{a_1=0}$$  \hspace{1cm} (2.8)

$$S_{21} = \left. \frac{b_2}{a_1} \right|_{a_2=0}$$  \hspace{1cm} (2.9)

$$S_{22} = \left. \frac{b_2}{a_2} \right|_{a_1=0}$$  \hspace{1cm} (2.10)

where $S_{11}$ is interpreted as the ratio of the reflected voltage wave to the incident voltage wave at port 1 with the output port properly terminated. The condition for a port being
properly terminated is that the impedance looking into the port must match the characteristic impedance of the transmission line attached to it. Definitions for the rest of the S parameters can be interpreted analogously.

**Linearity**

Linearity is a key requirement in the design of an LNA because the LNA must be able to maintain linear operation in the presence of a large interfering signal and when the input is driven by a large signal [14]. Intermodulation linearity, characterized by the input-referred 3rd-order intermodulation intercept point (IIP3), is crucial to prevent the intermodulation tones created by a large interfering signal from corrupting the signal of interest. Large-signal linearity, characterized by the input-referred 1dB voltage compression point (P1dB), is important as it determines the maximum input level that can be amplified linearly, i.e. the upper bound of the dynamic range of the LNA.

**Harmonic Distortion**

The input $V_i$ and output $V_o$ of a two-port network can be related by a power series [15]:

$$V_o = a_1 V_i + a_2 V_i^2 + a_3 V_i^3 + \ldots$$  \hspace{1cm} (2.11)

where $a_1, a_2, a_3$ are constants. If the input is driven by a sinusoidal signal as follows,

$$V_i(t) = V_i \cos(\omega_1 t)$$  \hspace{1cm} (2.12)

where $V_i$ and $\omega_1$ are the amplitude and frequency, respectively, then the output is equal to

$$V_o(t) = a_1 V_i \cos(\omega_1 t) + \frac{a_2 V_i^2}{2} \left[ \cos(2\omega_1 t) + 1 \right]$$

$$+ \frac{a_3 V_i^3}{4} \left[ \cos(3\omega_1 t) + 3\cos(\omega_1 t) \right] + \ldots$$  \hspace{1cm} (2.13)
The first term in (2.13) is the linear term, and is the ideal output if the two-port network is completely linear. Other terms in (2.13) are due to non-linearities, and they cause a DC shift as well as distortion at frequencies $2\omega_1$, $3\omega_1$, and higher harmonics, which result in either gain compression or gain expansion. It can also be observed from equation (2.13) that distortion is present in any signal level.

To quantify the amount of harmonic distortion, the following definitions are used:

\[ HD_2 = \frac{\text{Amplitude}_{2\omega_1}}{\text{Amplitude}_{\omega_1}} \quad (2.14) \]

\[ HD_3 = \frac{\text{Amplitude}_{3\omega_1}}{\text{Amplitude}_{\omega_1}} \quad (2.15) \]

where (2.14) represents the fractional second harmonic distortion, and (2.15) represents the third fractional harmonic distortion. Higher order fractional harmonic distortion definitions can be written in the same fashion.

The above definitions for the second and third order fractional harmonic distortion can be written in terms of the coefficients of the power series and the input signal by examining (2.13).

\[ HD_2 = \frac{a_2}{2a_1} V_1 \quad (2.16) \]

\[ HD_3 = \frac{a_3}{4a_1} V_1^2 \quad (2.17) \]
Intermodulation

Harmonic distortion, introduced previously, is the result of non-linearities due to a single sinusoidal input. It is quite possible in practice that two or more sinusoidal signals are applied to the input of a two-port, for example, the first signal representing the input and others representing large in-band interferences at the input of an LNA. When this occurs, another non-linearity called intermodulation results. To see the effects of both harmonic distortion and intermodulation, assume that the input signal is now equal to

\[ V_i(t) = V_1 \cos(\omega_1 t) + V_2 \cos(\omega_2 t) \]  

(2.18)

The output can be expanded in a power series and (2.18) can be substituted into (2.11). The linear first term can be expressed as

\[ a_1 V_i = a_1 [V_1 \cos(\omega_1 t) + V_2 \cos(\omega_2 t)] \]  

(2.19)

The second term is given by

\[ a_2 V_i^2 = \frac{a_2 V_1^2}{2} [\cos(2\omega_1 t) + 1] + \frac{a_2 V_2^2}{2} [\cos(2\omega_2 t) + 1] + a_2 V_1 V_2 [\cos(\omega_1 + \omega_2) t + \cos(\omega_1 - \omega_2) t] \]  

(2.20)

Expanding the third term gives

\[ a_3 V_i^3 = \frac{a_3 V_1^3}{4} [\cos(3\omega_1 t) + 3 \cos(\omega_1 t)] + \frac{a_3 V_2^3}{4} [\cos(3\omega_2 t) + 3 \cos(\omega_2 t)] + \frac{3}{4} a_3 V_1 V_2^2 [2 \cos(\omega_1 t) + 2 \cos(\omega_2 t) t + \cos(\omega_1 - \omega_2) t] + \frac{3}{4} a_3 V_1^2 V_2 [2 \cos(\omega_2 t) + 2 \cos(\omega_1 t) t + \cos(\omega_1 - \omega_2) t] \]  

(2.21)
It can be observed from (2.20) and (2.21) that harmonic distortion terms are produced as if each sine wave is applied separately. However, second order intermodulation terms are also produced at \((\omega_1 + \omega_2)\) and \((\omega_1 - \omega_2)\), and third order intermodulation terms are produced at \((2\omega_1 \pm \omega_2)\), and \((2\omega_2 \pm \omega_1)\). Fig. 2.5 shows the distortion terms in the frequency spectrum [12].

![Diagram showing frequency locations of distortion terms](image)

**Figure 2.5** Frequency locations of distortion terms.

The following two equations define fractional intermodulation:

\[
IM_2 = \frac{Amplitude_{\omega_1,\omega_2}}{Amplitude_{\omega_1,\omega_2}}
\]

(2.22)

\[
IM_3 = \frac{Amplitude_{(2\omega_1 \pm \omega_2)}(2\omega_2 \pm \omega_1)}{Amplitude_{\omega_1,\omega_2}}
\]

(2.23)

Using the definitions in (2.22) and (2.23), the fractional intermodulation terms are

\[
IM_2 = \frac{a_2}{a_1} V_1
\]

(2.24)

\[
IM_3 = \frac{3a_3}{4a_1} V_1^2
\]

(2.25)
where it is assumed that $V_1 = V_2$. From Fig. 2.5, it is apparent that the 3rd-order intermodulation distortion $IM_3$ signals are close to the signal of interest $F$, which makes the filtering out of $IM_3$ signals difficult when recovering the signal of interest. Therefore minimizing intermodulation distortion is a key objective in many RF circuit designs.

### 1 dB Compression Point

It is mentioned previously that the 3rd-order term in the power series can either cause gain compression or gain expansion. If we assume that the sign between $a_1$ and $a_3$ are different, then gain compression occurs. $P_{1dB}$ is a measure of the power of the input signal such that it causes the 3rd-order non-linearity to reduce gain by 1 dB from the ideal value. It can be expressed as:

$$20 \log \left( 1 + \frac{3a_3}{4a_1} V_1^2 \right) = -1dB \tag{2.26}$$

Solving for $V_1$ in (2.26) gives:

$$P_{1dB} = \sqrt{\frac{4}{3}} \frac{a_1}{a_3} \sqrt{0.11} \tag{2.27}$$

### 3rd Order Intercept Point

Another measure for the 3rd-order non-linearity in a two-port network is the 3rd-order intercept point. Since the 3rd-order non-linearity is proportional to the input signal cubed, while the fundamental is increasing only linearly with the input signal, there is a point at which the amplitudes of the fundamental and that of the 3rd-order intermodulation product meet. The input signal at which this occurs is defined as the input-referred 3rd-
order intercept point (IIP3), and is equal to when $IM_3$ equals 1. Solving for $V_1$ using (2.25), the following equation for IIP3 is obtained.

$$IIP3 = \sqrt[4]{\frac{4}{3}} \left| \frac{a_1}{a_3} \right|$$  \hspace{1cm} (2.28)

**Stability**

A critical requirement of a two-port network is that it must not produce an output with oscillatory behavior. The stability of a two-port network can be determined from its S-parameters and the load and source impedances. The Rollet stability factor, $K$, is often used for verifying stability [16]. Unconditional stability is satisfied under the following two conditions:

$$K > 1$$  \hspace{1cm} (2.29)

$$\Delta < 1$$  \hspace{1cm} (2.30)

where

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2 |S_{12}S_{21}|}$$  \hspace{1cm} (2.31)

$$\Delta = S_{11}S_{22} - S_{12}S_{21}$$  \hspace{1cm} (2.32)

However, $K$ alone is usually good enough to test for stability since most transistors are either unconditionally stable, satisfying (2.29) and (2.30), or conditionally stable with $K < 1$ and $|\Delta| < 1$ [14].
Noise in Two-Port Systems

In order to design a circuit for low noise, it is useful to determine the condition under which noise can be minimized. This condition is then used in Chapter 3 to relate noise performance to CMOS design parameters. For the analysis of noise in two-port systems, consider a noisy two-port network driven by a noisy source as shown in Fig. 2.5(a).

