AN ABSTRACT OF THE THESIS OF

Jianjun J. Zhou for the degree of Doctor of Philosophy in Electrical and Computer Engineering presented on August 18, 1998. Title: CMOS Low Noise Amplifier Design Utilizing Monolithic Transformers.

Abstract approved: _____
David J. Allstot

Full integration of CMOS low noise amplifiers (LNA) presents a challenge for low cost CMOS receiver systems. A critical problem faced in the design of an RF CMOS LNA is the inaccurate high-frequency noise model of the MOSFET implemented in circuit simulators such as SPICE. Silicon-based monolithic inductors are another bottleneck in RF CMOS design due to their poor quality factor.

In this thesis, a CMOS implementation of a fully-integrated differential LNA is presented. A small-signal noise circuit model that includes the two most important noise sources of the MOSFET at radio frequencies, channel thermal noise and induced gate current noise, is developed for CMOS LNA analysis and simulation. Various CMOS LNA architectures are investigated. The optimization techniques and design guidelines and procedures for an LC tuned CMOS LNA are also described.

Analysis and modeling of silicon-based monolithic inductors and transformers are presented and it is shown that in fully-differential applications, a monolithic transformer occupies less die area and achieves a higher quality factor compared to two independent inductors with the same total effective inductance. It is also shown that monolithic transformers improve the common-mode rejection of the differential circuits.

[1] Portions of this thesis have been submitted to the IEEE J. Solid-State Circuits for possible publication. Copyright may be transferred from the author to the IEEE without notice.

CMOS Low Noise Amplifier Design Utilizing Monolithic Transformers

by

Jianjun J. Zhou

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed August 18, 1998

Commencement June 1999

Doctor of Philosophy thesis of Jianjun J. Zhou presented on August 18, 1998

APPROVED:

Major Professor, representing Electrical and Computer Engineering

Head of Department of Electrical and Computer Engineering

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Jianjun J. Zhou, Author

ACKNOWLEDGMENTS

As I look back to my first day at Oregon State, to my first college day at Shanghai Jiao Tong University, to my first elementary-school day in my mother's arms, I finally realize how truly fortunate I am. I am not sure if such a couple of pages, appearing in the front of my thesis with the same style and format, could possibly express my thankful and sincere heart. So many terrific individuals have contributed to the education I have received throughout the years which made it possible!

First and foremost, I would like to thank my major advisor Professor David J. Allstot for his encouragement, patience, and support of the work and of my education. I hope he realizes how much I gained from his excellent teaching and great insight into almost every problem an IC engineer could possibly have encountered.

Thanks are due Professor Kartikeya Mayaram, Professor Virginia Stonick, Professor Andreas Weisshaar, and Professor Dwight Bushnell for taking time to serve on my committee, for reviewing the thesis, and for helpful guidance. Thanks also go to Professor Un-Ku Moon, for many enlightening and enjoyable discussions. In addition, I would like to thank many teachers I have had over the years for providing many helpful and engaging lectures and for their devotion to Electrical and Computer Engineering.

I would especially like to thank Richard Li, Enrique Ferrer, and Georgia Salgado of Motorola Inc., Ft. Lauderdale, Florida, for providing useful discussions and test and measurement assistance. Helpful discussions and technical exchanges with fellow researchers and graduate students Brian Ballweber, Hairong Gao, Ravi Gupta, Anping Liu, Hiok-Hion Ng, Hiok-Tiaq Ng, Yihai Xiang, and Ramsin Ziazadeh of Oregon State University are also deeply appreciated.

Special thanks to Wenjun Su and his wife Wenjing Zhang, Rugang Ding and his wife Lian Song, Yun Ge, Jeff Ward, Yunteng Huang, Bo Zhang, Bo Wang, and Haiqing Lin for their constant help and encouragement throughout my years at Oregon State University. Much thanks go to my old high-school and JiaoDa buddies Haiwen Xi, Xiaojie Yuan, Gong Jun, Xiao Bei, Lu Da, Omega, Zhang Zhenfan, Li Zuxin, and Lei Ming who have been providing numerous laughs and good times since 10+ years ago.

No words, however, could possibly be enough to express my thanks to my family: my great parents Zhou Wanyao and Yi Xueying, my little brother Zhou Jianmin, and my best friend and wife Qin, for their love and guidance, for supporting me through every step in my life. And I am here to say: I did it and it is for you!

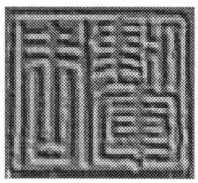


TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

CMOS LOW NOISE AMPLIFIER DESIGN UTILIZING MONOLITHIC TRANSFORMERS

CHAPTER 1. GENERAL INTRODUCTION

1.1 Scope

With the recent proliferation of wireless communication applications, there is an extensive effort to develop low cost, highly integrated CMOS RF circuits which meet the performance requirements of current and future communication system standards. This research is to support the development and implementation of low-power CMOS RF low noise amplifiers (LNA), which are a critical building block at the front-end of wireless communication systems.

The primary goal of this research is to design a fully-integrated 900MHz CMOS LNA with 15-20mW power consumption using a 3V power supply. To accomplish this goal, the high-frequency noise characteristics of the MOS transistor, high quality passive elements, and low-power circuit techniques have been exploited.

1.2 Wireless Receivers

The capability of electromagnetic waves to provide wireless distant communications has been a major factor in the explosive growth of communications during the twentieth century. In 1862, Maxwell predicted the existence of electromagnetic waves which was proven by Hertz 26 years later. The first wireless receiver was probably built with a tuned antenna and some iron dust at the end to observe a tiny spark generated by activating the transmitter [1]. Then with the advent of vacuum tube (replaced quickly by

transistor after its invention in 1947), active amplifiers were introduced into receivers to improve system sensitivity and selectivity (tuned amplifiers). Frequency translation schemes were employed thereafter to build so-called homodyne (coherent) and heterodyne (non-coherent) receivers which provided a potential improvement and cost reduction over previous receivers. Soon after that, the superheterodyne receiver was invented by Armstrong (this "poor" guy spent half of his life in the court to fight for his royalties and concluded his life with a suicide [2]).

The superheterodyne receiver makes use of the heterodyne principle of mixing an incoming signal with a local oscillator (LO) signal in a nonlinear element called a mixer (Fig. 1.1). However, rather than synchronizing the frequencies, the superheterodyne receiver uses a LO frequency offset by a fixed intermediate frequency (IF) from the desired signal. Although it does introduce a problem of spurious responses not present in other receiver types, the superheterodyne receiver predominates in most modern wireless communication applications in that it offers many advantages:

i) RF tuning can be done by varying the LO frequency, which eases the design of bandpass filters (BPF) and amplifiers at the RF front-end.

ii) Channel selectivity is accomplished at the IF section where narrow high-order filtering is more easily achieved.

iii) Amplification can be provided primarily at lower frequencies where high gain is generally more economical.

As shown in Fig. 1.1, the RF signal is fed from the antenna to a preselection filter which serves to attenuate the image signal and the undesired signals outside the service band. Insertion loss of the filter decreases the sensitivity of receiver and thus the filtering is often broken into two or more parts with intervening low noise amplifiers (LNA) to

provide sufficient selectivity, while minimizing the effects of the filter loss on noise performance. The gain of the low noise amplifier is needed to overcome the filter loss but must not be too high to retain system dynamic range. Next, the local oscillator converts the RF signal to the fixed IF at the mixer. The output from the mixer is applied to the IF filter for channel selectivity and then to the IF amplifier to obtain a suitable power level for demodulation.

There are several alternative receiver architectures worthy of mention [3]. A double-conversion superheterodyne receiver converts the incoming RF signal first to an IF

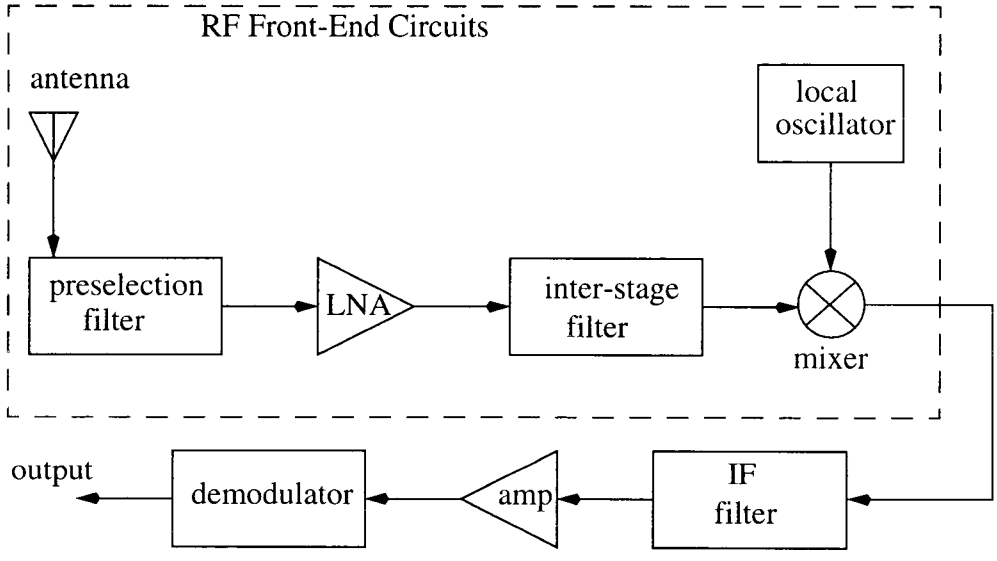


Figure 1.1: Block diagram of superheterodyne receiver.

at a relatively high frequency, and then to a lower second IF. This architecture further relaxes the bandpass filter design at the cost of more circuit complexity and power consumption. An ultra-low IF receiver uses carefully selected RF and LO frequencies so that the image signal falls within an unused portion of the RF spectrum. As a result, no image rejection bandpass filter is required. The homodyne (zero-IF) receiver, re-claiming its lost popularity in low-power integrated design, eliminates the need for image-rejection bandpass filters in that it makes the image signal exactly the same as the desired signal.

1.3 Integration of CMOS LNA's

A typical wireless receiver is required to perform selection, amplification and demodulation of received signals. The noise factor F (called noise figure NF if in decibel) of a receiver is a measure of its ability to amplify and demodulate weak signals and can be defined as the ratio of the input signal-to-noise-ratio (SNR) to the output signal-to-noise-ratio (SNR)

$$F = SNR_{in}/SNR_{out} \tag{1.1}$$

and the overall noise factor of a cascade system is given by

$$F = F_1 + \frac{F_2 - 1}{G_1} + \ldots + \frac{F_k - 1}{G_1 G_2 \ldots G_{k-1}} \tag{1.2}$$

where F_k and G_k are the noise factor and power gain, respectively, of the *kth* block in the system.

From Equation (1.2), it is clear that system noise performance depends primarily on the first function block which is the filtering block including the low noise amplifier as shown in Fig. 1.1. The minimum signal level that can be detected in a wireless receiver is the required output SNR (determined by the bit error rate in the following DSP block) plus the system noise floor which is equal to the sum of the input noise power in decibels and

the system noise figure. Therefore, the noise performance of the LNA is crucial in achieving high system sensitivity for a wireless receiver.

Design issues surrounding integrated CMOS receiver have been addressed in many articles [3]-[6]. While fine-line CMOS technology easily provides high frequency active devices for use in RF applications (e.g., 800MHz-2.4GHz), high quality passive components, especially inductors, present serious challenges to silicon integration. Several previous RF CMOS low noise amplifiers have reported impressive results [7]-[12]. However, since many of the CMOS implementations require inductors for narrowband tuning or matching, off-chip inductors have often been used owing to the relatively poor quality of monolithic inductors. In addition, the existing noise model for the MOS transistor in standard circuit simulators, such as SPICE, is insufficient to accurately predict the circuit noise performance at high frequencies. A lack of comprehensive understanding of the noise characteristics of MOS transistors at high frequencies is still an obstacle to the development and optimization of CMOS RF LNA designs [11].

Efforts towards the full integration of a CMOS LNA are described in this thesis. The inductance required to implement a fully-differential CMOS LNA is provided by monolithic transformers, instead of monolithic inductors, to achieve better circuit performance [13]. High frequency noise characteristics of the MOS transistor are investigated and a compact noise model is developed for circuit simulation. In addition, low-power circuit techniques have also been investigated.

1.4 Thesis Outline

High-frequency noise characteristics of the MOS transistor are investigated in Chapter 2. Various noise sources are carefully studied. The channel thermal noise and the

induced gate current noise are integrated into a compact noise circuit model for the MOS transistor used for hand analysis and SPICE simulations.

Basic principles of CMOS LNA design are presented in Chapter 3. Two measures of CMOS LNA circuit performance, i.e., noise figure and linearity, are discussed in detail. Basic architectures of CMOS LNA's are analyzed based on the compact noise circuit model developed in Chapter 2. The LC tuned CMOS LNA is found to have the best noise performance. Finally, design considerations of CMOS LNA's are discussed and design guidelines and procedures for CMOS LNA optimization are summarized.

Throughout the investigation of CMOS LNA architectures in Chapter 3, the importance of high-quality monolithic inductors is illustrated. Chapter 4 thus turns to the study of silicon-based monolithic inductors. First, a review of various implementations of monolithic inductors is given. Detailed analysis and modeling of the square spiral inductors is then described. The inductor circuit model developed can be used directly in an IC simulator, such as SPICE. Based on the circuit model, the inductor performance is analyzed and formulated to facilitate hand analysis. In addition, some alternative designs of spiral inductors are also discussed. It is concluded that the quality factor and self-resonant frequency of a monolithic spiral inductor are mainly limited by the parasitic series resistance and shunt capacitance.

In Chapter 5, a brief introduction to monolithic spiral transformers on silicon substrate is given first, followed by a detailed description of the characterization and modeling of a transformer consisting of two identical spiral inductors. Design guidelines for transformer optimization are then described, based on extensive simulation results. Computer simulation is also conducted to compare circuit performance between various transformers and inductors. When utilized in a differential application, it is found that a

transformer occupies less die area and thus has less parasitic series resistance and shunt capacitance, compared to two independent inductors with the same equivalent differential-mode inductance. As a consequence, the quality factor and the self-resonant frequency are improved. The improvements become more significant as the required equivalent inductance increases. In addition, a transformer provides additional common-mode rejection for the differential circuits.

Taking advantage of the transformers identified in Chapter 5, a 900MHz fully-differential CMOS LNA design is implemented and experimental results are presented in Chapter 6. The complete circuit and layout description are illustrated. The LNA which is fully integrated in a standard digital 0.6μm CMOS technology utilizes three monolithic transformers for on-chip tuning networks. Bias current re-use is used to reduce power dissipation and process-, voltage-, and temperature-tracking biasing techniques are employed. This chapter concludes with the discussion of experimental results.

Finally, conclusions are presented in Chapter 7.

CHAPTER 2. NOISE CHARACTERISTICS OF MOS TRANSISTORS

2.1 Introduction

The sensitivity of a wireless receiving system is mainly determined by the noise performance of the front-end circuits, specially, the low noise amplifier (LNA). Since the noise performance of an integrated circuit is determined by the noise behavior of the active components, it is important to understand the noise characteristics of MOS transistors in order to analyze and optimize CMOS LNA designs.

Conventional noise models for CMOS devices are deficient for RF circuit design. At high frequencies, the induced gate current noise is comparable to the channel thermal noise in MOS transistors. Additionally, gate resistance may contribute significant noise in a large transistor if it is not properly laid out. Furthermore, in submicron MOS transistors, there are many second-order noise mechanisms, such as hot carrier effects and substrate coupling, which may result in a large amount of excess device noise. Unfortunately, these effects are poorly modeled in the existing commercially available CAD tools such as HSPICE (version H96.1).