The noise factor of a two-port network is defined as

\[ F = \frac{\text{SNR}_{\text{IN}}}{\text{SNR}_{\text{OUT}}} = \frac{P_{n,\text{output}}}{P_{n,\text{source}}} \] (2.33)

where \( P_{n,\text{output}} \) is the noise power outputted by the two-port and \( P_{n,\text{source}} \) is the noise presented at the input of the two-port. An ideal noiseless two-port network contributes no noise; hence the noise factor is equal to one. Noise figure NF, which is noise factor...
expressed in decibel, often used to specify noise performance and has an ideally value of 0dB.

To simplify analysis, the noise of a two-port network can be modeled as a noise voltage and a noise current at the input as shown in Fig. 2.5(b). The signal source is represented by a current source $i_s$ in parallel with and an admittance $Y_s$ to simplify derivation. In this case, the noise factor can be expressed as [13]:

$$ F = \frac{i_n^2 + |i_n + Y_s v_n|^2}{i_s^2} $$  \hspace{1cm} (2.34)

In the derivation of (2.34), the assumption has been made that the noise from the source is not correlated to the noise from the two-port. However, since the exact nature of the source of the two-port is not known, the above assumption about correlation may not be reasonable. Therefore, the following definition for $i_n$ is needed:

$$ i_n = i_c + i_u $$  \hspace{1cm} (2.35)

where $i_c$ is the portion of $i_n$ that is correlated with $v_n$, while $i_u$ is the part of $i_n$ that is uncorrelated with $v_n$. The current $i_c$ is equal to $Y_c v_n$, where $Y_c$ is known as the correlation admittance and is given by

$$ Y_c = \frac{i_c}{v_n} $$  \hspace{1cm} (2.36)
Equation (2.34) contains independent noise sources, each of which may be treated as thermal noise produced by an equivalent resistance or conductance:

\[ R_n = \frac{v_n^2}{4kT\Delta f} \]  
\[ G_u = \frac{i_u^2}{4kT\Delta f} \]  
\[ G_s = \frac{i_s^2}{4kT\Delta f} \]  

where \( k \) is Boltzmann's constant (about \( 1.38 \times 10^{-23} \text{ J/K} \)), \( T \) is the absolute temperature in kelvins, and \( \Delta f \) is the noise bandwidth in hertz. Using (2.36) – (2.39), the noise factor can be expressed purely in terms of impedances and admittances:

\[ F = 1 + \frac{G_u + \left(G_c + G_s\right)^2 + \left(B_c + B_s\right)^2}{G_s R_n} \]  

To optimize for noise in a circuit, the minimum noise factor can be solved for by first taking the derivative of (2.40) with respect to the source conductance and susceptance and setting them to zero. The results for the optimal source conductance and susceptance are stated below.

\[ B_{s,\text{opt}} = -B_c \]  
\[ G_{s,\text{opt}} = \sqrt{\frac{G_u + G_c^2}{R_n}} \]  

Substituting (2.41) and (2.42) into (2.40) gives the following results for the minimum noise factor.

\[ F_{\text{min}} = 1 + 2R_n \left(G_{s,\text{opt}} + G_c\right) \]
The noise factor can then be expressed in terms of $F_{\text{min}}$ and the source admittances by

$$F = F_{\text{min}} + \frac{R_n}{G_s} \left[ (G_s + G_{s,\text{opt}})^2 + (B_s + B_{s,\text{opt}})^2 \right]$$  \hspace{1cm} (2.44)

The above analysis shows that a source impedance optimized for a minimum noise factor exists, but this source impedance is often not the same as the impedance that achieves maximum power transfer. In (2.44), the ratio $R_n/G_s$ appears as a multiplier in front of the second term. For a fixed source conductance, $R_n$ represents the sensitivity of the noise factor as $G_s$ and $B_s$ departs from their optimal values. A large $R_n$ implies a high sensitivity, which obligates the design to stay close to optimal noise matching. As discussed subsequently, operation at low bias currents is associated with large $R_n$, mainly due to small device transconductance $g_m$. This is an example of the difficulty in achieving high performance at low power consumption.
Chapter 3
CMOS LNA Fundamentals

This chapter describes the theories and considerations useful to the implementation of an LNA in the CMOS technology.

Noise Sources in CMOS

Before beginning an analysis of how to design for low noise, the origins of the noise should be identified and understood. This section discusses several important noise sources in CMOS transistors.

Thermal Noise

Thermal noise is due to the random thermal motion of the carriers in the channel [17]. It is commonly referred to as a white noise source because its power spectral density holds a constant value up to very high frequencies (over 1 THz) [13]. Thermal noise can be
modeled as a current source across the drain and source of a transistor, as depicted in Fig. 3.1 [12], and has a power spectral density of

\[
\frac{i_d^2}{\Delta f} = 4kT\gamma g_{d0}
\]  

(3.1)

where \( k \) is Boltzmann’s constant, \( T \) is the absolute temperature, \( \gamma \) is a bias dependent process parameter, and \( g_{d0} \) is the zero-\( V_{DS} \) drain-source conductance, and has the following definition:

\[
\alpha = \frac{g_m}{g_{d0}}
\]  

(3.2)

where \( g_m \) is the transconductance and \( \alpha \) is typically in the range between 0 and 1.

Equation (3.1) is a general equation that can be used in both the linear and saturation regions of operation simply by using different values for \( \gamma \). For long channel transistors, \( \gamma = 2/3 \) in the saturation region, and \( \gamma = 1 \) in the linear region. For short channel transistors, hot carrier effects may cause \( \gamma \) to be as high as 2 or 3 [18].

**Induced Gate Noise**

Induced gate noise is a high frequency noise source that is caused by the non-quasi static effects influencing the power spectral density of the drain current [17]. Thermal noise in the channel couples through the oxide capacitance to the gate terminal, causing a gate noise current to flow. This noise source is normally not included in standard noise analysis because at low frequencies it is negligible. However, it can dominate at RF
frequencies [13]. Induced gate noise, with its circuit model shown in Fig. 3.2(a), has a power spectral density given by

\[
\frac{i_g^2}{\Delta f} = 4kT\delta \frac{\omega^2 C_{gs}^2}{5g_{d0}}
\]  

(3.3)

where \( \omega \) is the frequency, \( C_{gs} \) is the gate-source capacitance, and \( \delta \) is a process parameter equal to 4/3 in long channel devices [17]. Since the thermal channel noise and induced gate noise stem from the same physical phenomenon, it can be assumed that the relation \( \delta = 2\gamma \) continues to hold for short channel devices [13].

![Diagram of induced gate noise model](image)

(a) Frequency dependent, (b) Frequency independent.

The power spectral density of (3.3) is frequency dependent. An equivalent frequency independent noise model is to express the induced gate noise as a voltage in series with the gate capacitance, as shown in Fig. 3.2(b). If a high quality factor \( Q \) is assumed, then
the models shown in Fig. 3.2(a) and (b) are equivalent, and (3.4) gives the power spectral density.

\[ \frac{V_g^2}{\Delta f} = 4kT\delta r_g \]  

(3.4)

Equation (3.4) shows the interesting result that the gate noise is equal to the noise of \( r_g \), a resistor placed at the gate, scaled with the constant \( \delta \).