A detailed investigation of MOS noise characteristics is given in this chapter. All major noise mechanisms are studied and a compact noise circuit model of MOS transistors at high frequencies is developed based on previously reported research results. The noise model is then integrated with HSPICE, using a popular script program (Awk/Perl), for simulation of the noise performance of CMOS RF circuits.

2.2 Channel Thermal Noise

Channel thermal noise which is one of the most significant noise sources in a MOS transistor has been well studied and documented [14]-[15]. It is caused by a random motion of free carriers in the conductive channel, analogous to that in normal resistors. The current flowing between the drain and the source terminals in a MOS transistor is based on the existence of a conductive channel formed by an inversion layer. If the drain-source voltage $V_{DS} = 0\text{V}$, this conductive channel can be treated as a normal resistance. In such a case, the thermal noise drain current $\overline{i_d^2}$ is given by Nyquist [16]

$$\overline{i_d^2} = 4kTg_0 \cdot \Delta f \tag{2.1}$$

where k is the Boltzman constant ($1.38 \times 10^{-23}\text{J/°K}$), T is the absolute temperature, g_0 is the channel conductance at zero drain-source voltage, and Δf is the bandwidth of interest.

Inspection of Equation (2.1) reveals that this expression is nonphysical since it indicates that the total current noise power approaches infinity if the bandwidth is increased without limit, which is contradictory to what is found in the real world. The more complete expression for the resistive thermal noise is [17]

$$\overline{i_d^2} = 4kTg_0p(f) \cdot \Delta f \tag{2.2}$$

where $p(f)$ is the Planck factor, given by

$$p(f) = \frac{hf/kT}{e^{hf/kT} - 1} \tag{2.3}$$

where h is Planck's constant, $6.63 \times 10^{-34}\text{J-s}$. As long as $hf/kT \ll 1$, $p(f)$ is very close to 1. Therefore, Equation (2.1) will remain valid for all practical frequencies. For example, at 290°K, $p(f)$ is greater than 0.999 up to 10GHz. However, as the frequency increases further, $p(f)$ decreases rapidly. Based on Equation (2.2), the total thermal noise power over an infinite bandwidth available from a resistor is approximately $4 \times 10^{-8}\text{W}$ for T=290°K.

In most applications, MOS transistors operate in the saturation region in which the conductive channel cannot be considered as a resistance. It has been shown that the thermal noise drain current for a MOS transistor in saturation can be expressed as [18]

$$\overline{i_d^2} = 4kT\frac{\mu}{L_{eff}^2}Q_N \cdot \Delta f \tag{2.4}$$

where μ is the effective carrier mobility, L_{eff} is the effective channel length, and Q_N is the total inversion layer charge. Q_N is a complex function of the gate-bulk voltage, drain potential, source potential, drain-source current, channel width and length, gate oxide capacitance, and bulk doping concentration. A complete and precise expression for Q_N is given in [19]. For simplicity, however, the channel thermal noise in a MOS transistor in saturation is often written as [20]

$$\overline{i_d^2} = 4kT\gamma g_m \cdot \Delta f \tag{2.5}$$

where g_m is the transconductance of the device and γ is the noise coefficient. In general, γ depends on all the terminal bias voltages and the basic transistor parameters. A numerical approach is required to interpret the dependencies of γ in expression (2.5). It has been shown that γ falls between 2/3 and 1 for long-channel MOSFETs (in which the effective channel length can be approximated by the drawn channel length between the drain and the source) if the bulk doping concentration is low and the gate oxide thickness is small [21]. When the bulk effect can be completely ignored, the value of γ is about 2/3.

For short-channel devices, however, γ may increase since channel length modulation is more pronounced and the effective channel length is relatively much shorter than the drawn channel length [22]. In addition, the carrier temperature increases with increasing field strength in the channel. Therefore, the high electric fields in submicron MOS devices produces hot carriers with temperatures higher than the lattice temperature. The presence of excess thermal noise that has been experimentally verified [23]-[24] is

attributed to this hot carrier effect. For example, γ was measured as high as 2 to 7 for a 0.7-μm channel length MOSFET depending on bias conditions [23].

2.3 Induced Gate Current Noise

At high frequencies the MOSFET should be considered as an RC distributed network. The conductive channel can be treated as a distributed resistance while the gate oxide capacitance represents a distributed capacitance. This means that the gate impedance of the device will exhibit a resistive component at high frequencies. This can be accounted for by shunting the gate oxide capacitance with a conductance g_g which can be expressed as [25]-[27]

$$g_g = \frac{4}{45}\frac{\omega^2(C_{ox} \cdot WL)^2}{g_m} \tag{2.6}$$

In saturation, g_g can be simplified to

$$g_g = \frac{1}{5}\frac{\omega^2 C_{gs}^2}{g_m} \tag{2.7}$$

assuming $C_{gs} = \frac{2}{3}(C_{ox} \cdot WL)$.

The conductance g_g has noise associated with it. Called induced gate current noise $\overline{i_g^2}$, this noise is caused by the random motion of free carriers in the channel coupling through the gate capacitance. If the device were a passive device, the noise would simply be the resistive thermal noise $\overline{i_g^2} = 4kTg_g \cdot \Delta f$. However, since the resistive channel can not be considered as a homogeneous resistance when the device is in saturation, the calculation of the induced gate current noise is rather complex. As has been shown in [25]-[27], it is approximately given by

$$\overline{i_g^2} = 4kT\beta g_g \cdot \Delta f \tag{2.8}$$

where β is the coefficient of gate noise, equal to 4/3 for a long-channel MOSFET. For short channels in which hot carrier effects cannot be ignored, β may be larger.

The current fluctuations through the gate and drain are correlated since both are generated by the random motion of free carriers in the channel. The correlation coefficient has a complex value. For long-channel devices, a first-order approximation to the correlation coefficient c is given theoretically by [27]

$$c = \frac{\overline{i_g i_d^*}}{\sqrt{\overline{i_g^2} \cdot \overline{i_d^2}}} = 0.395j \tag{2.9}$$

More accurate calculation shows that the correlation coefficient c has a real part which is significant at high frequencies and can be approximated by [28]

$$c = \frac{\overline{i_g i_d^*}}{\sqrt{\overline{i_g^2} \cdot \overline{i_d^2}}} = 0.13\frac{\omega C_{gs}}{g_m} + 0.35j \tag{2.10}$$

For typical radio frequencies, however, c can be simply assumed to be 0.35j.

Assuming the transfer functions of i_g and i_d are H_g and H_d respectively, the output noise due to the current fluctuations through the gate and drain can be calculated by

$$\begin{aligned} N_o &= \overline{(H_g i_g + H_d i_d) \times (H_g i_g + H_d i_d)^*} \\ &= |H_g|^2 \overline{i_g^2} + |H_d|^2 \overline{i_d^2} + \overline{H_g i_g H_d^* i_d^*} + \overline{H_d i_d H_g^* i_g^*} \\ &= |H_g|^2 \overline{i_g^2} + |H_d|^2 \overline{i_d^2} + 2Re(cH_g H_d^*) \cdot \sqrt{\overline{i_g^2} \cdot \overline{i_d^2}} \end{aligned} \tag{2.11}$$

where Re stands for the real part. It is clear from Equation (2.11) that the correlation between $\overline{i_g^2}$ and $\overline{i_d^2}$ can be computed if the transfer functions of them are known.

2.4 Resistive Gate Thermal Noise

The thermal noise generated in the resistive gate of the MOS device is an often overlooked additional noise component which should be taken into consideration when laying out wide MOS devices [29]-[31].

To calculate the resistive gate thermal noise, consider a general layout for the gate structure shown in Fig. 2.1. We shall only calculate the thermal noise contributed from the poly-gate over thin oxide (active area). The resistance of interconnect metal and the poly-gate over the thick oxide are neglected due to their small value. An elementary section Δx at position x in the i-th poly-gate finger, as shown in Fig. 2.1, has a thermal noise voltage ΔV given by

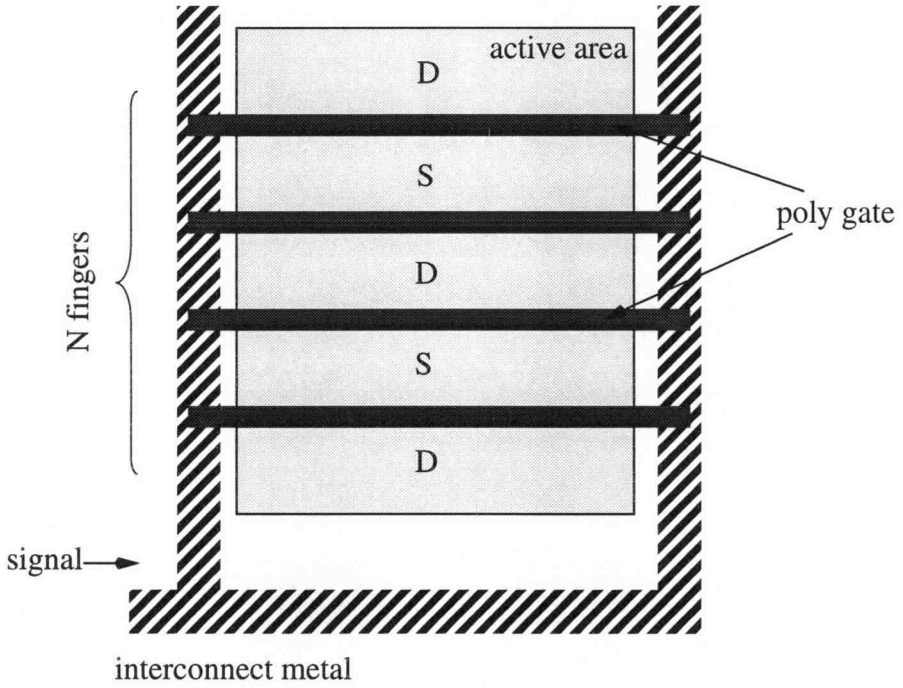
$$\Delta V = \sqrt{4kT\frac{R_{sq}\cdot\Delta x}{L}\cdot\Delta f} \tag{2.12}$$

where R_{sq} is the sheet resistance of the poly-gate and L is the channel length. This thermal noise voltage ΔV will cause a voltage fluctuation along the i-th gate finger. At position x', the voltage fluctuation $\delta V_i(x')$ is given by

$$\delta V_i(x') = \delta V_i(0) + \frac{x'}{W_i}\Delta V, \qquad 0<x'<x \tag{2.13}$$

$$\delta V_i(x') = \delta V_i(W_i) - \frac{W_i - x'}{W_i}\Delta V, \qquad x<x'<W_i \tag{2.14}$$

where W_i is the width of the i-th gate finger, equal to W/N. $\delta V_i(0)$ and $\delta V_i(W_i)$ are the voltage fluctuation at two ends of the i-th gate finger. Since both ends are tied together and have a dc path to ground via interconnect metal, $\delta V_i(0)$ and $\delta V_i(W_i)$ are equal to zero. Therefore, the channel current fluctuation Δi_c due to ΔV can be calculated by



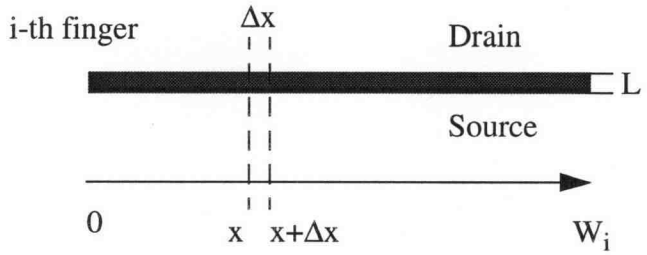


Figure 2.1: Finger structure of a MOS transistor with large *W/L*.

$$\begin{aligned}\Delta i_c &= \int_0^{W_i} g_{m0} \cdot \delta V_i(x')dx' \\ &= g_{m0}\left[\int_0^x \left(\frac{x'}{W_i}\Delta V\right)dx' - \int_x^{W_i}\left(\frac{W_i - x'}{W_i}\Delta V\right)dx'\right] \\ &= g_{m0}\Delta V \cdot (x - W_i/2)\end{aligned} \tag{2.15}$$

where g_{m0} is unit-width transconductance of the transistor, defined as $g_{m0} = g_m/W$.

Since the thermal noise voltage generated by every section Δx in the *i*-th gate finger can be considered un-correlated, the total channel current noise generated by the *i*-th gate finger is given by

$$\begin{aligned}\overline{i_c^2} &= \int_0^{W_i} (\Delta i_c)^2 \\ &= \int_0^{W_i} g_{m0}^2 \cdot (x - W_i/2)^2 \cdot 4kT\frac{R_{sq} \cdot \Delta x}{L} \cdot \Delta f \\ &= 4kT\frac{R_{sq}}{L}\frac{g_{m0}^2 W_i^3}{12} \cdot \Delta f\end{aligned} \tag{2.16}$$

Hence the channel current noise generated by all N gate fingers is

$$\begin{aligned}\sum_1^N \overline{i_c^2} &= 4kT\frac{R_{sq}}{L}\frac{g_{m0}^2 W_i^3}{12} \cdot \Delta f \cdot N \\ &= 4kT\frac{R}{12N^2}g_m^2 \cdot \Delta f\end{aligned} \tag{2.17}$$

where $R = R_{sq} \cdot W/L$ is the total poly-gate resistance. Equation (2.17) indicates that the resistive gate thermal noise can be modeled by a series resistance at the gate given by

$$R_g = \frac{R}{12N^2} \tag{2.18}$$

From Equation (2.18) it is clear that in order to minimize the noise associated with the resistive poly-gate, the number of gate fingers *N* should be chosen as large as possible. For example, assuming the total poly-gate resistance *R* of a large device is 500Ω, the effective gate thermal resistance R_g is only about 0.4Ω if *N* is chosen as 10.

It is worthy of mention that the above equations hold only if both sides of the gate fingers are tied together. If one side of the gate fingers is left open (assuming at $x'=0$), the voltage fluctuation due to Δx along the i-th gate finger would be ΔV for $0<x'<x$ and 0 for $x<x'<W_i$. As a consequence, the channel current fluctuation Δi_c due to Δx is $g_{m0}\Delta V \cdot x$ and the total channel current noise generated by the i-th gate finger is now given by

$$\begin{aligned} \overline{i_c^2} &= \int_0^{W_i} (\Delta i_c)^2 \\ &= 4kT\frac{R_{sq}}{L}\frac{g_{m0}^2 W_i^3}{3} \cdot \Delta f \end{aligned} \tag{2.19}$$

This indicates the series gate resistance is increased by a factor of four to

$$R_g = \frac{R}{3N^2} \tag{2.20}$$

Therefore, for low noise design, it is desirable to tie both sides of the gate fingers together using a low-resistance interconnect such as metal.

2.5 Additional Noise Sources

2.5.1 Flicker (1/f) Noise and Shot Noise

Flicker noise was first observed by Johnson in 1925 [32]. Since its spectrum varies as $1/f^{\alpha}$, with α close to unity, it is also often called 1/f noise. Despite continuous pursuit, the physical mechanism behind flicker noise is still not very clear. Since flicker noise decreases rapidly with an increase in frequency, it is negligible in RF CMOS LNA's compared to the thermal noise. Hence, the treatment of flicker noise will not be discussed in this thesis. There are many papers in the literature on the subject of flicker noise in MOSFETs both theoretical and experimental [33]-[35], where detailed analysis and modeling of flicker noise can be found.

Shot noise, also known as Schottky noise, is a result of the discontinuous character of electronic charges that constitute a dc current flowing through a potential barrier, such as a *pn* junction. A simple yet general form of the shot noise is [36]

$$\overline{i_q^2} = 2qI_{DC} \cdot \Delta f \tag{2.21}$$

where q is the electronic charge, 1.6×10^{-19} C, and I_{DC} is the dc current in amperes. Formula (2.21) is valid for the frequencies lower than $f_T = 1/(2\pi\tau)$, where τ is the electron transit time in the device. Beyond f_T, the shot noise rolls off rapidly. Shot noise is one of the dominant noise sources in a bipolar transistor. Fortunately, in MOSFETs, shot noise is generally negligible since only the dc gate leakage current contributes shot noise and this gate leakage current is very small.

2.5.2 Substrate Resistance Noise

The thermal noise voltage across the distributed substrate resistance induces a fluctuating substrate potential. These random variations couple to the MOSFET channel, giving rise to fluctuations in the channel current [37]. In a standard CMOS process, a lightly doped epitaxial layer generates a significant resistance to ac ground under the device channel. The noise voltage generated across this resistance R_{sub} is given by

$$\overline{v_{sub}^2} = 4kTR_{sub} \cdot \Delta f \tag{2.22}$$

Thus, the device channel experiences current fluctuations given by

$$\overline{i_{sub}^2} = 4kTR_{sub}g_{mb}^2 \cdot \Delta f \tag{2.23}$$

where g_{mb} is the substrate transconductance, given by

$$g_{mb} = \frac{\alpha g_m}{2\sqrt{V_{SB} + |2\phi_F|}} \tag{2.24}$$

where ϕ_F is the difference between the quasi-Fermi level and the intrinsic level and α is the body-effect constant, with a typical value about $0.5V^{1/2}$.

Estimation of R_{sub} is quite complicated due to its distributed nature and its dependence on device layout. For a certain device, experiments showed that the substrate resistance noise added 25% more noise power to that already existing due to the channel thermal noise [37]. Two approaches can be used to minimize its contribution. First, reduce R_{sub} by changing device layout or by using a thinner and less lightly doped epitaxial material. Second, bias the substrate at high potential to reduce g_{mb}. In a typical CMOS process, g_{mb} is practically reduced to zero when V_{SB} is about -2V. In such a case, the substrate contribution to the channel noise is negligible.

In addition, impact ionization in the channel gives rise to an additional noise mechanism in MOSFETs. This impact-ionization-generated current appears as gate (or substrate) current flowing between the gate (or substrate) and the drain of the MOS transistor [38]-[39]. Also, noise from other components and its coupling through the chip interconnects, substrate, and package, can be detrimental to circuit performance [40].

High precision noise calculations including all noise sources is not only problematic given the disparate properties of the many noise sources as discussed above, it is also unnecessary from a practical viewpoint. First, many existing noise models are either qualitative or first-order approximations. Second, there are inevitable process variations of the parameters, and third, the noise performance of a circuit is typically determined by one or two dominant noise sources.

2.6 Small-Signal Noise Circuit Model

Having investigated the noise sources in a MOSFET, we are ready to develop a general small-signal noise circuit model of a MOSFET in order to facilitate the analysis and design of low noise CMOS circuits.

The induced gate current noise can be accounted for by placing a noise current source $\overline{i_g^2}$ in parallel with the gate capacitance. Note that the conductance g_g should also be included in the gate circuit to account for the distributed nature of the MOSFET channel at high frequencies. Such a circuit model is shown in Fig. 2.2 (a) where a shunt noise current $\overline{i_g^2}$ and a shunt conductance g_g have been added. Shaeffer and Lee [11] have derived the Thevenin equivalent circuit for (a), as shown in Fig. 2.2 (b). First, transform the parallel RC network into an equivalent series RC network. The impedance looking into the parallel RC network is

$$\begin{aligned} Z &= \frac{1}{g_g + j\omega C_{gs}} \\ &= \frac{1}{5g_m[1+(\omega C_{gs}/5g_m)^2]} + \frac{1}{j\omega C_{gs}[1+(\omega C_{gs}/5g_m)^2]} \end{aligned} \tag{2.25}$$

We observe that

$$(\omega C_{gs}/5g_m)^2 = \left(\frac{\omega}{5\omega_T}\right)^2 \ll 1 \tag{2.26}$$

usually holds for all practical frequencies. Here $\omega_T=g_m/C_{gs}$ is the transistor unity-gain frequency which indicates an upper limit on the maximum frequency at which the transistor can be effectively used (Appendix B). Thus Equation (2.25) can be simplified as

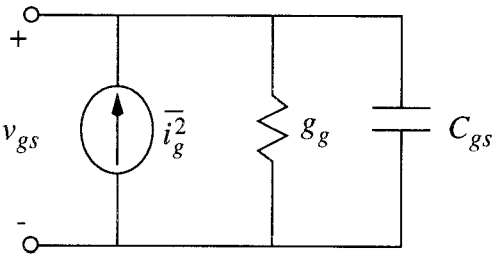
$$Z \approx \frac{1}{5g_m} + \frac{1}{j\omega C_{gs}} = r_g + \frac{1}{j\omega C_{gs}} \tag{2.27}$$

This indicates the equivalent series RC network consists of a frequency-independent resistance $r_g = 1/5g_m$ and the gate capacitance C_{gs}.

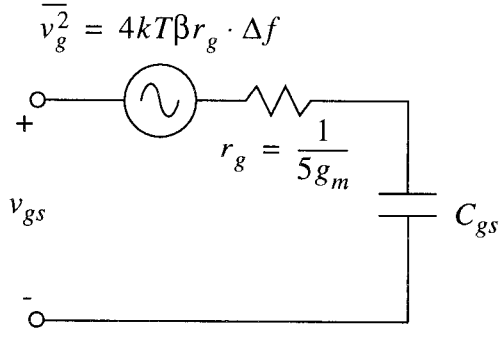
The Thevenin equivalent noise voltage is then given by

$$\begin{aligned} \overline{v_g^2} &= \overline{i_g^2} \cdot \left| \frac{1}{j\omega C_{gs}} + r_g \right|^2 \\ &= 4kT\beta r_g \cdot \Delta f \cdot \left[1 + \left(\frac{\omega}{5\omega_T} \right)^2 \right] \approx 4kT\beta r_g \cdot \Delta f \end{aligned} \tag{2.28}$$

where



(a)



(b)

Figure 2.2: Representations of the induced gate current noise in MOS transistor.

$$r_g = \frac{1}{5g_m} \tag{2.29}$$

This may seem attractive at first because $\overline{v_g^2}$ is similar to a resistive thermal noise, unlike $\overline{i_g^2}$ which is frequency-dependent. However, further observation shows that this treatment does not guarantee convenience in the analysis of circuit noise performance because, as shown in Fig. 2.2 (b), the critical voltage v_{gs} (which determines the output noise current) is still frequency-dependent, considering the impedance of the driving-source is usually not capacitive. Moreover, it is difficult to adopt Fig. 2.2 (b) into a circuit simulator since v_{gs} is no longer the voltage across C_{gs}. Therefore, we will not use Fig. 2.2 (b) in the following noise analysis.

Fig. 2.3 shows the complete small-signal noise circuit of a MOSFET based on the previous discussion. As shown, the channel thermal noise and the induced gate noise are included. The resistive gate thermal noise is modeled by the series resistance R_g. Flicker noise is not included since this noise model is intended to be used for RF CMOS LNA circuits. In addition, substrate resistance noise and other second-order noise sources are not included since they are mainly concerned with the layout or process parameters and have little connection with the circuit design parameters.

Given the noise circuit model, it is possible to calculate the noise performance of any RF LNA circuit composed of MOS transistors by the superposition of contributions of the individual noise sources. For more accurate analysis, the correlation between $\overline{i_g^2}$ and $\overline{i_d^2}$ should be taken into account and the substrate transconductance g_{mb}, the drain-source conductance g_{ds}, and bulk capacitances should also be added to the noise circuit. An optimal choice of the basic design parameters for each MOS transistor in low-noise CMOS circuits, such as W, L, and I_{DS}, can then be made with the aid of computer simulation.

Finally, the layout should be optimized in order to minimize additional noise contributions related to the layout parameters, such as the resistive gate thermal noise and substrate resistance thermal noise.

2.7 Modified Noise Simulation in SPICE

Though much progress has been achieved in high-frequency integrated circuit simulation using the standard circuit simulator SPICE, the noise models implemented in SPICE are not accurate enough at radio frequencies. The existing MOS noise models in

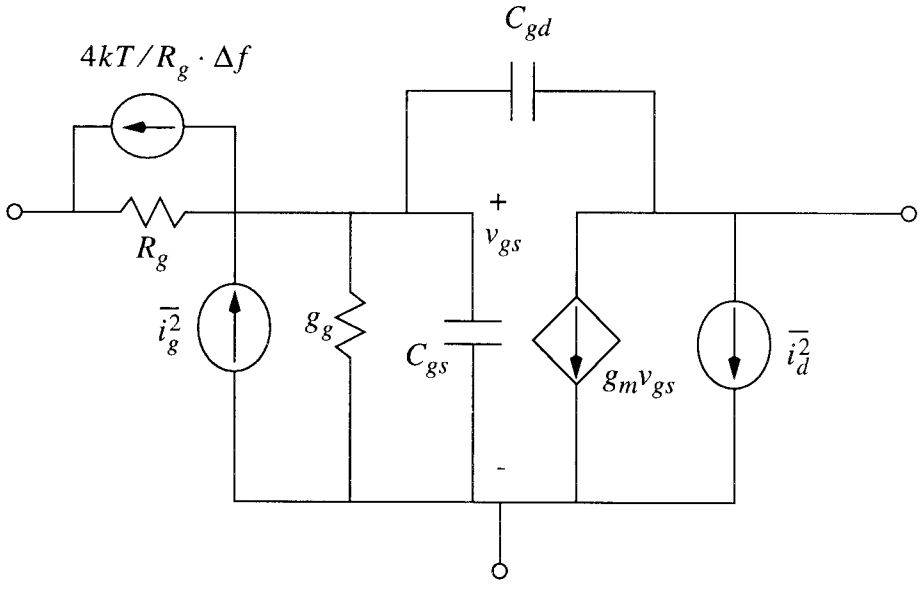


Figure 2.3: Small-signal noise circuit of a MOS transistor.

SPICE account for neither the gate resistance R_g nor the conductance g_g and its associated induced noise $\overline{i_g^2}$. Therefore, some modifications should be made when using SPICE to simulate the noise performance of RF circuits, in order to accommodate the high-frequency noise characteristics of MOS transistors.

To include the effect of the resistive gate in SPICE simulations, resistance R_g can be simply added in series with the gate of each MOS transistor in the circuit. Effects of the gate conductance g_g and its associated induced noise $\overline{i_g^2}$ on circuit performance are more complicated due to their frequency dependence. However, using a script program such as Awk or Perl, we can easily modify SPICE simulations to include the effects of the gate conductance g_g and the gate induced noise $\overline{i_g^2}$ at every single frequency point. Post-simulation data processing can then be adopted by extracting and plotting the single-frequency results to get a good presentation of the frequency response of the circuit. The following procedure is listed to illustrate one such method of the modified noise simulation using HSPICE.

Step 1. Form HSPICE input file with gate resistance R_g included: `input1`

Step 2. Conduct ac noise simulation in HSPICE at N frequency points over a specified frequency range (`f1` to `f2`) using `input1`: `result1`

Step 3. Extract C_{gs} and g_m from `result1` for all MOSFETs in the circuit; compute corresponding $g_g = (\omega^2 C_{gs}^2)/(5g_m)$ and $\overline{i_g^2} = 4kT\beta g_g$ (noise power within 1Hz) at N frequency points from `f1` to `f2`

Step 4. For a single frequency, modify `input1` to add g_g between the gate and source of all MOSFETs: `input2`

Step 5. Modify *input2* to add an unit ac current source in parallel with g_g for one MOSFET; conduct ac analysis in HSPICE at the specified single frequency; multiply the output voltage power by $\overline{i_g^2}$

Step 6. Repeat step 5 for all MOSFETs (to save simulation time, we may only repeat with MOSFETs that have significant noise contribution)

Step 7. Repeat steps 4-6 for *N* frequency points from *f1* to *f2*: *result2*

Step 8. Total output noise = output due to the resistive thermal noise: *noise1* (in *result1*) + output due to the channel thermal noise: *noise2* (in *result1*) + output due to the induced gate current noise: *noise3* (in *result2*)

Step 9. Data processing and plotting

The above procedure is elementary and somewhat tedious. For accurate noise simulation of CMOS RF circuits, it is an effective approach to include the effects of the induced gate current noise. Note that the correlation between the channel thermal noise and the induced gate current noise can also be computed using the transfer functions of the two noise sources obtained in SPICE (refer to Equation (2.11)). Nevertheless, efforts need to be made to improve the noise models of MOSFETs implemented in SPICE so that the noise simulation will be more convenient and accurate.

CHAPTER 3. PRINCIPLES OF CMOS LNA DESIGN

In this chapter, we shall discuss the basic principles of CMOS LNA design. Beginning with the discussion of some basic concepts, such as noise figure and linearity, we will study various CMOS LNA topologies in the radio frequency range (e.g., 800MHz-2.4GHz). It is concluded that a CMOS LNA with LC series tuning at its inputs offers the possibility of achieving the best noise performance. The design considerations of the LC tuned CMOS LNA are then discussed and a detailed description of the optimization techniques for basic device parameters is presented.

3.1 Basic Concepts

3.1.1 Noise Factor and Noise Figure

Before studying a CMOS LNA, it is necessary to understand the most popular figure of merit for noise performance, noise figure (NF). Friis [41] defined the noise factor (F) of a network to be the ratio of the signal-to-noise-ratio (SNR) at the input to the signal-to-noise-ratio (SNR) at the output; the noise figure (NF) is the logarithmic equivalent in decibels

$$NF = 10\log\left(\frac{SNR_{in}}{SNR_{out}}\right) = 10\log(F) \tag{3.1}$$

Thus the noise figure of a network is the decrease or degradation in the signal-to-noise ratio as the signal passes through the network. A perfect amplifier would amplify only the noise at its input along with the signal. A realistic amplifier, however, also adds some extra noise from its own components and degrades the signal-to-noise ratio. Equation (3.1) implies that a lower NF is achieved when the device noise contributes less to the total output noise; i.e., the input noise contributes a larger portion to the total output noise.

Noise figure is a meaningful parameter if and only if the input noise is well defined. The usual assumption is that input noise is the thermal noise power available from a resistor (typically 50Ω for wireless receiving systems) at a reference temperature, usually taken as 290°K [42] (close to the temperature seen by receiving antennas directed across the atmosphere at the transmitting antenna). Noise figure is generally a function of frequency but it is usually a bandwidth invariant parameter so long as the bandwidth is narrow enough to resolve variations with frequency. If the bandwidth is large, a frequency average of the spot noise figures over the band of measurement should be used. The spot noise figure, however, is the viable measure of a device, such as an LNA, for most wireless receiver applications due to the narrow-band characteristics.

It is worthwhile to mention what the noise figure does not characterize. The noise figure is not a measure of the noise performance of networks with one port, e.g., oscillators. Noise figure also has nothing to do with modulation. It is independent of the modulation format and of the fidelity of modulators and demodulators. One weakness of the noise figure is its being meaningful for a given device only in conjunction with a specified source impedance. Thus it cannot be used as a basis for evaluating the noise performance or for comparison of devices with different source impedances.

3.1.2 Nonlinearity

The nonlinearity of a device limits the maximum signals that may be processed. For a CMOS LNA, nonlinearity can be characterized either by the 1-dB compression point, defined as the input power at which the output power gain drops by 1-dB relative to the small-signal gain, or by the input-referred third-order intercept point (IIP3), the input power at which the third-order intermodulation term extrapolated from the small-signal values is equal to the fundamental.