As discussed previously, gate noise and drain noise are partially correlated due to the fact that they are from the same source. The correlation coefficient \( c \) can be expressed as in (3.5):

\[ c = \frac{i_g i_d^*}{\sqrt{i_g^2 + i_d^2}} \]  

(3.5)

The theoretical value for \( c \) is \(-0.395\) for long channel devices [17]. Precise measurements of \( c \) are difficult to carry out, but the best published measurements reveal that its magnitude stays within a factor of 2 of this theoretical value, even for devices with drawn channel lengths as small as 0.13 \( \mu \text{m} \) [13].

**Distributed Gate Noise**

The distributed gate resistance of the CMOS transistor also contributes to the noise figure of an LNA. This noise source is modeled as a resistor at the gate and has a noise power spectral density equal to

\[ \frac{V_g^2}{\Delta f} = 4kTR_g \]  

(3.6)
where $R_g$ is the gate resistance, and is given by

$$R_g = \frac{R_{sq} W}{3n^2 L} \quad (3.7)$$

In (3.7), $R_{sq}$ is the sheet resistance of the gate material, $n$ is the number of fingers, and the factor $1/3$ results from the assumption that each finger is only contacted at only one end. If both ends are contacted, then the factor reduces to $1/12$.

**Intrinsic MOSFET Two-port Noise Parameters**

In chapter 2, the expression for the noise factor is derived, herein reproduced as (3.8) for convenience.

$$F = F_{\text{min}} + \frac{R_n}{G_s} \left[ (G_s + G_{s,\text{opt}})^2 + (B_s + B_{s,\text{opt}})^2 \right] \quad (3.8)$$

Viewing the MOSFET as a two-port network, with the gate and source forming a port and the drain and source forming another, it is useful to express $F_{\text{min}}, R_n, G_{s,\text{opt}},$ and $B_{s,\text{opt}}$ in terms of MOSFET device parameters [13]:

$$F_{\text{min}} \approx 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_r} \sqrt{\delta (1 - |c|^2)} \quad (3.9)$$

$$R_n = \frac{\gamma G_{d0}}{g_m} \quad (3.10)$$

$$G_{s,\text{opt}} = \alpha \omega C_{gs} \sqrt{\frac{\delta}{5\gamma}} (1 - |c|^2) \quad (3.11)$$

$$B_{s,\text{opt}} = -\omega C_{gs} \left( 1 - \alpha \frac{\sqrt{\delta}}{\sqrt{5\gamma}} \right) \quad (3.12)$$
This completes the noise analysis that relates the noise factor with design variables $g_m$, $\omega$, $C_{gs}$, and $g_{do}$. Examining (3.9) – (3.12), $W$ can also be related to the four noise parameters, which permits noise consideration in transistor sizing.

\[ F_{\text{min}} \text{ no width dependence} \]  
\[ R_n \propto 1/W \]  
\[ G_{sp} \propto W \]  
\[ B_{sp} \propto W \]  

Aside from the classical noise matching (CNM) presented above, a number of CMOS LNA design optimization techniques are also well established such as simultaneous noise and input matching (SNIM), power-constrained simultaneous noise and input matching (PCSNIM), and power-constrained noise optimization (PCNO). A good overview of these techniques is presented in [5].

**Inductive Source Degeneration**

An LNA must provide an input matching to a typically 50$\Omega$ element such as a band-select filter or an antenna. In a fully integrated receiver, LNA output matching is often not required as it is connected to the next on-chip stage in the receive chain. To minimize the number of off-chip components, an LNA should implement the elements required for input matching on chip. One popular approach is to use inductive degeneration.
**Input Impedance Match**

Input impedance matching by inductive source degeneration is popular as matching to the signal source does not introduce additional noise (as in the case of using a shunt input resistor) and does not restrict the value of \( g_m \) (as in the case of the common-gate configuration).

![Inductive source degeneration circuit](image)

**Figure 3.3** Inductive source degeneration.

The circuit shown in Fig. 3.3 has an input impedance equal to

\[
Z_{in}(\omega) = \frac{1}{j\omega C_{gs}} + \frac{g_m L_s}{C_{gs}}
\]

(3.17)

where ideal inductors and capacitors have been assumed. From (3.17), in order to achieve an input impedance match, the following condition must be satisfied:

\[
R_s = \frac{g_m L_s}{C_{gs}} \approx \omega T L_s
\]

(3.18)

where \( \omega T \approx g_m/C_{gs} \) is the transit frequency of the transistor. Once \( L_s \) is chosen based on gain, linearity and input matching requirements, \( L_g \) can then be chosen such that \( L_g, L_s, \) and \( C_{gs} \) resonate at \( \omega_0 \). In other words, the following condition for \( L_g \) must hold:

\[
\omega_0 = \frac{1}{\sqrt{(L_g + L_s)C_{gs}}}
\]

(3.19)
Circuit Transconductance

Transconductance is important for gain. To find the transconductance $G_m$ of the circuit shown in Fig. 3.3, first note that the input matching network forms a series RLC tank. The $Q$ of the tank is

$$Q_{in} = \frac{1}{\omega_0 \left( R_s + \frac{g_m L_s}{C_{gs}} \right) C_{gs}} \quad (3.20)$$

where $\omega_0$ is the resonant frequency defined in (3.19). At resonance, the voltage across the capacitor is equal to

$$v_{gs} = Q_{in} v_s \quad (3.21)$$

and the short circuit output current is equal to

$$i_{out} = g_m v_{gs} \quad (3.22)$$

where $g_m$ is the transconductance of the device. Using (3.20) – (3.22), the overall circuit transconductance can be solved for, and is given by the following equations:

$$G_m = \frac{g_m}{\omega_0 \left( R_s + \frac{g_m L_s}{C_{gs}} \right) C_{gs}} \quad (3.23)$$

$$\approx \left( \frac{\omega_T}{\omega_0} \right) \frac{1}{\left( R_s + \omega_T L_s \right)} \quad (3.24)$$

It can be observed from (3.24) that $G_m$ is dependent on CMOS process technology through the transit frequency.
Circuit Topologies

The key specifications for characterizing the performance of an integrated LNA are gain, noise, linearity, power consumption, stability, and input matching. These specifications depend on the circuit topology. Topologies such as common-source, common-gate, cascode, and distributed amplifiers have been used for different performance requirements. Compared to the common-source configuration, common-gate is more suitable for wide-band operation, but suffers from relatively high NF [19]. Distributed amplifiers are also capable of wide-band operation, but they suffer from relatively high power consumption [20].

For narrow-band operation, which is the focus of this work, the common-source and cascode amplifiers are the most suitable. In the common-source configuration, as shown in Fig. 3.4(a), the signal is applied to the gate and the output is taken from the drain. The cascode amplifier is a common-gate amplifier stacked on top of a common-source
amplifier, as shown in Fig. 3.4(b). The cascode amplifier consists of an input transistor $M_1$ and a cascode transistor $M_2$ with a gate bias voltage $V_B$. Compared to the cascode amplifier, the common-source amplifier suffers from the following shortcomings:

1) Poor isolation between input and output, due to the gate-to-drain parasitic capacitance $C_{gd}$, increases the chance of instability significantly.

2) As CMOS technology scales, the MOSFET output resistance $r_o$ decreases, causing noticeable performance degradation. For a common-source amplifier, $r_o$ appears in parallel with the load impedance in small-signal operation, which reduces the output impedance, and lowers the gain of the LNA. A possible solution is to increase the gate length $L$, which results in a degradation of NF.

To alleviate the shortcomings of the common-source topology, the cascode topology is often used. The addition of the cascode device reduces the effect of the $C_{gd}$ of $M_1$ by presenting a low impedance node at the drain of $M_1$, improving stability. The cascode device also performs impedance transformation so that the output impedance of the amplifier is improved by a factor approximately equal to the intrinsic gain of the cascode device.

The load inductor, $L_d$, is designed to resonate at the operating frequency with the output node capacitance. The input and output tanks can be aligned to provide a narrowband gain, but can also be offset from each other to provide a broader and flatter frequency
response. One shortcoming of the cascode topology is that the extra transistor consumes voltage headroom. As a result, the load should not consume a large voltage headroom. An inductive load $L_d$, as opposed to a resistive load, is preferred. An inductive load has the added benefit of increasing the gain by resonating with the capacitances associated with the output node and also improving frequency selectivity.