Any pseudo-linear network can be characterized by a transfer function, the output voltage/current as a function of an input voltage/current. The transfer function may be characterized as a power series

$$S_o = a_0 + a_1 \cdot S_i + a_2 \cdot S_i^2 + a_3 \cdot S_i^3 + \ldots \tag{3.2}$$

assuming an input signal S_i and an output signal S_o.

Using Equation (3.2), the nonlinearity of an amplifier may be analyzed. Two types of input signals will be considered in the analysis. The first is a single frequency input (single-tone), S_i=$A\cos\omega t$. The other is a pair of unrelated inputs added to form a two-tone input, S_i=$A(\cos\omega_1 t+\cos\omega_2 t)$. A practical amplifier may, of course, be subjected to more complicated inputs. It is sometimes useful to consider more complicated input signals, containing three or more input tones. However, the analysis would become predictably messy [43]. Fortunately, most of the salient features of the nonlinearity are suitably characterized with the single-tone or two-tone inputs.

3.1.2.1 Gain Compression

The gain of a circuit can be obtained based on Equation (3.2). Let S_i be a single-tone signal ($A\cos\omega t$), then

$$\begin{aligned} S_o &= a_0 + a_1 \cdot A\cos\omega t + a_2 \cdot A^2\cos^2\omega t + a_3 \cdot A^3\cos^3\omega t + \ldots \\ &= a_0 + \frac{a_2 A^2}{2} + \left(a_1 A + \frac{3a_3 A^3}{4}\right)\cos\omega t + \frac{a_2 A^2}{2}\cos 2\omega t + \frac{3a_3 A^3}{4}\cos 3\omega t + \ldots \end{aligned} \tag{3.3}$$

From the above expansion, we can see that the gain of the circuit is $\left(a_1 + \frac{3a_3 A^2}{4}\right)$, neglecting other higher-order terms. The small-signal gain is a_1 when the effect of the amplitude A of the input signal S_i can be ignored. In most circuits, as A increases, the gain begins to drop from the small-signal gain a_1, which is usually referred to as gain

compression or saturation. The 1-dB compression point, defined as the input signal amplitude that causes the gain to drop by 1dB from the small-signal power gain, can then be calculated by

$$20\log\left|a_1 + \frac{3a_3A^2}{4}\right| = 20\log|a_1| - 1 \tag{3.4}$$

which indicates that the amplitude of S_i at the 1-dB compression point is approximately

$$A_{cp} = \sqrt{0.145\left|\frac{a_1}{a_3}\right|} \tag{3.5}$$

Please note Equation (3.5) is the first-order approximation for a pseudo-linear circuit. When higher-order terms are taken into consideration, the 1-dB compression point is usually lower than that expected from Equation (3.5) (refer to Fig. 3.1).

3.1.2.2 Intermodulation

Now, let S_i be a two-tone signal $A(\cos\omega_1 t+\cos\omega_2 t)$. We have

$$\begin{aligned} S_o = {} & a_0 + a_1A(\cos\omega_1 t + \cos\omega_2 t) + a_2A^2(\cos\omega_1 t + \cos\omega_2 t)^2 \\ & + a_3A^3(\cos\omega_1 t + \cos\omega_2 t)^3 + \ldots \end{aligned} \tag{3.6}$$

After simplification and collection of terms, we obtain

$$\begin{aligned} S_o = {} & a_0 + a_2A^2 + \left(a_1A + \frac{9a_3A^3}{4}\right)(\cos\omega_1 t + \cos\omega_2 t) \\ & + a_2A^2[\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t] \\ & + \frac{a_2A^2}{2}(\cos 2\omega_1 t + \cos 2\omega_2 t) + \frac{a_3A^3}{4}(\cos 3\omega_1 t + \cos 3\omega_2 t) \\ & + \frac{a_3A^3}{4}[\cos(\omega_1 + 2\omega_2)t + \cos(\omega_1 - 2\omega_2)t \\ & + \cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t] + \ldots \end{aligned} \tag{3.7}$$

It is seen that the output signal exhibits some components that are not harmonics of the input frequencies when a two-tone signal is applied to the network. This phenomenon is called intermodulation (IM). Intermodulation is a troublesome effect in a wireless receiving system. For example, the third-order intermodulation (IM3)

$$\frac{a_3 A^3}{4}[\cos(\omega_1 + 2\omega_2)t + \cos(\omega_1 - 2\omega_2)t + \cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t]$$

is of great importance for a superheterodyne receiver (IM2 is more important for a zero-IF or direct-conversion receiver). If the input tones (ω_1 and ω_2) are close to each other, the sum frequency terms in IM3 are close to the third harmonic and no more of a problem than harmonic distortion, for it may be filtered from the system. However, the difference frequency terms in IM3 are very close to that of the input tones and may eventually lie in the signal band. In a wireless receiving system, a weak signal accompanied by two strong interferers (for example, from an adjacent channel) would be corrupted by the third-order intermodulation terms.

A valuable figure of merit is the third-order intercept point (IP3). As illustrated in Fig. 3.1, the third-order intercept point is defined to be at the intersection of two extrapolated lines from the small-signal fundamental and IM3 curves. Please note we do not need to consider the higher-order effects for the third-order intercept point since the intercept is evaluated by extrapolating trends observed with sufficiently small amplitude inputs in both simulations and experiments. By this definition, we can easily compute the input-referred third-order point (IIP3) by setting the amplitude of the IM3 equal to the amplitude of the fundamental using small-signal results from Equation (3.7)

$$|a_1 A| = \left|\frac{a_3 A^3}{4}\right| \tag{3.8}$$

which gives the amplitude of the input signal at IP3 as

$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|} \tag{3.9}$$

Fig. 3.1 shows the simulation results for an amplifier with a 10dB small-signal gain, modeled by a hyperbolic tangent function. As shown, the fundamental curve has a slope of 1:1 and the IM3 curve has a slope of 3:1 when the input signal is sufficiently small, because, as indicated in Equation (3.7), the amplitude of fundamental ($a_1 A$) increases in proportion

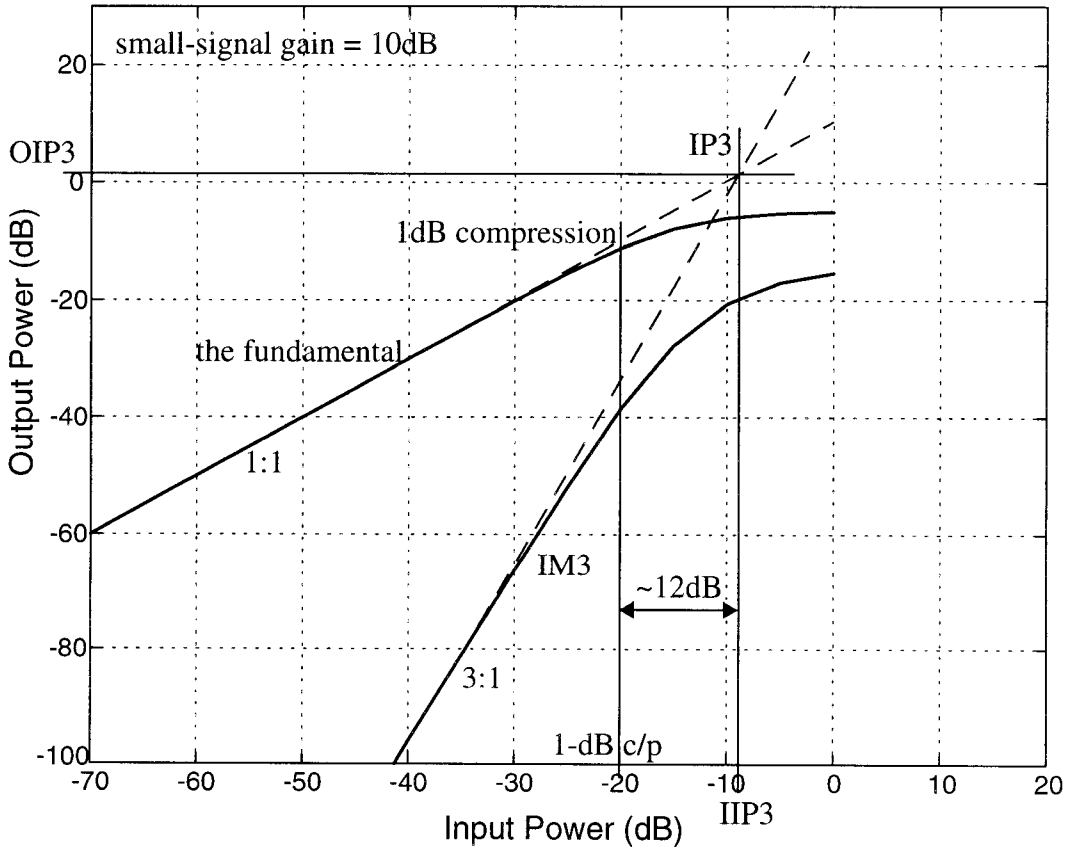


Figure 3.1: Graphical interpretation of the nonlinearity of an amplifier.

to A, while the amplitude of the IM3 ($a_3A^3/4$) increases in proportion to A^3. It is clear that the input-referred third-order intercept point (IIP3) is different from the output-referred third-order intercept point (OIP3) by the small-signal gain of the amplifier, i.e., 10dB in this example.

It is also observed the 1-dB compression point occurs at a lower input power than IIP3. This is usually true for most practical circuits. The relationship between the 1-dB compression point and IIP3 can be determined based on the foregoing analysis [44]. Combining Equations (3.5) and (3.9), we obtain

$$\frac{A_{IP3}}{A_{cp}} = \sqrt{\frac{4}{3}} / \sqrt{0.145} = 9.64\text{dB} \tag{3.10}$$

Equation (3.10) indicates that the input-referred third-order intercept point is expected to be about 10dB higher than the 1-dB compression point. However, as mentioned before, the 1-dB compression point may be lower than that expected from Equation (3.5), taking higher-order effects into consideration. As a consequence, the difference between the 1-dB compression point and IIP3 may be higher than 10dB. For example, Fig. 3.1 shows that the amplifier has an IIP3 of about 12dB higher than its 1-dB compression point. Typically IIP3 is about 10-15 dB beyond the 1-dB compression point for amplifiers in current CMOS technologies [45].

3.2 CMOS LNA Architectures

The essential theory and practical considerations for the design of low-noise amplifiers and various architectures for practical implementations have been discussed in the literature [46]-[47]. The selection of the best LNA topology involves complex trade-offs between noise performance, power consumption, available gain, input matching, and

linearity. For a CMOS LNA, the common-gate input stage has the same noise sources as a common-source stage. However, the total performance is inferior. For this reason, we shall focus on the study of CMOS LNA topologies with a common-source input stage. Based on the noise model of MOS transistors established in Chapter 2, we will analyze the noise performance of various CMOS LNA's using the concept of noise figure.

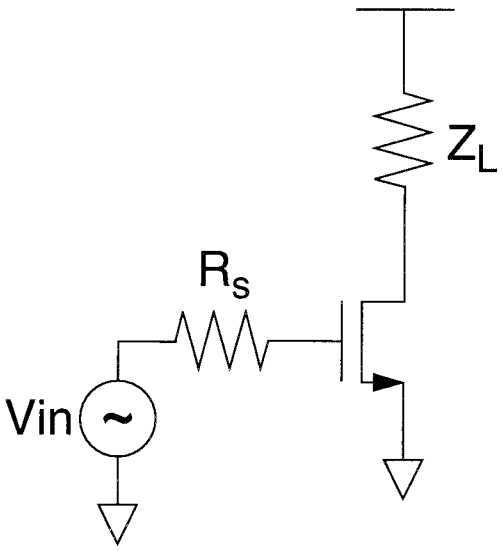
Please note that the noise performance of a CMOS LNA, besides being dependent on the amplifier, is also a function of the signal source impedance. A classical approach is to obtain the minimum noise figure from a given device by using the optimum source impedance [46], [48]. Though extensively used in discrete RF LNA designs, this approach does not offer guidance for the optimization of active devices.

3.2.1 Single-Transistor CMOS LNA

The simplest architecture of a CMOS LNA is the single-transistor implementation, as shown in Fig. 3.2. The small-signal noise equivalent circuit for the single-transistor CMOS LNA is developed based on the high-frequency noise model of MOS transistors. Note we neglect the gate-drain capacitance to simplify the analysis. In addition, since

$$g_g = \frac{\omega^2 C_{gs}^2}{5g_m} = \omega C_{gs} \cdot \frac{\omega}{5\omega_T}$$

is usually much smaller than ωC_{gs} for all practical frequencies, it is ignored here. The gate resistance R_g is a layout-related parameter and can always be reduced to a negligible value by a special gate structure, such as multi-finger gate as discussed in Sec. 2.4. It is also ignored here to further simplify the analysis.



$$\overline{i_g^2} = 4kT\beta g_g \cdot \Delta f \qquad g_g = \frac{1}{5}\frac{\omega^2 C_{gs}^2}{g_m}$$

$$\overline{i_s^2} = (4kT/R_s) \cdot \Delta f \qquad \overline{i_d^2} = 4kT\gamma g_m \cdot \Delta f$$

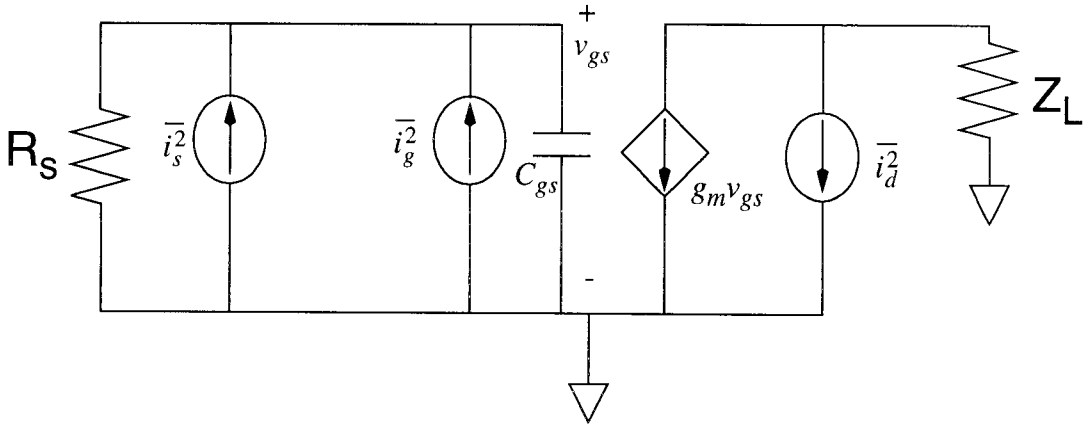


Figure 3.2: Single-transistor CMOS LNA.

The computation of noise figure is troublesome when using the definition directly. By simple arrangements of Equation (3.1), however, we can obtain a more useful formula for noise figure calculation, as shown below

$$\begin{aligned} NF &= 10\log\left(\frac{S_i/N_i}{S_o/N_o}\right) \\ &= 10\log\left(\frac{N_o}{N_i \cdot Gain}\right) \\ &= 10\log(F) \end{aligned} \tag{3.11}$$

where N_o is the total output noise into load Z_L and N_i is the input noise, which is the thermal noise associated with the source impedance R_s, modeled by a noise current source, $\overline{i_s^2} = (4kT/R_s) \cdot \Delta f$. Neglecting the noise contribution from the load impedance Z_L and assuming the correlation coefficient c between the induced gate current noise and the channel thermal noise is purely imaginary, we can derive the total output noise current N_o using KCL/KVL

$$\begin{aligned} N_o &= \overline{i_d^2} + (\overline{i_s^2} + \overline{i_g^2})\frac{g_m^2 R_s^2}{1+Q^{-2}} + \frac{2|c|g_m R_s}{Q+Q^{-1}}\sqrt{\overline{i_d^2} \cdot \overline{i_g^2}} \\ &= 4kT\Delta f\left[\gamma g_m + \left(\frac{1}{R_s} + \beta g_g\right)\frac{g_m^2 R_s^2}{1+Q^{-2}} + \frac{2|c|g_m R_s}{Q+Q^{-1}}\sqrt{\gamma g_m \beta g_g}\right] \\ &= 4kT\Delta f\left[\gamma g_m + \frac{g_m^2 R_s}{1+Q^{-2}} + \frac{\beta g_m}{5}\frac{1}{1+Q^2} + \frac{2|c|g_m}{1+Q^2}\sqrt{\frac{\gamma\beta}{5}}\right] \end{aligned} \tag{3.12}$$

where Q is the quality factor of the input capacitance C_{gs}, given by

$$Q = \frac{1}{\omega R_s C_{gs}} \tag{3.13}$$

The output noise current due to the source impedance R_s $(N_i \cdot Gain)$ is obtained as

$$N_i \cdot Gain = \overline{i_s^2}\frac{g_m^2 R_s^2}{1+Q^{-2}} = 4kT\frac{g_m^2 R_s}{1+Q^{-2}}\Delta f \tag{3.14}$$

Combining Equations (3.11)-(3.14), the noise factor F can be obtained as

$$F = \frac{N_o}{N_i \cdot Gain}$$
$$= 1 + \frac{\gamma}{R_s} \cdot \frac{1 + Q^{-2}}{g_m} + \frac{\beta}{5R_s} \cdot \frac{Q^{-2}}{g_m} + 2|c|\sqrt{\frac{\gamma\beta}{5}} \cdot \frac{Q^{-2}}{R_s g_m} \qquad (3.15)$$

Some conclusions can be drawn from Equation (3.15). It is clear that the noise factor F is dependent on the source impedance R_s, transistor transconductance g_m, and Q which is determined by R_s, capacitance C_{gs}, and signal frequency ω. Circuit designers usually have little control over parameters γ, β, and c since they are primarily technology-dependent (γ and β actually depend on biasing conditions. Unfortunately, we have little knowledge about this dependency at the present time). For a given R_s, the effective way to reduce the noise factor is to maximize g_m by increasing either the bias current I_D or W/L of the transistor. However, because g_m is proportional to $\sqrt{I_D \cdot W/L}$, there is no advantage in increasing I_D beyond a value dictated by other considerations such as power consumption. In addition, a large I_D may cause excessive heat dissipation which reduces the effective g_m and increases the noise temperature of the transistor. Increasing W/L, on the other hand, may actually degrade the noise performance due to the corresponding increase of C_{gs} which leads to a reduction in Q. It is also worthy of mention that a higher signal frequency will result in higher noise factor because of the degradation of Q.

The input impedance of the single-transistor CMOS LNA, ignoring the gate conductance g_g, is purely capacitive, given by

$$Z_{in} = r_g + \frac{1}{j\omega C_{gs}} \approx \frac{1}{j\omega C_{gs}} \qquad (3.16)$$

This mismatch to the source impedance R_s (typically 50Ω) will cause a large reflection from the LNA.

3.2.2 LC Tuned CMOS LNA

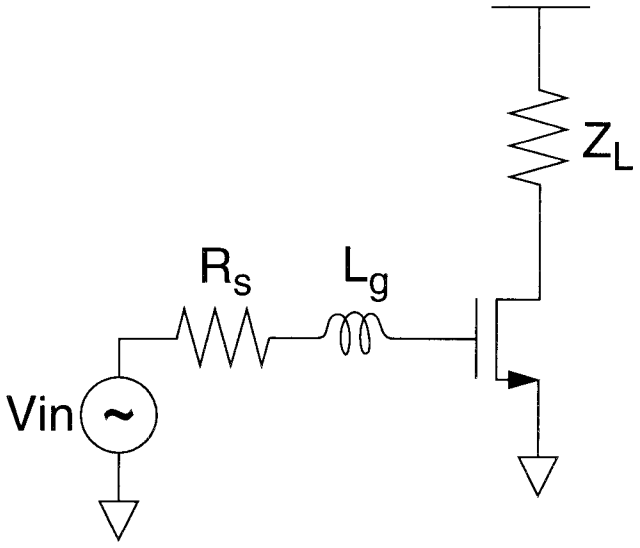
In order to improve the noise performance of a MOS transistor, we can reduce the device noise contribution by increasing the transistor conductance g_m. However, as discussed in the foregoing subsection, this approach requires higher power consumption and usually offers a limited improvement constrained by the technology. The definition of noise figure (Equation (3.1)) implies that a lower noise figure can be achieved if the noise contribution from the input noise source becomes a larger portion in the total output noise, even if the device noise contribution is not decreased. Since the input noise source is usually given (typically 50Ω), better noise performance is often achieved by using an input LC series resonant network to boost the input noise power at the gate of the MOS transistor (the input signal power gets boosted too) without adversely affecting the device noise.

Fig. 3.3 illustrates the first-order analysis of the resulting LC tuned amplifier. A series inductor is inserted between the signal source and the transistor. The inductor is modeled by an inductance L_g and its parasitic series resistance R_L (this model is taken for simplification. A more complicated inductor model will be discussed in the next chapter). By adding the inductor L_g in series with the gate capacitance C_{gs} to form a series resonant network, the total output noise current N_o is now given by

$$N_o = 4kT\Delta f\left[\gamma g_m + \frac{g_m^2 R' + \beta g_g g_m^2 (R'^2 + \omega^2 L_g^2)}{(1-\omega^2 L_g C_{gs})^2 + Q^{-2}} + 2|c|\frac{g_m R' Q^{-1} - g_m \omega L_g (1-\omega^2 L_g C_{gs})}{(1-\omega^2 L_g C_{gs})^2 + Q^{-2}}\sqrt{\gamma g_m \beta g_g}\right] \tag{3.17}$$

Due to the parasitic series resistance R_L, the quality factor Q of the input capacitance C_{gs} is degraded and given by

$$Q = \frac{1}{\omega R' C_{gs}} \tag{3.18}$$



$$\overline{i_g^2} = 4kT\beta g_g \cdot \Delta f \qquad g_g = \frac{1}{5}\frac{\omega^2 C_{gs}^2}{g_m}$$

$$\overline{i_s^2} = (4kT/R_s) \cdot \Delta f \qquad \overline{i_L^2} = (4kT/R_L) \cdot \Delta f \qquad \overline{i_d^2} = 4kT\gamma g_m \cdot \Delta f$$

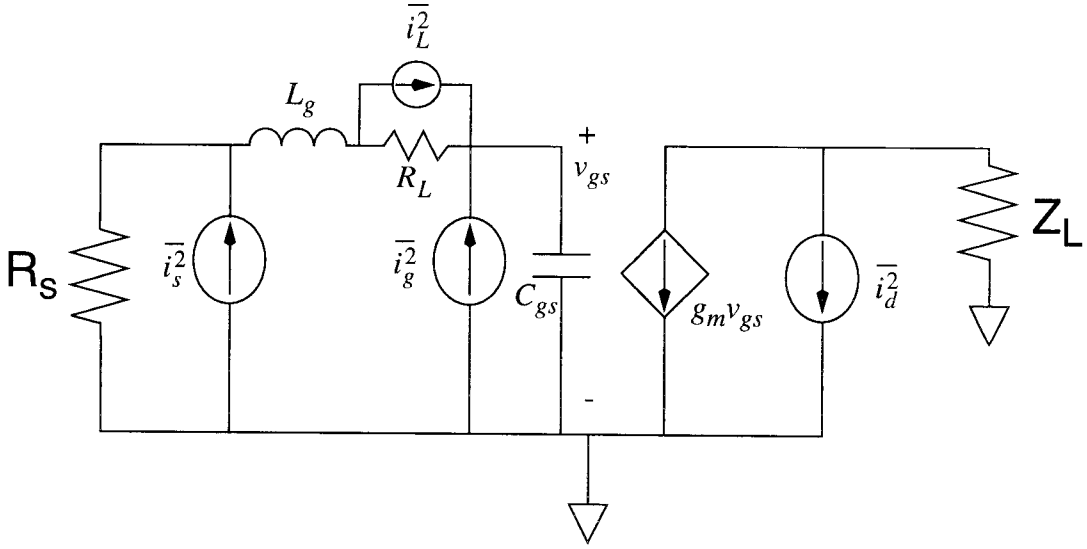


Figure 3.3: LC tuned CMOS LNA.

where $R'=R_s+R_L$. The output noise current due to the source impedance R_s $(N_i \cdot Gain)$ is obtained as

$$N_i \cdot Gain = 4kT \frac{g_m^2 R_s}{(1-\omega^2 L_g C_{gs})^2 + Q^{-2}} \Delta f \tag{3.19}$$

Combining Equations (3.17)-(3.19), the noise factor F is then obtained

$$\begin{aligned} F &= \frac{N_o}{N_i \cdot Gain} \\ &= 1 + \frac{R_L}{R_s} + \frac{\gamma}{R_s} \cdot \frac{(1-\omega^2 L_g C_{gs})^2 + Q^{-2}}{g_m} + \frac{\beta}{5R_s} \cdot \frac{\omega^4 L_g^2 C_{gs}^2 + Q^{-2}}{g_m} \\ &\quad +2|c|\sqrt{\frac{\gamma\beta}{5}} \cdot \frac{Q^{-2} - \omega^2 L_g C_{gs}(1-\omega^2 L_g C_{gs})}{R_s g_m} \end{aligned} \tag{3.20}$$

Equation (3.20) may look complicated but it provides guidance on how to select the optimal inductance L_g for a minimum noise factor. If the induced gate noise current is negligible, then it is obvious that the noise factor is minimized by selecting L_g so that $1-\omega^2 L_g C_{gs} = 0$ at the frequency of interest. The presence of gate noise current makes the selection of L_g a bit difficult. After a simple rearrangement and collection of terms in Equation (3.20), we obtain the following terms which are related to inductance L_g

$$\frac{\gamma}{R_s} \cdot \frac{(1-\omega^2 L_g C_{gs})^2}{g_m} + \frac{\beta}{5R_s} \cdot \frac{\omega^4 L_g^2 C_{gs}^2}{g_m} - 2|c|\sqrt{\frac{\gamma\beta}{5}} \cdot \frac{\omega^2 L_g C_{gs}(1-\omega^2 L_g C_{gs})}{R_s g_m}$$

Our goal is to make the above expression minimum (so also the noise factor) at the frequency of interest by properly selecting the inductance L_g. By setting the first derivative with respect to the inductance L_g to zero, it shows that the above expression is minimum and equal to

$$\frac{(1-|c|^2)\gamma\beta/5}{\gamma + \beta/5 + 2|c|\sqrt{\gamma\beta/5}} \cdot \frac{1}{R_s g_m}$$

when

$$\omega^2 L_g C_{gs} = \frac{\gamma + |c|\sqrt{\gamma\beta/5}}{\gamma + \beta/5 + 2|c|\sqrt{\gamma\beta/5}} \tag{3.21}$$

Now the minimum noise factor is obtained

$$F = 1 + \frac{R_L}{R_s} + \frac{\gamma}{R_s}\cdot\frac{Q^{-2}}{g_m} + \frac{\beta}{5R_s}\cdot\frac{Q^{-2}}{g_m} + 2|c|\sqrt{\frac{\gamma\beta}{5}}\cdot\frac{Q^{-2}}{R_s g_m} + \frac{(1-|c|^2)\gamma\beta/5}{\gamma + \beta/5 + 2|c|\sqrt{\gamma\beta/5}}\cdot\frac{1}{R_s g_m} \tag{3.22}$$

Compared to that of the single-transistor CMOS LNA, the minimum noise factor of the LC tuned CMOS LNA is lowered by

$$\frac{(\gamma + |c|\sqrt{\gamma\beta/5})^2}{\gamma + \beta/5 + 2|c|\sqrt{\gamma\beta/5}}\cdot\frac{1}{R_s g_m} - \frac{R_L}{R_s} \tag{3.23}$$

The parasitic series resistance R_L in the inductor introduces additional thermal noise which degrades the circuit noise performance. It increases the noise factor by R_L/R_s as indicated in Equation (3.22). Generally speaking, a large on-chip inductance is not desirable not only because of its low self-resonant frequency, but also because of its large parasitic resistance and other shunt parasitics which increase signal loss and generate excessive thermal noise. It is clear from expression (3.23) that the LC tuned CMOS LNA achieves better noise performance than the single-transistor CMOS LNA only if

$$R_L < \frac{1}{g_m}\cdot\frac{(\gamma + |c|\sqrt{\gamma\beta/5})^2}{\gamma + \beta/5 + 2|c|\sqrt{\gamma\beta/5}} \tag{3.24}$$

Taking $|c| = 0.35$ (Equation (2.10)), for long-channel devices, in which γ=2/3 and β=4/3, R_L should be less than about $0.54/g_m$. A lower R_L results in a more significant improvement in noise performance for the LC tuned CMOS LNA. Therefore, the quality factor of the inductance L_g is critical for the noise performance of the LC tuned CMOS LNA. For a given inductance, less parasitics are desirable.

Now we shall turn to the input impedance of the LC tuned CMOS LNA. Since the capacitance C_{gs} is partly tuned out by the inductance L_g, the gate conductance g_g may not be negligible. To take g_g into consideration, recall that the parallel network of g_g and C_{gs} can be converted to the series network of r_g and C_{gs}, where r_g=1/5g_m. The input impedance of the LC tuned CMOS LNA is then given by

$$\begin{aligned} Z_{in} &= R_L + r_g + j\omega L_g + \frac{1}{j\omega C_{gs}} \\ &= R_L + \frac{1}{5g_m} + \frac{\beta/5 + |c|\sqrt{\gamma\beta/5}}{\gamma + \beta/5 + 2|c|\sqrt{\gamma\beta/5}} \cdot \frac{1}{j\omega C_{gs}} \end{aligned} \tag{3.25}$$

Note that we can get rid of the capacitive term in the above equation by making $1 - \omega^2 L_g C_{gs} = 0$. However, the noise factor is degraded a bit and is given by

$$\begin{aligned} F = 1 + \frac{R_L}{R_s} + \frac{\gamma}{R_s} \cdot \frac{Q^{-2}}{g_m} + \frac{\beta}{5R_s} \cdot \frac{Q^{-2}+1}{g_m} \\ + 2|c|\sqrt{\frac{\gamma\beta}{5}} \cdot \frac{Q^{-2}}{R_s g_m} \end{aligned} \tag{3.26}$$

3.2.3 Inductive Source Degeneration

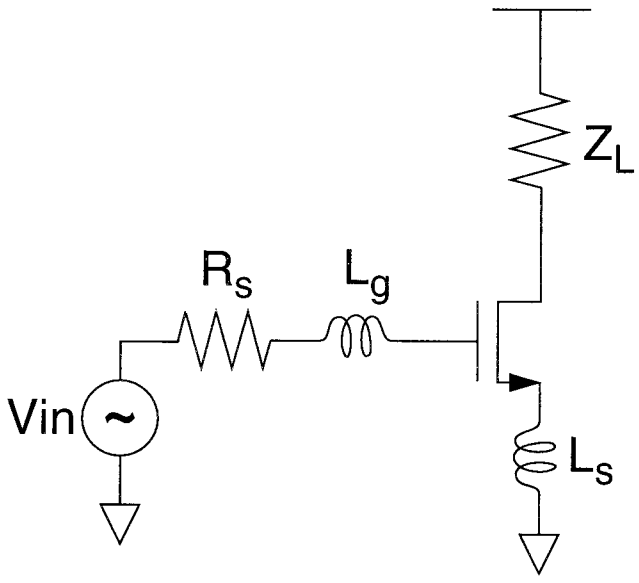
The input impedance matching of a CMOS LNA is a somewhat confusing issue. Traditional LNA designs usually utilize conjugate matching between the LNA and the signal source to achieve a maximum input power. However, for the CMOS LNA in which the output power is determined by the voltage across the input gate capacitance, the conjugate matching does not guarantee a maximum output power. From the noise point of view, as we discussed before, best noise performance is achieved while the input impedance has a capacitive term (Equation (3.25)). For these reasons, we may expect that the input impedance matching is not as useful for a CMOS LNA as it is for traditional designs (e.g., GaAs and BJT implementations), because it does not provide the maximum power gain while degrading the noise performance. This is not quite true, however. As a

matter of fact, the LNA's dominant in modern systems, even in CMOS technology, are designed to have the input impedance matched to the source impedance, which is typically a resistance of 50Ω. One possible reason for this is that the bandpass filter following the antenna (e.g., the duplexer) is usually implemented in a doubly terminated structure, which requires the same source and load impedance. If its load impedance (the input impedance of the LNA) deviates significantly from its source impedance (50Ω), the bandpass filter's characteristics may exhibit considerable loss and ripple [49].