Since on-chip inductors have a limited range of values, the capacitor $C_m$ of Fig. 3.4(b) can be added between the gate and source terminals of $M_1$ so that the values of $L_g$ and $L_s$ are reduced to permit on-chip implementation. Reducing the inductor value also improves $Q$, which reduces input losses and improves LNA noise figure. Adding $C_m$ has another advantage of providing an extra degree of freedom for choosing the gate width $W$ of $M_1$, which decouples impedance matching requirements from power consumption. A drawback of adding extra gate-to-source capacitance is the degradation of $f_T$ of $M_1$. However, as discussed subsequently, at operating frequencies well below $f_T$, this trade-off is reasonable. It is also interesting to note that adding $C_m$ does not degrade distortion performance [10].

**Performance of the Cascode Amplifier**

Since the cascode amplifier depicted in Fig. 3.4(a) is a robust topology for narrow-band applications, a detailed study of its gain, noise, and linearity performance is provided in this section.
**Voltage and Power Gain**

Assuming the LNA is input matched, the voltage gain and power gain can be expressed as (3.25) and (3.26), respectively [17].

\[
A_v = \left( \frac{\omega_T}{\omega_0} \right) \frac{R_L}{2R_s} \tag{3.25}
\]

\[
G_T = \frac{P_L}{P_{_{av}}} = \left( \frac{\omega_T}{\omega_0} \right)^2 \frac{R_L}{R_s} \tag{3.26}
\]

where \( R_L \) is the resistance at the output due to the finite \( Q \) of \( L_d \) and the finite output resistance of the transistor, and \( R_s \) is the impedance of the signal source.

**Noise**

The noise factor \( F \) of the LNA can be expressed as [1]

\[
F = 1 + \frac{R_s}{R_s} + \gamma \chi \beta_m R_s \left( \frac{\omega_0}{\omega_T} \right)^2 \tag{3.27}
\]

\( F \) can be expressed in a more intuitive form if power/impedance matching is assumed at the input (which is often the case). The derivation is done in [21], and the result is repeated below:

\[
F \approx 1 + \frac{\gamma}{\alpha} \left( \frac{\omega_0}{\omega_T} \right)^2 \left( g_m R_s + \frac{2}{5} \right) + \frac{\alpha \delta (1 - |c|^2)}{5 g_m R_s} \tag{3.28}
\]

where the portion of the gate noise that is correlated with the drain noise has been neglected. In (3.28), the second term is the contribution from the channel drain noise, and the last term is the contribution from the uncorrelated portion of the gate noise. It can be
observed that as $\omega_T$ increases (for example with improving technology), the drain noise contribution becomes less significant if the operating frequency is kept constant.

**Linearity**

A purely sinusoidal input signal can produce a distorted output signal with higher-order harmonics due to the nonlinearity of the MOSFET. These harmonics are mainly induced by higher-order derivatives of the current-voltage ($I_D-V_{GS}$) characteristics. An important figure of merit for linearity is the 3rd-order harmonic intercept voltage VIP3, which is the extrapolated input voltage amplitude at which the 1st- and 3rd-order output amplitudes are equal. The input-referred VIP3 can be expressed as [9]:

$$V_{ip3} = \sqrt{\frac{24}{g_m}}$$  \hspace{1cm} (3.29)

For linearity, VIP3 is used as a key design parameter and its value is representative of the input signal amplitude that the system can process with reasonably low distortion. It can be obtained by taking the third derivative of the $I_D-V_{GS}$ characteristics in respect to $V_{GS}$.

Figure 3.5 gives the VIP3 as a function of gate bias $V_{GS}$ under different values of drain bias $V_{DS}$ [22]. A sharp peak is observed near the threshold voltage as gate bias is varied, which reflects the so called “sweet spot” of gate biases for high MOSFET linearity [9]. This is because, during the transition from subthreshold to strong inversion, the increase of $g_m$ with $V_{GS}$ is at its highest, and the 2nd-order nonlinearity coefficient $g_{m2}$ reaches its peak, while the 3rd-order nonlinearity coefficient $g_{m3}$ becomes zero.
At the system level, the input 3\textsuperscript{rd}-order intermodulation point (IIP3) of the cascode amplifier can be written as [22]

\[
IIP3[dbm] = IIP3_{in}[dBm] - 20 \log_{10} \frac{1}{\omega_0 C_{gs} R_j} 
\]  

(3.30)

The first term in (3.30) is the intrinsic IIP3 of the device, and arises from the fact that short channel CMOS transistors exhibit velocity saturation, which gradually linearizes the ideal quadratic relationship of the long channel drain current equation. The second term results from the extra voltage boost across the $C_{gs}$ due to the series resonance tank, which increases gain but degrades IIP3. This outlines a tradeoff between gain and linearity.
Inductor Design

Inductors are widely used in RF circuitry to resonate with capacitors and to provide impedance transformations. The design of on-chip inductors is important as the inductor may dominate the frequency selectivity of a RF circuit and consumes a large area. The frequency selectivity of an inductor, characterized by the quality factor $Q$, is a major design parameter to be optimized. As evident in the following subsections, there is a tradeoff between $Q$, practical inductance values, and inductor area.

Physical Dimensions

On-chip inductors can be implemented in different shapes such as squares, octagons, and circles. The layout of a square spiral inductor is depicted in Fig. 3.6.

Figure 3.6  The layout of a square spiral inductor.
Although it may seem counterintuitive, the $Q$ depends very little on the shape [13]. Instead, it mainly depends on the following parameters:

\[ \begin{align*}
N & : \text{The number of turns in the spiral} \\
D & : \text{The inductor's inner radius (For a square spiral, D is the shortest distance between the center and the inner side of the spiral)} \\
W & : \text{The width of the wire} \\
S & : \text{The spacing between two wires} \\
T & : \text{The thickness of the wire. Once the metal layer is chosen, T is fixed.}
\end{align*} \]

**Inductor Figures of Merit**

In this section, three common figures of merit used to describe the characteristics of an inductor are discussed: inductance, quality factor, and self-resonant frequency.

**A. Inductance**

The calculation of the inductance of a structure is based upon the self-inductance of a conductor, and the mutual inductance between two conductors [23]. The total inductance of a spiral structure is equal to the sum of all of the self-inductances of the wire segments and the positive and negative mutual inductances between the wire segments. Two segments have a positive mutual inductance if the direction of current flow in them is the same and a negative mutual inductance if the direction of current flow in them opposes each other. Inductance calculation is often complicated and is best performed by a computer simulation program such as ASITIC [24].
B. Quality Factor

The quality factor $Q$ is a measure of the amount of energy loss in a circuit component or network and has implications on their frequency selectivity. It is defined as

$$Q = 2\pi \frac{\text{Energy}_{\text{stored}}}{\text{Energy}_{\text{dissipated}}} = \frac{\omega_0}{BW}$$

(3.31)

where $\omega_0$ and $BW$ are the resonant frequency and bandwidth, respectively, of the component or network. An ideal inductor has a $Q$ of infinity. When the loss is significant (caused by for example interconnect resistance, substrate loss, and the skin effect), the peaking of a signal at resonance degrades, which in turn lowers the amplifier gain. The degradation of the frequency selectivity, which results in a large $BW$ may be undesirable as, for example, an LNA should ideally reject out-of-band signals while only amplifying the signal of interest. The quality factor can be improved by adding a patterned ground shield between the inductor and the substrate to reduce capacitive coupling [25].

C. Self-Resonant Frequency

The self-resonant frequency is the frequency at which the inductor resonates with its own parasitic capacitance. Below the self-resonant frequency, the inductor behaves like an inductor, at the self-resonant frequency the inductor behaves like a resistor, and above the self-resonant frequency, the inductor behaves like a capacitor. In general, a physically larger inductor tends to have a lower self-resonant frequency [13]. Therefore, the requirement that the inductor operates at most at half its self-resonant frequency places a limit to the size of the inductor.
Inductor Modeling

Inductors can be modeled with an electromagnetic field solver tool such as ASITIC. A frequency-dependent \( \pi \) model as shown in Fig. 3.7 can be used for narrow-band applications. \( R \) models the resistance of the interconnect, \( L \) models the inductance, \( C_s \) and \( R_s \) respectively model the capacitance and resistance from the inductor metal to the substrate.