To obtain an input impedance of 50Ω for the CMOS LNA, an inductive source degeneration may be used [50]-[51], [10]-[11]. The modified LC tuned CMOS LNA is shown in Fig. 3.4, in which an inductance L_s is added between the ground and the source of the MOS transistor. This series feedback inductance L_s contributes a noiseless resistive part to the input impedance of the CMOS LNA. It is preferred to the resistive feedback found in wideband amplifiers for impedance matching, because unlike feedback resistors, the inductor L_s does not degrade the noise performance if its parasitics are negligible. It is not difficult to show the input impedance of the CMOS LNA has the following form

$$\begin{aligned} Z_{in} &= R_L + r_g + j\omega L_g + \frac{1}{j\omega C_{gs}} + \left(1 + \frac{g_m}{j\omega C_{gs}} + g_m r_g\right) \cdot j\omega L_s \\ &= R_L + \frac{1}{5g_m} + \frac{g_m L_s}{C_{gs}} + j\omega(L_g + 1.2L_s) + \frac{1}{j\omega C_{gs}} \end{aligned} \tag{3.27}$$

Here we ignore the parasitic series resistance associated with L_s to simplify the analysis. Except for the additional thermal noise that is brought with it, the parasitic series resistance with L_s also contributes a real term and a capacitive term to the input impedance of the CMOS LNA.



$$\overline{i_g^2} = 4kT\beta g_g \cdot \Delta f \qquad g_g = \frac{1}{5}\frac{\omega^2 C_{gs}^2}{g_m}$$

$$\overline{i_s^2} = (4kT/R_s) \cdot \Delta f \qquad \overline{i_L^2} = (4kT/R_L) \cdot \Delta f \qquad \overline{i_d^2} = 4kT\gamma g_m \cdot \Delta f$$

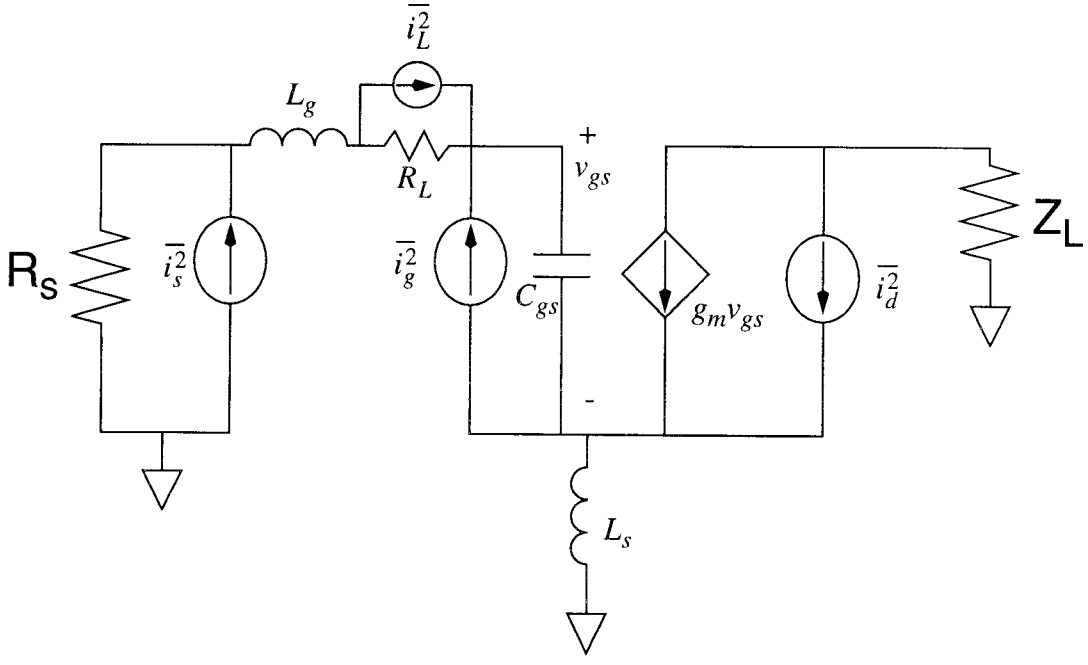


Figure 3.4: LC tuned CMOS LNA with inductive source degeneration.

From Equation (3.27), it is clear that two criteria for input impedance matching are:

$$R_L + \frac{1}{5g_m} + \frac{g_m L_s}{C_{gs}} = R_s = 50\Omega \tag{3.28}$$

and

$$j\omega(L_g + 1.2L_s) + \frac{1}{j\omega C_{gs}} = 0 \tag{3.29}$$

Clearly the input impedance can be matched to the source impedance only at one frequency.

Assuming Equations (3.28) and (3.29) are satisfied by carefully selecting L_g and L_s to obtain the matched input impedance, the total output noise current N_o is now given by

$$N_o = 4kT\Delta f\left[\gamma g_m \frac{R'^2}{4R_s^2} + \frac{g_m^2 R' + \beta g_g g_m^2 [R'^2 + 1/(\omega^2 C_{gs}^2)]}{4R_s^2\omega^2 C_{gs}^2} + 2|c|\frac{g_m R'^2\sqrt{\gamma g_m \beta g_g}}{4R_s^2\omega C_{gs}}\right] \tag{3.30}$$

where $R'=R_s+R_L$. The output noise current due to the source impedance R_s $(N_i \cdot Gain)$ is obtained as

$$N_i \cdot Gain = 4kT\frac{g_m^2 R_s}{4R_s^2\omega^2 C_{gs}^2}\Delta f \tag{3.31}$$

The noise factor F of this source-degenerated CMOS LNA can then be obtained as

$$\begin{aligned} F &= \frac{N_o}{N_i \cdot Gain} \\ &= 1 + \frac{R_L}{R_s} + \frac{\gamma}{R_s}\cdot\frac{Q^{-2}}{g_m} + \frac{\beta}{5R_s}\cdot\frac{1+Q^{-2}}{g_m} \\ &\quad +2|c|\sqrt{\frac{\gamma\beta}{5}}\cdot\frac{Q^{-2}}{R_s g_m} \end{aligned} \tag{3.32}$$

where

$$Q = \frac{1}{\omega R' C_{gs}}$$

Compared to Equation (3.22), the noise performance degrades only slightly.

The major downside of the inductive source degeneration, however, is the degradation of the amplifier gain. Neglecting the parasitic resistance R_L and the gate conductance g_g, the effective transconductance G_m of the LC tuned CMOS LNA can be expressed as $G_m = g_m \cdot Q$ without the source degeneration, and $G_m = g_m \cdot Q/2$ with the source degeneration. Therefore, the source degeneration results in approximately a 6dB loss in the power (or voltage) gain of the LC tuned CMOS LNA.

3.3 Design Considerations of CMOS LNA

3.3.1 Optimization of Device Parameters

The analysis of the previous sections can now be drawn upon in designing the CMOS LNA. Our goal here is to develop optimization techniques for CMOS LNA design. Particularly, the optimization of the device parameters for minimum noise factor shall be discussed.

To make things easier, we re-write the noise factor formula (3.22) in a simplified version

$$\begin{aligned} F &= 1 + \frac{R_L}{R_s} + \gamma A \cdot \frac{Q^{-2}}{R_s g_m} + \frac{(1-|c|^2)\beta/5}{A} \cdot \frac{1}{R_s g_m} \\ &= 1 + \frac{R_L}{R_s} + \frac{\gamma F_d}{R_s} \end{aligned} \tag{3.33}$$

where

$$F_d = A \cdot \frac{Q^{-2}}{g_m} + \frac{(1-|c|^2)\beta/(5\gamma)}{A} \cdot \frac{1}{g_m} \tag{3.34}$$

and $A = 1 + \beta/(5\gamma) + 2|c|\sqrt{\beta/(5\gamma)}$.

From the noise factor expression (3.33), it is clear that only the last term F_d depends on device parameters. Clearly a large bias current is desired for a low noise factor because the transconductance g_m is proportional to the square root of bias current. However, the LNA usually suffers from a power consumption constraint, which puts a limit on the available bias current. Given the maximum bias current I_D, it is still possible to improve the noise factor by optimizing other device parameters, such as the device width and length.

Recall

$$g_m = \sqrt{2\mu C_{ox}(W/L)I_D} \tag{3.35}$$

We adopt the long-channel formula to simplify the analysis. Also Q can be rewritten as

$$Q = \frac{1}{\omega R' C_{gs}} = \frac{3}{2\omega R'(C_{ox} \cdot WL)} \tag{3.36}$$

assuming $C_{gs} = \frac{2}{3}(C_{ox} \cdot WL)$. Substituting Equations (3.35) and (3.36) into (3.34), F_D can then be rewritten as

$$F_d = A \cdot \frac{(2\omega R' C_{ox}/3)^2}{\sqrt{2\mu C_{ox} I_D}} \cdot W^{1.5}L^{2.5} + \frac{(1-|c|^2)\beta/(5\gamma)}{A\sqrt{2\mu C_{ox} I_D}} \cdot W^{-0.5}L^{0.5} \tag{3.37}$$

It is clear that minimum channel length L should be used for minimum F_d and thus minimum noise factor F. To obtain the optimized device width W, take the first derivative of expression (3.37) with respect to W and set it to zero. After making tedious simplifications, we finally obtain an expression for the width of the optimum device:

$$W_{opt} = \frac{\sqrt{3(1-|c|^2)\beta/(5\gamma)}}{2A} \cdot \frac{1}{\omega R' C_{ox} L} \tag{3.38}$$

For a long-channel device, γ=2/3, β=4/3. As we know, for a short-channel device, γ and β may be much higher and dependent on the bias condition. However, we may assume that the ratio β/γ remains roughly constant regardless of the shrinkage of the channel length or the variation of biasing. Taking $|c| = 0.35$ (Equation (2.10)), then Equation (3.38) can be further simplified as

$$W_{opt} \approx \frac{1}{3\omega R' C_{ox} L} \tag{3.39}$$

This implies

$$Q_{opt} = \frac{3}{2\omega R'(C_{ox} \cdot W_{opt} L)} \approx 4.5 \tag{3.40}$$

The optimized noise factor can then be expressed as

$$F_{min} \approx 1 + \frac{R_L}{R_s} + \frac{0.3\gamma}{g_m R_s} \tag{3.41}$$

or

$$F_{min} \approx 1 + \frac{R_L}{R_s} + 1.3\gamma\left(1 + \frac{R_L}{R_s}\right)\left(\frac{\omega}{\omega_T}\right) \tag{3.42}$$

or

$$F_{min} \approx 1 + \frac{R_L}{R_s} + \frac{0.4\gamma\sqrt{R_s + R_L}}{R_s}\sqrt{\frac{\omega}{\mu I_D}} \cdot L \tag{3.43}$$

Equations (3.41) and (3.42) show the relationship between the minimum noise factor theoretically achievable and the transistor g_m and ω_T, respectively. Note that g_m and ω_T are related to the optimal device width determined using Equation (3.39). In this sense, Equations (3.41) and (3.42) may be misleading if not carefully referred. For example, given all device parameters except the width, one may argue based on Equation (3.42) that increasing ω_T by reducing the device width would result in better noise performance, which is, of course, not true. Equation (3.43), however, is probably most useful for CMOS LNA designs in that it shows clearly the dependence of the minimum noise factor on the

basic device parameters, namely, the bias current I_D, the channel length L, and the signal frequency ω. Based on Equation (3.43), basic conclusions can be drawn for CMOS LNA design for a minimum noise factor:

i) The noise factor decreases with the shrinkage of channel length. Therefore, a minimum channel length should be used in a given technology. It is also expected that as the current CMOS technology continues scaling down, the noise performance of a CMOS LNA can be further improved and eventually will be limited only by parasitic effects associated with the passive components, interconnects, or packaging.

ii) The noise factor decreases with an increase in the bias current I_D (power consumption). Therefore, in order to achieve good noise performance, a CMOS LNA usually dissipates a large amount of power.

iii) Given the minimum channel length and the maximum bias current, the device width should be chosen using Equation (3.39) in order to achieve the minimum noise factor. Please note this optimum device width is frequency dependent. It is also a function of the parasitic series resistance R_L of the inductance L_g.

iv) Having chosen the device length and width, we can pick the inductance L_g based on Equation (3.21). If inductive source degeneration is employed, inductances L_g and L_s can be determined using Equations (3.28) and (3.29).

v) Because we must include the parasitic series resistance R_L even before determining L_g, a gradual refinement of the optimum device width and the inductance value is necessary. One may assume R_L in Ω is approximately equal to L_g in nH, for monolithic inductors in CMOS technology, as a starting point.

vi) The CMOS LNA discussed here has narrowband characteristics. The noise factor increases as the signal frequency increases. Given a technology and a power

consumption, a CMOS LNA designed for higher-frequency applications would have poorer noise performance.

The foregoing conclusions provide primitive guidelines for narrowband CMOS LNA design. However, the design should also be verified and further optimized with the aid of computer simulation. Obviously, accurately modeling of the monolithic inductors, besides the modeling for high-frequency MOS noise characteristics, would play a critical role.