![Pi-model of inductor](image)

Figure 3.7 Pi-model of inductor.
Chapter 4

Designing for Power-Efficient Operation

The main objective of this work is to create a design methodology for power-efficient LNAs that can operate at sub-mW power consumption levels. This chapter describes device biasing and sizing, the two fundamental design steps, and presents a step-by-step design procedure.

The following discussion is illustrated with a collection of graphs that relate device characteristics such as transconductance and power consumption to circuit design parameters such as gate voltage $V_{GS}$ and gate width $W$. These graphs offer perspectives to explore the design space particularly useful for selecting the biasing condition for the amplifier, which is the first step of the proposed design procedure. Data from the graphs has been obtained from SpectreRF simulations using a commercial 90nm process design kit.

Device Biasing

Since circuit performance is strongly tied to device biasing and that proper biasing involves knowledge of device characteristics, it is important to incorporate the knowledge of the device from the beginning of the design cycle to minimize parameter tuning at the end.
Terminal $I-V$ Characteristics

Figure 4.1  $I_D$ vs. $V_{GS}$ for a 90nm nFET ($V_{DS} = 1V$).

Figure 4.1 shows the drain current of a 90nm nMOSFET as a function of its gate voltage for three transistor sizes, specified in units of micrometers. In this process, the threshold voltage $V_{TH}$ is around 0.35V. All results thereafter are based on such an nMOSFET. It is apparent that deep submicron effects such as velocity saturation occur at $V_{GS} > 0.6V$, rendering the $I_D-V_{GS}$ curve more linear as opposed to quadratic. This suggests that linearity of a circuit can be improved with a strong gate bias voltage. Also evident from Fig. 4.1 is that $I_D$ is proportional to the transistor width $W$ for a given length $L$. However, according to the square law, transistor sizes 20/0.1 and 40/0.2 should have the same current, which is not the case for this 90nm nFET. With the aid of this graph, the designer can evaluate the extent of the deviation of the $I-V$ relationship from the classical model.
Figure 4.2  $I_D$ vs. $V_{DS}$ for a 90nm nFET ($W/L = 20 \mu m/0.1\mu m$, $V_{DS} = 1V$).

The ability for a transistor to function as a robust current source is crucial to amplifier design. Fig. 4.2 shows the drain current characteristics of a 90nm nFET. For a given $V_{GS}$, $I_D$ is plotted as a function of $V_{DS}$. This plot explicitly reveals channel length modulation, evident by the dependency of $I_D$ on $V_{DS}$ when the transistor operates in saturation. Channel length modulation manifests itself as non-linearity and gain reduction at the circuit level.

If the power consumption of a MOSFET is indicated by the product of $I_D$ and $V_{DS}$, then a set of constant power contours, which is also depicted in Fig. 4.2, can be plotted in the $I_D$-$V_{DS}$ design space. The plot shows three power levels at 0.1, 0.5, and 1mW, with the lower left corner representing a region of lower power consumption. The intersection of power contours and $I_D$-$V_{DS}$ curves provides different combinations of $V_{GS}$, $V_{DS}$, and $I_D$, which in turn facilitates the choice of power-constrained bias selection.
The cut-off frequency $f_T$, also called transit frequency, has been widely used as a measure of operating frequency of the device. It is defined as the frequency at which the current gain of the device is equal to unity and is given by

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \quad (4.1)$$

Figure 4.3 illustrates $f_T$ as a function of $V_{GS}$ for different transistor sizes. The curves for transistor sizes 20/0.1 and 40/0.1 overlap, indicating that $f_T$ for a fixed $L$ is not a function of $W$. It can be inferred from (4.1) that the increase in transconductance $g_m$ that results from increasing $W$ is offset by the increase in parasitic capacitances. Also evident from Fig. 4.3 is that $f_T$ is relatively low at a low $V_{GS}$, which is necessary for power-efficient operation, and since a high $f_T$ is necessary for a low noise figure, a tradeoff is needed between noise and power. Also, $f_T$ is degraded when the device operates in the subthreshold region and when non-minimum $L$ is used.
The transconductance $g_m$ of a device is important to amplifier design as the gain of an amplifier is the product of its transconductance and output resistance. Fig. 4.4 depicts $g_m$ as a function of $V_{GS}$. $g_m$ is obtained by differentiating the DC drain current $I_D$ with respect to $V_{GS}$: $g_m$ increases rapidly with $V_{GS}$ until it saturates at a $V_{GS}$ well above $V_{TH}$.

To relate device performance with power consumption, it is useful to define transconductance efficiency $g_m/I_D$ [7]. Fig. 4.5 shows $g_m/I_D$ as a function of $V_{GS}$.
As can be seen, $g_m/I_D$ is, to the first order, invariant across transistor sizes, indicated by the overlapping of the curves for the $L=0.1\mu m$ cases and that $g_m/I_D$ is insensitive to $W$ when the device is turned on. The fact that $g_m/I_D$ is independent from transistor size is significant as this removes transistor size from the power efficiency optimization equation, leaving $V_{GS}$ as the primary variable to be considered. This means that circuit design that optimizes transconductance efficiency can be broken down into two sequential steps: first determining bias condition for maximum efficiency, then sizing transistors based on the absolute power requirement. As shown in Fig. 4.5, to exploit the high $g_m/I_D$, it is ideal to bias the transistor in the subthreshold region. However, as previously shown in Fig. 4.3, $f_T$ in this region may be insufficient. A good compromise is to operate in moderate inversion.
As mentioned previously, an amplifier’s gain is a function of the amplifier’s output resistance, which in turn is a function of the device’s output resistance $r_o$. $r_o$ can be expressed as $1/g_{ds}$, where $g_{ds}$ is the drain-to-source conductance. If $r_o$ is small, the fact that it appears in parallel with the amplifier’s load can significantly reduce the output resistance of the amplifier, hence degrading its gain. For the older technologies, $r_o$ has been high enough that it can be ignored. As devices move to deep submicron, $r_o$ decreases due to channel length modulation, with a small value in the order of hundreds of ohms for practical choices of $V_{GS}$. As can be seen from Fig. 4.6, $r_o$ is a strong function of $V_{GS}$, $W$, and $L$. Also, a large $W$, often required for noise matching at the input stage, degrades $r_o$ substantially. This posts a challenge for LNA design and is often overcome by an increase in power consumption.
To evaluate the ability of a device to provide gain, the intrinsic gain $g_m r_o (= g_m/g_{ds})$ is plotted as a function of $V_{GS}$ in Fig. 4.7. The intrinsic gain represents the theoretical maximum gain achievable by a single transistor. Curves for $L=0.1\mu$m overlap each other. As shown in Fig. 4.7, $g_m/g_{ds}$ improves significantly as $L$ increases. However, using non-minimum $L$ is often not a good practice in designing LNAs as it degrades the noise performance. This outlines a tradeoff between gain and noise performance. If gain is compromised to achieve low noise, then multiple gain stages may be required. This adds power consumption into the gain-noise tradeoff.

**Transistor Sizing**

The biasing condition, which corresponds to a particular $V_{GS}$, determines the drain current density $I_D/W$. Once $I_D/W$ is found, transistor sizing can be performed based on the current density and the drain current allowed by the power specification. Since the objective is to
design for low noise, the dependency of noise on $W$ is examined. Fig. 4.8 shows the NF of a cascode and a common-source LNA for different gate widths.

As can be seen, NF decreases as $W$ increase, which suggests a tradeoff between NF and power consumption. Also, NF is significant for small $W$, indicating that low noise is fundamentally difficult to achieve for low-power circuits.

Figure 4.8 reveals an interesting limitation of conventional power-constrained noise optimization methods proposed to target low-power design. This technique is based on first finding an optimal gate width, then biasing the device with the amount of drain current allowed by the power constraint [13]. In sub-mW designs, the large optimal gate width combined with a small drain current often force the device to be biased in the subthreshold region. When subthreshold operation is inadequate, such as insufficient
frequency response, the technique is no longer applicable. An effective noise optimization technique for the ultra-low-power design space is needed.

**Step-by-step LNA Design Methodology**

A step-by-step design methodology for power-efficient inductively degenerated common-source or cascode LNAs is proposed. The cascode LNA is depicted in Fig. 4.9. The key difference between the proposed methodology and the conventional ones [1] [5] is that it starts from device biasing instead of device noise characteristics. Beginning the design procedure with device biasing has the advantage that all of gain, noise, linearity, and power are taken into consideration at the start of the design, instead of optimizing for noise while later possibly resorting to compromising other aspects severely. Also, it is important to note that the primary objective of the proposed design methodology is to reduce power consumption. The performance of the LNA, especially noise, is inevitably suboptimal.

![Figure 4.9 Cascode amplifier. (a) Schematic, (b) Simplified small-signal model for input matching analysis.](image-url)
Step1: Choosing the bias $V_{GS}$

Since power consumption is the focus of this design methodology, the first step exploits the notion of transconductance efficiency $g_m/I_D$. As $V_{GS}$ increases, $g_m/I_D$ reduces but $g_m$ (and also $f_T$) improves. This suggests that there is an optimal value of $V_{GS}$ for a given application. Ideally, $V_{GS}$ should be chosen to be a low value to maximize $g_m/I_D$, which leads to a power-efficient circuit. But the lower bound of $V_{GS}$ is governed by designing for sufficient $g_m$, which translates proportionally to amplifier gain, and sufficient $f_T$, which provides enough bandwidth for the amplifier operating frequency.

$V_{GS}$ can also determine noise performance in three ways. First, since biasing for higher $f_T$ leads to lower device NF, $V_{GS}$ should ideally be large. Second, there exists a characteristic current density of 0.15mA/µm that yields minimum device NF [6]. Having a lower device NF in turn lowers the overall amplifier NF. Since the characteristic current density corresponds to strong inversion, $V_{GS}$ should ideally be large. Third, as often seen in sub-mW designs, the width of the input transistor is often made small to satisfy the power requirement. But this small $W$ is often not optimal for noise matching. By lowering $V_{GS}$, although reducing $f_T$ and deviating from the characteristic current density, $W$ can be made larger to further approach an optimal noise match. Therefore, an optimal $V_{GS}$ exists and its selection is nontrivial.
$V_{GS}$ can also determine linearity performance. The fact that the device exhibits superb VIP3 linearity performance when biased in moderate inversion gives the designer more incentive to choose $V_{GS}$ from a narrow range that corresponds to moderate inversion.

Since $g_m$ and $f_T$ fall dramatically as $V_{GS}$ enters the subthreshold values, having $V_{GS}$ slightly above $V_{TH}$ is often a good choice for RF operations (frequency roughly below 10GHz). For operation at a higher frequency, a higher $V_{GS}$ is often needed to achieve an $f_T$ close to the maximum achievable by the technology.

As can be seen from the above discussion, all of gain, noise, and linearity are simultaneously affected by biasing, reflecting the interdependent nature of analog circuit design.

**Step 2: Calculate $I_D$**

Calculate the drain current $I_D$ from the target power consumption (excluding biasing circuits) $P_{DC}$ and target supply voltage $V_{DD}$, namely $I_D = P_{DC} / V_{DD}$.

**Step 3: Transistor Sizing**

The widths of the transistors $W$ can then be readily calculated from the drain current $I_D$ obtained in Step 2 with the aid of (4.2), the expression for $I_D$ of a short-channel device, where $v_{sat} \approx 10^7 \text{ cm/s}$ is the saturation velocity and $E_c$ is the critical field ($E_c \approx 6 \times 10^4 \text{ V/cm}$ for electrons and $24 \times 10^4 \text{ V/cm}$ for holes). In LNA design, non-minimum $L$ is rarely chosen as a small $L$ is critical for providing a low NF and gain at RF.
\[ I_D = W_{sat} C_{ox} \frac{(V_{GS} - V_{TH})^2}{(V_{GS} - V_{TH}) + E_s L} (1 + \lambda V_{DS}) \] (4.2)

**Step 4: Determine \( g_m \)**

Transconductance \( g_m \) can be obtained by taking the partial derivative of \( I_D \) with respect to \( V_{GS} \) for (4.2) or by running a DC simulation.

**Step 5: Determine Gate Capacitance \( C_{gs} \)**

Decide whether or not additional gate-to-source capacitance \( C_m \) is beneficial. If the circuit is designed to operate at high frequency, \( C_{gs} \) should be minimized to improve \( f_T \). If the circuit is designed for operation at relatively low frequency, the lower resonance frequency of the circuit requires larger combined inductance and capacitance. Since large on-chip inductors cost significant area and cannot be made with high \( Q \) in current standard technologies, it is easier to add capacitance.

**Step 6: Impedance matching**

Inductive degeneration is used for input matching. Fig. 4.9(b) is a simplified small-signal model showing the components for matching, where \( C_{gs1} \) denotes the gate-source capacitance of \( M_1 \). The input impedance of the LNA \( Z_{in} \) can be expressed as

\[ Z_{in} = \frac{1}{j \omega C_{gs}} + j \omega (L_s + L_g) + \frac{g_m L_s}{C_{gs}} \] (4.3)
where $g_m$ is the small-signal transconductance, $C_{gs} = C_m \parallel C_{gs1}$ is the effective capacitance between the gate and source and $\omega$ is the operating frequency. Since $Z_{in}$ is to be matched
to the source impedance, which is typically 50Ω in an RF system, the real and imagery parts of Z_{in} can be expressed as follows:

\[
\text{Re}\{Z_{in}\} = \frac{g_m L_s}{C_{gs}} = R_s = 50\Omega
\] (4.4)

\[
\text{Im}\{Z_{in}\} = \frac{1}{\omega C_{gs}} + \omega (L_s + L_g) = 0
\] (4.5)

Examining (4.4), given \( g_m \) is determined by \( V_{GS} \) and \( W, L_s \) and \( C_{gs} \) can be designed. It is better to design \( L_s \) first as the gain and linearity of the amplifier is dependent on it. A larger \( L_s \) adds more source degeneration and reduces the gain but improves the linearity [10]. Another practical reason is that inductors that are suitable for on-chip implementation have a smaller range of values. Since \( g_m \) is known from Step 4, once \( L_s \) is found, \( C_{gs} \) can be readily calculated. Then \( L_g \) can be calculated from (4.5).

**Step 7: Designing the Load \( L_d \) and \( C_{tune} \)**

\( L_d \), \( C_{tune} \), and the parasitic capacitance at the drain of the cascode device should resonate at the frequency of operation. \( C_{tune} \) is often implemented using a bank of capacitors to provide variable capacitance for channel tuning and calibration for process variation. \( L_d \) is often chosen to be as large as possible for on-chip implementation as a large \( L_d \) improves gain.

Step 7 concludes the design methodology. The performance and yield of the design may be further optimized by using CAD tools such as a design optimizer or a yield optimizer.
To demonstrate the application of the design methodology described in Chapter 4, an LNA is designed and simulated in a commercial 90nm CMOS technology to operate at the 2.4GHz band. This chapter presents the design rationale and the corresponding simulation results.

Circuit Topology and Impedance Matching

To alleviate the shortcomings of the common-source topology as discussed in Chapter 3, the cascode topology as discussed previously is used, hereby reproduced as Fig. 5.1 for convenience.

![Figure 5.1 Schematic of a cascode amplifier.](image)
The cascode amplifier consists of an input transistor $M_1$ and a cascode transistor $M_2$ with a gate bias voltage $V_B$. Since the cascode amplifier consists of two stacked transistors, the load should not consume a large voltage headroom. An inductive load $L_d$, as opposed to a resistive load, is preferred. An inductive load has the added benefit of boosting the gain by resonating with the capacitances associated with the output node. Inductive degeneration is used for input matching.

### Power-Efficient LNA Design

The biasing of the transistors has strong implications on LNA performance such as gain, noise, and linearity. When a short-channel MOSFET is biased in saturation, $I_D$ can be expressed by (4.2), herein reproduced as (5.1) below:

$$I_D = W \nu_{sat} C_{ox} \frac{(V_{GS} - V_{TH})^2}{(V_{GS} - V_{TH}) + E_c L} (1 + \lambda V_{DS})$$

(5.1)

where $\nu_{sat} \approx 10^7 \text{cm/s}$ is the saturation velocity and $E_c$ is the critical field ($E_c \approx 6 \times 10^4 \text{ V/cm}$ for electrons and $24 \times 10^4 \text{ V/cm}$ for holes). As shown by (5.1), when $L$ is kept to its minimum, $V_{GS}$ and $W$ are key design parameters that directly link to power consumption.