3.3.2 Fully-Differential v/s Single-Ended

Most traditional LNA designs are single-ended possibly because the incoming RF signal from the antenna is single-ended in nature. However, fully-differential LNA designs have become more and more popular recently. As shown in Fig. 3.5, one important shortcoming of single-ended LNA architecture is that the ground parasitic impedance has significant effect upon the circuit performance. For example, even a small ground inductance will significantly change the input impedance of the amplifier and thus degrade its performance. In a fully-differential LNA architecture, however, the ground parasitics are largely irrelevant to circuit performance because of the existence of a virtual ground if differential signals are applied to the LNA. In addition, the single-ended LNA is sensitive to any undesired signal or noise coming from other circuitry within the same die. A fully-differential LNA, on the other hand, exhibits good common-mode rejection to such disturbances. This consideration is particularly important in state-of-the-art wireless system design, in which efforts are ongoing to integrate the whole transceiver circuitry in one single chip. This means that the LNA should work with circuitry containing largely mixed-signal function blocks, where both the power supply and substrate may introduce a large amount of undesired signal and noise. Another advantage providing by the fully-

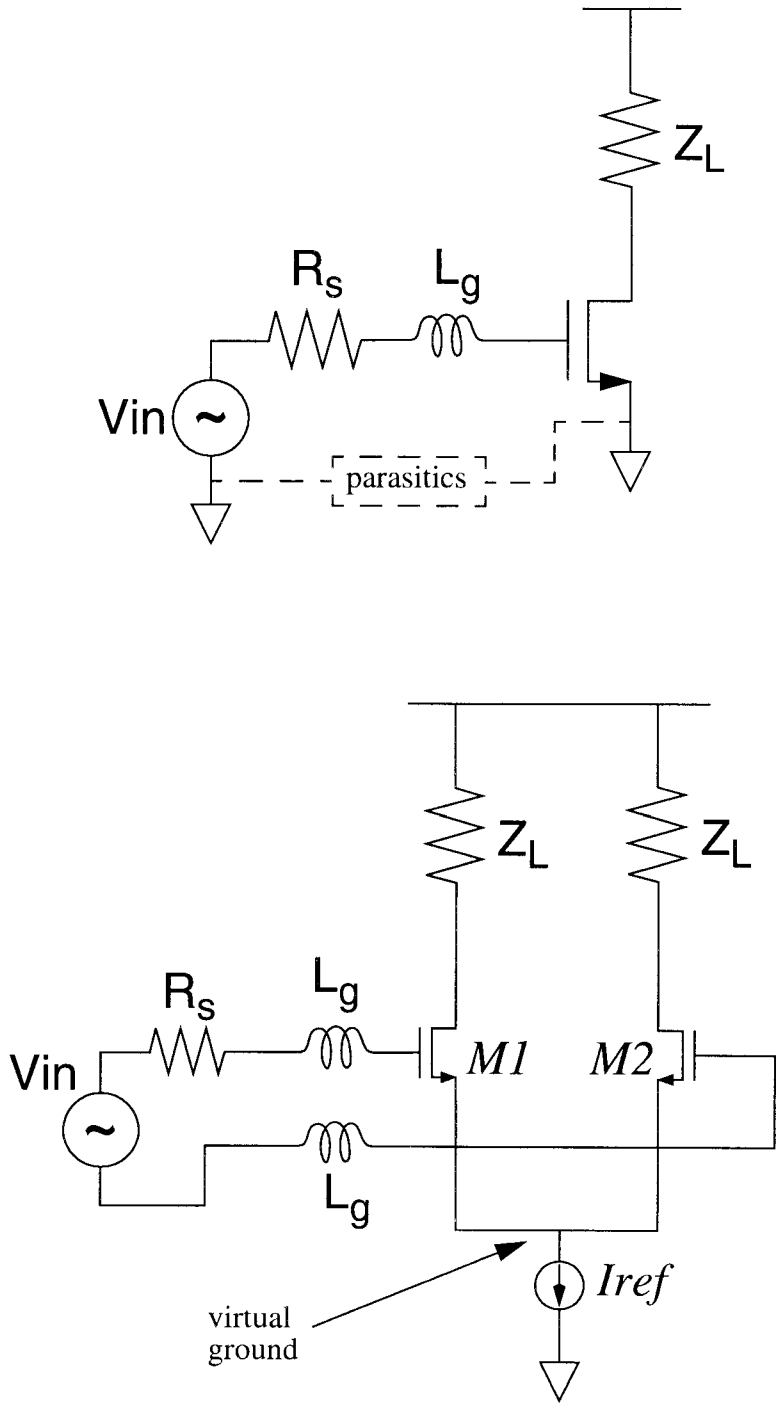


Figure 3.5: Single-ended and fully-differential CMOS LNA's.

differential LNA is that its differential output can be fed directly into the following doubly-balanced mixer, eliminating the need for the unbalanced-to-balanced conversion between them.

The fully-differential LNA has several drawbacks too. The power consumed is twice that of a single-ended counterpart in order to achieve the same g_m or ω_T. Even so, the noise performance is still worse because the device noise contribution is roughly double that in a single-ended LNA. For example, if a single-ended CMOS LNA has a noise figure of 2.5dB for a given power consumption, the fully-differential CMOS LNA would only achieve approximately 4dB noise figure even with twice the power consumption.

CHAPTER 4. SILICON-BASED MONOLITHIC INDUCTORS

Passive devices often determine the overall size, topology, and performance of RF circuits. As the size and cost of active devices continue to shrink, improvements in passive device performance become more urgent. As shown in the previous chapters, monolithic inductors are the key components in the realization of a high-performance CMOS LNA. The quality factors of these inductors determine the performance of the LC tuning circuits and thus the overall LNA circuit performance.

Much progress towards the integration of high quality silicon-based inductors has been reported [52]-[63]. Although many innovative structures and design techniques have been proposed [59]-[63], most monolithic inductors have achieved only moderate quality. The basic problem is that since only planar structures are practical in fine-line digital CMOS technologies, long metal traces, with unavoidable high resistive losses, are required. In addition, a monolithic inductor usually consumes a large die area so that significant losses in the conductive silicon substrate due to capacitive and magnetic coupling further degrade the performance.

In this chapter, a review of monolithic inductor implementations is given first, with emphasis on spiral inductors, the most widely used silicon-based monolithic inductors in RF IC's. Detailed analysis and modeling of the square spiral inductors are then described. The model developed can be used directly in an IC simulator, such as SPICE. Based on the circuit model, the inductor performance can be analyzed and formulated to facilitate hand analysis. In addition, some alternative designs of spiral inductors are also discussed.

4.1 Implementations of Monolithic Inductors

4.1.1 Active Inductors

A straightforward way to integrate an inductor is to realize the equivalent inductance using active elements [64]-[65], as shown in the general implementation in Fig. 4.1. The basic concept here is to convert a capacitive impedance to an inductive impedance using transconductors. From Fig. 4.1, the voltage-current relationship can be described as

$$\frac{V}{I} = \frac{I_1/g_{m1}}{-(-g_{m2}V_1)} = \frac{j\omega C}{g_{m1} \cdot g_{m2}} \tag{4.1}$$

Hence, the equivalent inductance is

$$L_{eq} = \frac{C}{g_{m1} \cdot g_{m2}} \tag{4.2}$$

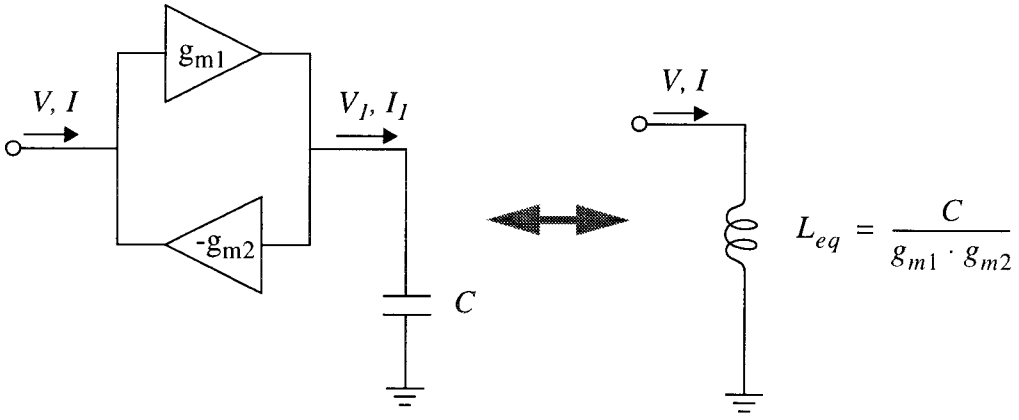


Figure 4.1: General implementation of an active inductor.

Active inductors are easily integrated on chip and their size is relatively small and independent of the inductance value. On the other hand, passive inductors occupy large die area and thus have large parasitics, especially when a large inductance is required. Therefore, active inductors can usually achieve larger inductance and higher self-resonant frequency. In addition, active inductors are nearly lossless and thus can achieve a high quality factor. Moreover, as shown in Equation (4.2), the active inductance is determined by transconductances and thus is electronically tunable. This feature is advantageous in many RF IC designs such as tunable oscillators.

Despite the advantages that active inductors may provide, they are not practical for LNA design due to their excessive noise contribution which is usually comparable to the total noise of an LNA [66]. This adverse feature makes the noise-reducing LC tuning network (discussed in Chapter 3) useless.

4.1.2 Bondwire Inductors

Bondwire inductors take advantage of the parasitic inductances associated with bondwires in an IC package. Due to the low series resistance of the bondwires, high quality factor inductors are achievable by careful design. Also since the parasitic capacitance to the substrate is reduced to just the capacitance of bondpads, a high self-resonant frequency can be realized.

The self- and mutual inductances of bondwire inductors can be calculated using inductance extraction simulators or the first-order formulas given in [67]. A simple estimation of the typical bondwire inductance is about 1nH/mm. A CMOS oscillator design using bondwire inductors is described in [68]. The very low series resistance of gold bondwires enables low phase noise and low power designs.

The bondwire inductance is determined by its length and spacing to other bondwires. Accurate control of these physical dimensions is difficult even in a modern IC technology. Due to fabrication variations and uncertainties, the bondwire inductance usually has a fairly large error from that theoretically expected, which makes the design unpredictable. Also the relatively low yield and reliability of the bonding process compared to chip fabrication processes increases the cost of an RF IC using bondwire inductors.

4.1.3 Spiral Inductors

Spiral inductors have been used extensively in microwave integrated circuits (MICs) and are usually deposited on a ceramic substrate (hybrid MICs) or a GaAs substrate (monolithic MICs). They are also the most widely used monolithic inductors in silicon integrated circuits. However, the properties of spiral inductors in silicon technology are much different from those in MIC processes because of the different metallization and the very lossy silicon substrate.

Silicon-based monolithic spiral inductors are implemented using one or more metal traces (usually aluminum or possibly gold or copper in an expensive process) in square spiral structures. As shown in Fig. 4.2, the simplest layout of a square spiral inductor consists of a series of spiral turns (four in Fig. 4.2) on the topmost metal layer (e.g., metal3 in a 3-metal digital CMOS process) to provide the lowest metal resistance and parasitic capacitance to the lossy substrate. Connection to the spiral center is made with vias and a cross-under of some lower metallization layer, e.g., metal2.

For a first-order approximation, we may ignore the presence of the lossy silicon substrate and calculate the inductance of the silicon-based spiral inductor in similar ways for spirals operating in free space as described in [67][69]. However, the performance of

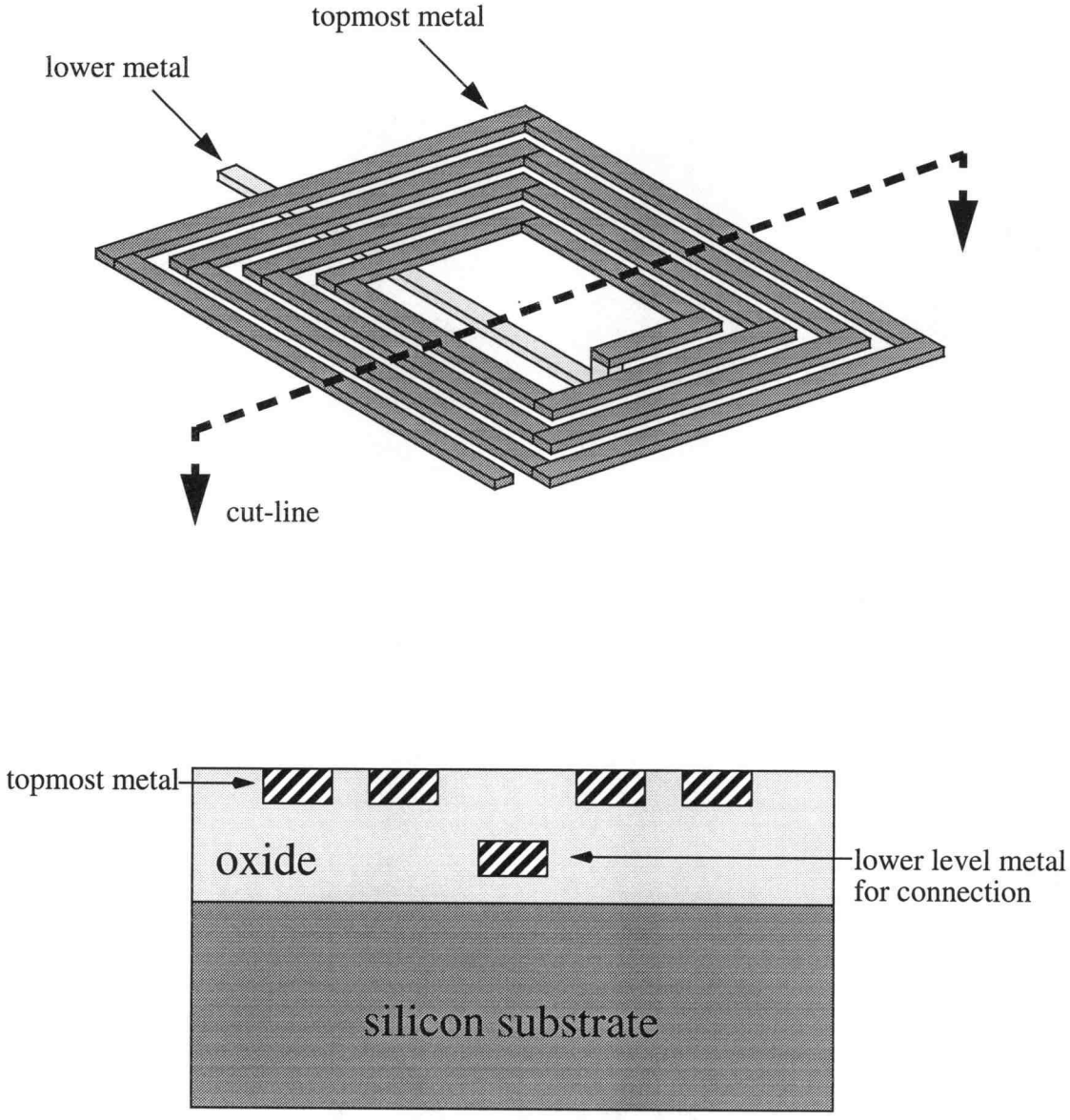


Figure 4.2: Layout and cross-section of a square spiral inductor.

practical silicon-based inductors is degraded significantly at radio frequencies by parasitic capacitances and resistive losses associated with the conductive substrate. Therefore, more accurate modeling of spiral inductors is required for the computer simulation and optimization of LC-tuned RF circuits.

4.2 Modeling of Spiral Inductors

The accurate modeling of silicon-based spiral inductors requires a complete analysis of the self- and mutual inductances and the parasitic resistances and capacitances. The most complicated and important effects for silicon technology, the capacitive and inductive coupling effects to the conductive Si substrate must also be included.

4.2.1 Scalable Circuit Model

To simplify the analysis, and more importantly, to develop a physical and scalable lumped-element circuit model, each segment (sixteen in Fig. 4.2) of the spiral inductor is treated as a microstrip line that can be represented by a traditional lumped-element π-model including all mutual coupling effects from the other segments [56]-[57]. A complete lumped-element equivalent circuit for a microstrip line is shown in Fig. 4.3. As shown, L represents the self-inductance of the microstrip line (a straight conductor) on a silicon substrate and the mutual inductance contributed from other microstrip lines in the spiral structure. R models the resistive loss associated with the microstrip line, including the metal resistive loss and the resistive losses caused by the magnetically induced eddy current in the heavily-doped Si substrate. The shunt capacitance C_{ox} models the oxide capacitance between the microstrip line and the substrate. In addition, the substrate parasitics are modeled by shunt capacitance C_{si} and resistance R_{si}.

The lumped-element circuit model for a silicon-based spiral inductor can then be formed by correctly connecting the equivalent circuits for all segments (i.e., microstrip lines) in the spiral structure. An illustrative example of such a scalable model is shown in Fig. 4.4. As we can see, the scalable circuit model for a 2-turn spiral inductor consists of 8 lumped-element subcircuits (I-VIII as indicated, but only 4 subcircuits I-IV are shown in Fig. 4.4), corresponding to 8 microstrip segments in the spiral structure. Please note that in addition to the mutual inductive coupling effects which are modeled in L, there are also mutual capacitive coupling effects between adjacent microstrip segments through the side-wall capacitance. These capacitive coupling effects are modeled by the lumped capacitors (C_c) between adjacent microstrip segments [57], as indicated in Fig. 4.4.