The drain current of a 90nm cascode LNA is plotted in Fig. 5.2 to quantify the sensitivity of power consumption to $V_{GS}$ and $W$ (supply voltage = 1V). Fig. 5.2(a) depicts the drain current of a cascode LNA with both transistors sized to 25µm. The threshold voltage of the technology used is around 0.4V. In the typical analog design space for this technology (i.e., $V_{GS}$ in the range of 0.4V to 0.7V), power consumption increases 6.4x whereas, in Fig. 5.2(b) a change of $W$ from 10µm to 40µm ($V_{GS}$ held constant at 0.4V) leads to a power increase of 4.2x. It is interesting to note that, given a drain-to-source voltage, the
performance of the MOSFET is strongly tied to the drain current density $I_D/W$, which is mainly controlled by $V_{GS}$. The following subsections describe the sensitivity of gain, noise, and linearity to $V_{GS}$ and $W$.

Figure 5.2 $I_D$ of a cascode LNA. (a) $I_D$ vs. $V_{GS}$ ($W_1=W_2=25\mu m$), (b) $I_D$ vs. $W$ ($V_{GS}=0.4V$).

Gain

In conventional RF and microwave design approaches [26], [27], primarily developed for high-speed bipolar circuits, $f_T$ is an indispensable design parameter. For CMOS technologies, there is a characteristic current density associated with operating in strong inversion that yields an optimal $f_T$ [6]. For the 90nm CMOS technology used in this work, $V_{GS} \approx 0.7V$ is required to reach the optimal value.
Figure 5.3 Voltage gain of a cascode LNA. (a) $A_v$ vs. $V_{GS}$, (b) $A_v$ vs. $W$.

Figure 5.3(a) shows the simulated voltage gain $A_v$ of the cascode amplifier depicted in Fig. 5.1 at 2.4GHz. While sweeping $V_{GS}$, $V_B$ is also modified such that the ratio $V_{GS1}/V_{GS2}$, hence $g_{m1}/g_{m2}$, is relatively constant. It may seem counterintuitive that $A_v$ only increases slightly as $V_{GS}$ changes from 0.4V to 0.7V, which corresponds to roughly doubling $f_T$ and $g_{m1}$. The reason lies in the fact that the operating frequency is sufficiently low compared to the maximum $f_T$ of the technology and therefore the benefits of increasing $f_T$ are not as pronounced. This suggests that the use of $f_T$ as a design tool may have less influence in modern RF design. In addition, as $V_{GS}$ increases, $I_D$ also increases, which reduces the overall output impedance of the cascode structure. Since gain is the product of output impedance and transconductance, the reduction of output impedance partially counteracts the effects of increasing $g_{m1}$ on the overall gain. For comparison, Fig. 5.3(b) shows moderate increase in $A_v$ as the widths of both transistors are increased.
Noise

Figure 5.4 NF of a cascode LNA. (a) NF vs. $V_{GS}$, and (b) NF vs. $W$.

As discussed in Chapter 4, there is currently no robust noise optimization technique specifically developed for the sub-mW regime. In the absence of such a technique, an alternative design approach is to use the knowledge gained in Chapter 4 to establish an initial biasing point and determine the sensitivity of NF with respect to $V_{GS}$ and $W$. For this purpose, a 90nm cascode LNA is simulated. Fig. 5.4(a) shows NF as $V_{GS}$ of $M_1$ is
swept from 0.3V to 0.7V ($W = 25\mu m$). $V_B$ is adjusted accordingly as mentioned earlier. As can be seen from Fig. 5.4(a), NF is inadequate when operating in the subthreshold region ($V_{GS} = 0.3V$). For a change of $V_{GS}$ from 0.4V to 0.7V, NF is reduced by 0.6dB at the expense of a 6.4× increase in the power consumption. Fig. 5.4(b) shows NF as the width of both transistors are changed simultaneously from 10μm to 40μm ($V_{GS} = 0.4V$), which results in a 3.4dB NF improvement at the expense of 4.2× the power consumption. This suggests that increasing $W$ is a more effective method for reducing NF.

It is also instructive to use the SpectreRF simulator to obtain a noise summary that shows the noise contribution of components. Fig. 5.5 depicts the RF signal source and the LNA with the inductor models explicitly shown. The noise contributions from inductor parasitic resistances are significant, suggesting high-$Q$ inductors are highly desirable.

Figure 5.5 Noise contribution of the signal source and LNA components.
Linearity

As discussed in Chapter 2, there are two types of linearity performance for an LNA, which are characterized by the input-referred $3^{rd}$-order intermodulation intercept point (IIP3) and the input-referred 1dB voltage compression point (P1dB).

Figure 5.6  \textit{IIP3 vs. $V_{GS}$} of the cascode LNA.

Figure 5.7  \textit{P1dB vs. $V_{GS}$} of the cascode LNA.

Figure 5.6 and Fig. 5.7, respectively, show the IIP3 and P1dB of the cascode LNA with 25µm transistors. Since linearity is a function of gain, Fig. 5.6 and Fig. 5.7 also show the
linearity of the LNA with gate widths adjusted to maintain a constant gain across $V_{GS}$. $V_B$ is adjusted accordingly as mentioned earlier. Two observations can be made. First, IIP3 of the LNA is indeed the highest in moderate inversion, suggesting that the VIP3 peak can be exploited in amplifier design. Second, the IIP3 and P1dB of the LNA degrade as $V_{GS}$ increases. This is an interesting observation since previous analysis and measurements reveal that the MOSFET IIP3 performance improves as $V_{GS}$ increases [9], [10]. Lastly, it has been reported that linearity of a MOSFET in moderate inversion is not well studied [9] and that the IIP3 peaking for a CMOS short-channel transistor may not be easily applicable for RF LNA designs [10]. This is due to the fact that source degeneration tends to improve the overall linearity but dampen the peak.

Simulation Results

The cascode LNA has been simulated using the BSIM3v3 model provided for the ST Microelectronics 90nm CMOS process. On-chip inductors are modeled and designed using the ASITIC electromagnetic field solver. Fig. 5.8 depicts the $\pi$-model for a 5nH spiral inductor used as $L_g$ and $L_d$ in Fig. 5.1, implemented using the thick metal layer of a 90nm technology. This inductor has a $Q$ of 7.3. To further improve $Q$, two metal layers can be used to reduce the series resistance of the inductor. $L_s$ is modeled similarly.

![Figure 5.8 π-model of a 5nH spiral inductor.](image)
In this design, $C_m$ and $C_{tune}$ are assumed to be implemented using high-$Q$ metal-insulator-metal (MIM) capacitors. The widths of both transistors are sized equally. $V_{GS}$ is chosen to be 0.4V, slightly above the threshold voltage to exploit the high $g_m/I_D$ in moderate inversion. Table 1 is a summary of component values.

<table>
<thead>
<tr>
<th>Summary of LNA Component Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{DD}$ (V)</td>
</tr>
<tr>
<td>$L_R = L_d$ (nH)</td>
</tr>
<tr>
<td>$L_g$ (nH)</td>
</tr>
<tr>
<td>$C_m$ (fF)</td>
</tr>
<tr>
<td>$C_{tune}$ (fF)</td>
</tr>
<tr>
<td>$W_1/L_1$ (μm)</td>
</tr>
<tr>
<td>$W_2/L_2$ (μm)</td>
</tr>
<tr>
<td>$V_{in,DC}$ (V)</td>
</tr>
<tr>
<td>$V_B$ (V)</td>
</tr>
</tbody>
</table>

Figure 5.9 shows the voltage gain $A_v$ and NF of the LNA. The voltage gain is 22.7dB in the 2.4GHz band with a 3dB bandwidth of around 300MHz.

![Figure 5.9 Gain and NF of the proposed LNA.](image-url)
As depicted in Fig. 5.10, an $S_{11}$ of $-14.7$ dB provides a good input impedance matching to 50Ω. NF is 2.8 dB which is acceptable for short-range applications. IIP3 and P1dB are +5.14 dBm and -10 dBm, respectively. Table 2 summarizes the performance of the LNA. The circuit consumes 943 $\mu$W from a 1V supply.

![Figure 5.10 $S_{11}$ of the proposed LNA.](image)

**TABLE 2**

<table>
<thead>
<tr>
<th>Summary of LNA Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gain (dB)</strong></td>
</tr>
<tr>
<td><strong>NF (dB)</strong></td>
</tr>
<tr>
<td><strong>$S_{11}$ (dB)</strong></td>
</tr>
<tr>
<td><strong>IIP3 (dBm)</strong></td>
</tr>
<tr>
<td><strong>P1dB (dBm)</strong></td>
</tr>
<tr>
<td><strong>$P_{DC}$ (μW)</strong></td>
</tr>
<tr>
<td><strong>$f_c$ (GHz)</strong></td>
</tr>
<tr>
<td><strong>Gate $L$ (μm)</strong></td>
</tr>
</tbody>
</table>
Performance Summary

Table 3 shows a performance comparison between the proposed LNA and low-power CMOS LNAs from recent literature. The performance comparison is further illustrated on the gain-power design space as in Fig. 5.11.

### Table 3

**COMPARISON OF CMOS LOW-POWER LNAs**

<table>
<thead>
<tr>
<th></th>
<th>This work</th>
<th>[3]</th>
<th>[28]</th>
<th>[2]</th>
<th>[29]</th>
<th>[30]</th>
<th>[31]</th>
<th>[32]</th>
<th>[33]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain (dB)</td>
<td>22.7</td>
<td>13.6</td>
<td>13*</td>
<td>12</td>
<td>9.2*</td>
<td>11.5</td>
<td>10.1</td>
<td>12.8</td>
<td>19</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>2.8</td>
<td>4.6</td>
<td>3.6</td>
<td>1.8</td>
<td>3.6</td>
<td>3.4</td>
<td>2.9</td>
<td>1.4</td>
<td>2.8</td>
</tr>
<tr>
<td>$S_{11}$ (dB)</td>
<td>-14.7</td>
<td>-5</td>
<td>&lt;-10</td>
<td>-18</td>
<td>-10</td>
<td>-14</td>
<td>-10.1</td>
<td>-14.5</td>
<td>-14.5</td>
</tr>
<tr>
<td>$IIp3$ (dBm)</td>
<td>5.14</td>
<td>7.2</td>
<td>-10</td>
<td>-3</td>
<td>-7.25</td>
<td>- 4</td>
<td>+4</td>
<td>+13.3</td>
<td>-</td>
</tr>
<tr>
<td>$P1dB$ (dBm)</td>
<td>-10</td>
<td>-0.2</td>
<td>-</td>
<td>-15.8</td>
<td>-8</td>
<td>-7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$P_{DC}$ (mW)</td>
<td>0.94</td>
<td>0.26</td>
<td>0.72</td>
<td>0.9</td>
<td>1</td>
<td>4</td>
<td>11.7</td>
<td>14.4</td>
<td>15</td>
</tr>
<tr>
<td>$f_c$ (GHz)</td>
<td>2.4</td>
<td>1</td>
<td>0.96</td>
<td>2.4</td>
<td>5.5</td>
<td>5.7</td>
<td>2.4</td>
<td>2.0</td>
<td>2.4</td>
</tr>
<tr>
<td>Gate $L$ ($\mu$m)</td>
<td>0.09</td>
<td>0.18</td>
<td>0.13</td>
<td>0.18</td>
<td>0.09</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
</tbody>
</table>

* Power gain, 1 Subthreshold design, 2 UWB design, 3 Dual-band design, # Measurement

---

**Figure 5.11** Graphical illustration comparing recently published LNAs and this work.
A direct comparison of LNAs listed in Table 3 is a challenge as most LNAs have specifications that are tailored to work with the specific radio architecture and application. An attempt is made to compare the proposed LNA to comparable low-power, relatively narrow-band LNAs. There are several aspects to note in this comparison. First, the proposed work is based on simulation results and the others are based on measurements. When a chip is fabricated, its performance tends to degrade due to unaccounted parasitics and the assumptions made during circuit design. Second, not all the LNAs compared operate at the same frequency. Typically, more power is required to provide gain at a higher frequency, for example, the LNAs in [3] and [28] have lower power consumption. Third, all LNAs use the cascode topology or its variants, except for the ones in [3] and [28]. The LNA in [3] uses a common-source topology, which exploits the relatively high MOSFET output resistance in 0.18μm CMOS. This LNA design may have difficulty delivering gain when ported to the 90nm technology. Also, the fact that it is targeted for lower frequency operation enables a subthreshold design, giving it an advantage in transconductance efficiency over other LNAs. The LNA in [28] uses the common-gate topology as it needs to provide a large bandwidth for the targeted ultra-wideband application. Since this design and the dual-band LNA in [31] need to cover a greater bandwidth, a higher power consumption and a lower gain is inevitable. Fourth, the LNA in [32] employs an extra transistor for post-linearization, which results in larger power consumption but the best linearity in the comparison. From Fig. 5.11, it can be seen that the LNA designed based on the proposed approach demonstrates a competitive gain and noise figure amongst low-power LNAs.
Chapter 6
Conclusions and Future Works

Conclusion

In this thesis, a design methodology for low-power CMOS RF inductively degenerated LNAs has been introduced. The design approach presented in this dissertation differs from existing techniques in that it begins with determining how device biasing affects gain, noise, linearity, and power consumption of the amplifier to ensure a well-rounded design. Design plots have been used to help the circuit designer in understanding the fundamental characteristics of the MOSFET in preparation to applying the design methodology. To demonstrate the technique, a fully integrated RF CMOS LNA is design and simulated. This 2.4GHz 90nm cascode LNA achieves a voltage gain of 22.7dB, NF of 2.8dB, IIP3 of +5.14dBm, and P1dB of -10dBm, while consuming 943μW from a 1V supply. This study has revealed that for designs that operate well below $f_T$, a balanced design can be achieved by biasing the device just above the threshold voltage to exploit the high $g_m/I_D$. This achieves low power consumption while avoiding the performance degradation associated with the subthreshold region. This observation is significant as many RF circuits for applications operating in the low-GHz range implemented in a modern CMOS technology can be designed based on the proposed technique.
**Future Work**

Although this work has provided solutions to the challenges of power-efficient design, it has also uncovered many interesting topics worthy of further investigation. They often require a deeper understanding and characterization of the device and are listed as follows:

1) Conventional power-constrained noise optimization methods proposed to target low-power design have their limits. The techniques is based on first finding an optimal gate width, then biasing the device with the amount of drain current allowed by the power constraint. In sub-mW designs, the large optimal gate width combined with a small drain current often force the device to be biased in the subthreshold region. When subthreshold operation is inadequate, such as insufficient frequency response, the technique is no longer applicable. An effective noise optimization technique for the ultra-low-power design space is needed.

2) The proposed design methodology has been devised with a focus of maximizing transconductance efficiency, which translates to power-efficient amplifier designs. Although current low-power applications tend to have a more relaxed noise requirement, it is worthwhile to incorporate noise formulation into the transconductance efficiency framework so that the establishment of biasing condition can include a quantitative noise analysis.
3) The effectiveness of the proposed design methodology has thus far been verified with SpectreRF simulations. Silicon implementation is the next logical step to confirm simulation results.

4) There is only a narrow range of biasing conditions that can exploit the high MOSFET linearity performance in moderate inversion. This narrow range demands for accurate biasing circuits; therefore, conventional LNA designs have not exploited this effect. Biasing circuits with improved accuracy may make the use of the high MOSFET linearity performance possible and reliable.

5) On-chip inductors are area intensive and do not scale well with technology. They are of low quality factor and contribute substantially to the noise figure of the amplifier. Improved on-chip implementation is needed.

6) Process variation is expected to increase as CMOS technology scales. Advanced layout techniques and robust calibration schemes for mitigating process variation is needed.

There is much work to be done in the area of low power, deep submicron CMOS LNA design. The landscape of analog circuit design is rapidly changing as technology scales to quantum levels, opening the designer to a new world of challenges and possibilities.
References