Since the properties of a microstrip line are much better known than those of a spiral inductor, the electrical parameters of the lumped-element circuit for a microstrip line can

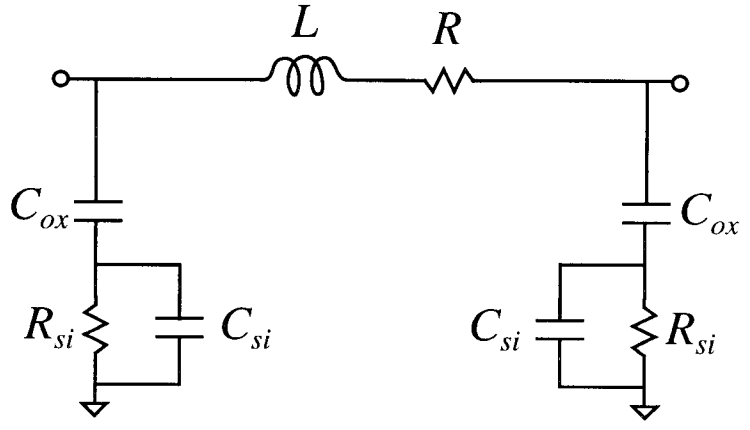


Figure 4.3: Lumped-element circuit model for a microstrip line.

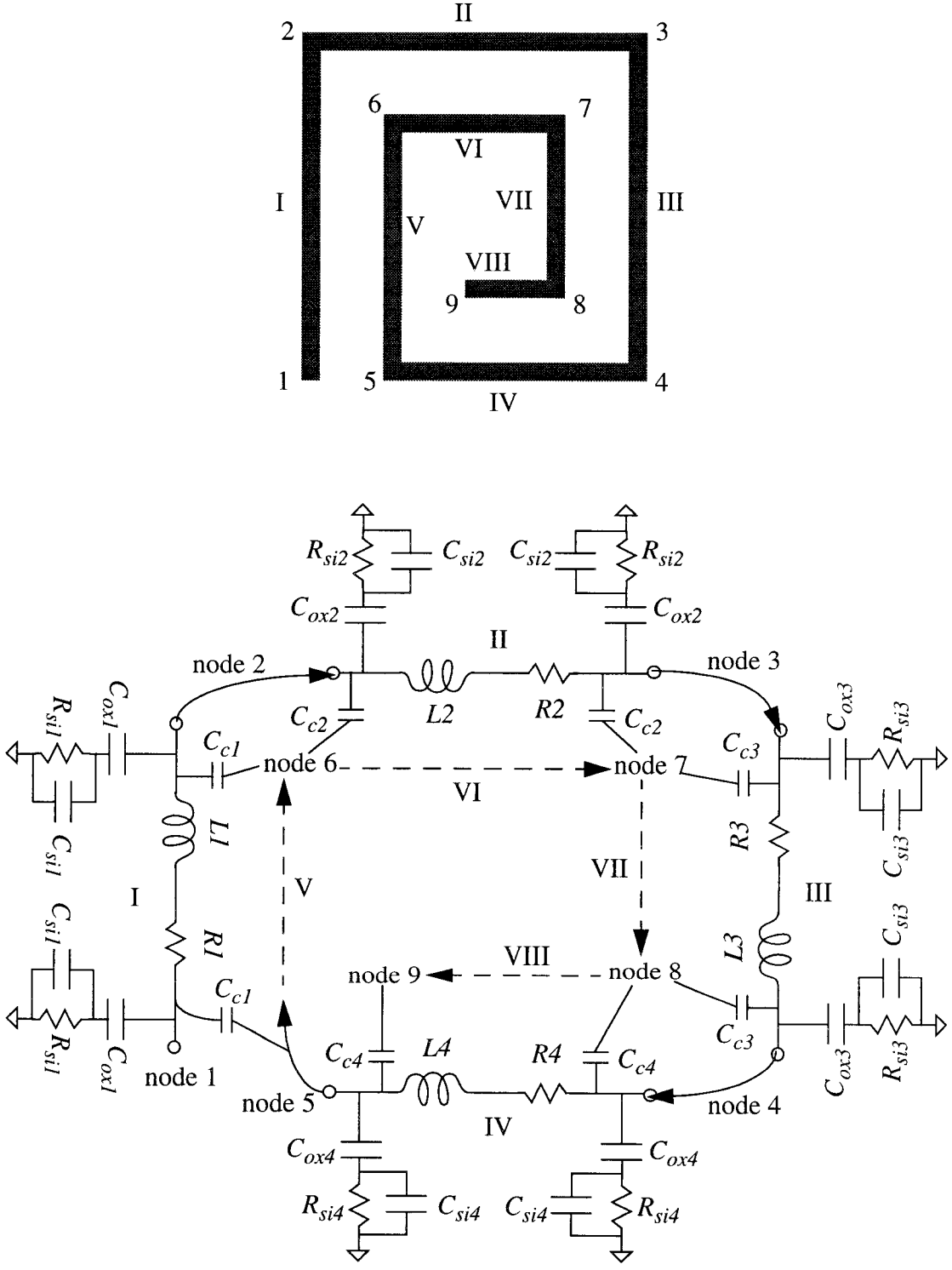


Figure 4.4: Scalable lumped-element circuit model for a spiral inductor.

be relatively easily determined by numerical analysis or even closed-form expressions (theoretical or empirical). Furthermore, since most electrical parameters have a physical meaning as discussed before, the lumped-element circuit model for the spiral inductor derived in this way can be scaled to reflect changes in dimensions or fabrication technology. This approach saves both development time and the cost associated with monolithic circuit design. Other approaches to modeling the silicon-based spiral inductors, such as parameter fitting of lumped-element equivalent circuits to the measured S-parameters of many fabricated spiral inductors [62][70], may obscure some of the circuit components, and the model derived is not scalable.

A detailed discussion of the properties of microstrip lines on a silicon substrate is given in the following subsections. An extraction procedure for all the electrical parameters in the scalable lumped-element circuit model for the spiral inductor is also described.

4.2.2 Self- and Mutual Inductance

Based on the work of Grover [69], Greenhouse developed a set of formulas to calculate the inductance of rectangular spiral inductors in free space [67]. For a microstrip segment in the spiral inductor, its self-inductance and the mutual inductance from all other microstrip segments can be calculated. Only the mutual coupling effects from all other parallel microstrip segments need to be calculated; the coupling from perpendicular segments is negligible.

Assuming that the width is much larger than the thickness, the self-inductance for a microstrip line is given by [67]

$$L_{self} = 0.02 \cdot l \cdot \left[\ln\left(\frac{2 \cdot l}{w+t}\right) + \frac{w+t}{3 \cdot l} + 0.50049 \right] \qquad (4.3)$$

where L_{self} is the self-inductance in nanohenries and l, w, and t are the length, width, and thickness of the microstrip line in micrometers, respectively. Note that the frequency dependence due to skin-effect is ignored in this analysis.

The mutual inductance M between two parallel microstrip lines is a function of the length of the microstrip lines and of the distance between them. The mutual inductance between the two parallel microstrip lines shown schematically in Fig. 4.5 is given by [67]

$$M = \frac{1}{2}(M_{(l+x)} + M_{(l+y)} - M_{(x)} - M_{(y)}) \tag{4.4}$$

and

$$M_{(l)} = 2 \times 10^{-4} \cdot l \cdot \left[\ln\left(\frac{l}{d} + \sqrt{1 + \frac{l^2}{d^2}}\right) - \sqrt{1 + \frac{d^2}{l^2}} + \frac{d}{l} \right] \tag{4.5}$$

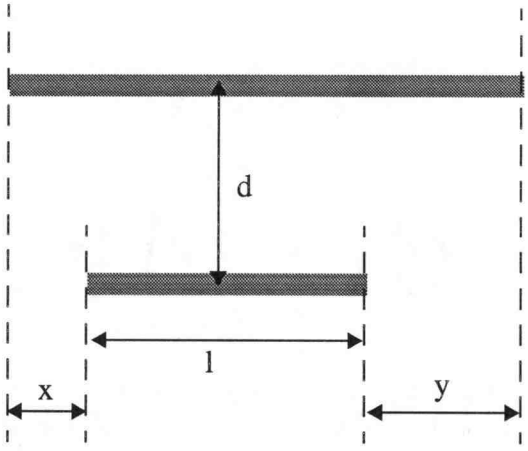


Figure 4.5: Two parallel microstrip lines.

where M is the mutual inductance in nanohenries and l, x, and y are the lengths in micrometers, and d is the distance in micrometers between the two line centers as indicated in Fig. 4.5. Note that the mutual inductance is positive when currents flow in two parallel microstrip lines in the same direction and negative when currents flow in opposite directions.

The total inductance L of a microstrip line equals its self-inductance plus the vector sum of all the mutual inductances. This method is accurate for the ideal case of the inductor in free space with no ground plane present [67]. However, because of the presence of a ground plane in a silicon-based technology, the mutual inductances from a mirror spiral under the ground plane must also be taken into account for accurate inductance calculations [71][72]. In addition, propagation delays around the spiral will cause phase differences between the currents in each segment. Both of these effects, not considered by Greenhouse, will lower the total inductance of the microstrip line [72].

Fig. 4.6 shows the two effects of the image spiral and the phase shift in current flow. The image spiral, mirrored by the ground plane and located at distance D from the actual spiral, contributes a net negative mutual inductance because the current flow is in the opposite direction in the return path. Because of the distributed nature of the spiral inductor, there is a phase shift in the current flow along the microstrip segments. At higher frequencies, the phase shift increases. As a consequence, for each microstrip line, the actual mutual inductance contributed from other microstrip segments is frequency-dependent. The mutual inductance adds progressively less to the total inductance as the frequency increases and eventually subtracts from the total inductance if the phase shift becomes more than 180° (although both positive and negative mutual inductance are affected by the phase shift problem, the net effect is a reduced mutual inductance).

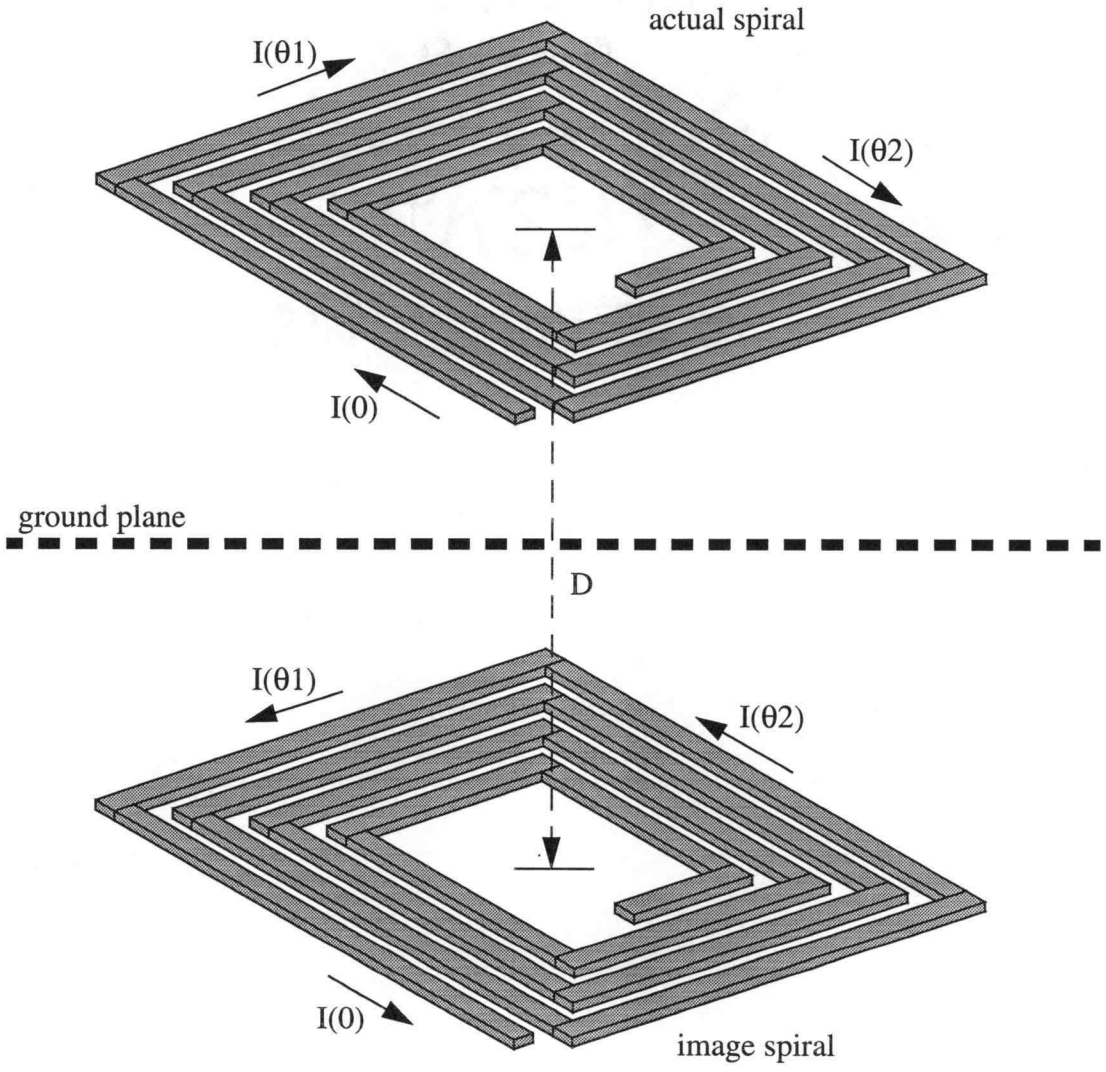


Figure 4.6: Effects of the image spiral and the phase shift in current.

Krafesik and Dawson proposed an improved method to calculate the inductance by accounting for the two effects [72]. The image spiral is treated the same as microstrip lines, contributing a mutual inductance which can be calculated using the Greenhouse formulas. Since usually the length of a segment is much smaller than the signal wavelength (otherwise we partition the long segment into several shorter ones), the phase shift along a segment can be lumped into a phasor, which is inserted between two connecting microstrip segments to account for the phase difference between them. The phasor can be computed using the lumped-element circuit model of a microstrip line once other lumped electrical parameters are determined.

The aforementioned method can accurately predict the total inductance and can be easily programmed. However, we adopt a three-dimensional inductance extraction program, *FastHenry*, developed at MIT [73] for the computation of the total inductance of a segment. For one segment in the spiral inductor, the self inductance L_{self} and the mutual coupling coefficients ($k_1 ... k_n$) to other segments can be easily calculated using *FastHenry*. The effect of mutual inductance contributed from other segments can then be modeled by dependent voltage sources as shown in Fig. 4.7. By doing so, the phase shift problem is now accounted for in the dependent voltage sources ($k_1 V_1 ... k_n V_n$), where $V_1 ... V_n$ are the voltages across the self-inductances of other microstrip segments, which would exactly reflect the current flow phase. When imported to a circuit simulator such as SPICE, this circuit would accurately model the total effective inductance of the microstrip segment, taking into consideration the current phase shift along the spiral. In addition, the image spiral effect due to the presence of a ground plane can also be easily computed using *FastHenry*. However, for a highly conductive substrate as in modern digital CMOS technology, it is difficult to determine the effective ground plane. To precisely predict the effects of image spiral and eddy current on the total effective inductance, we should know the exact location of the effective ground plane. Fortunately, these effects are relatively

small. Hence, for simplicity, we can consider the effective ground plane in the middle of the conductive substrate, without a great loss of accuracy in inductance calculations.

4.2.3 Resistive Losses

The resistive losses associated with the microstrip line are caused by the metal trace resistance and the magnetically induced eddy current in the heavily-doped silicon substrate.

The metal resistance is approximately constant at low frequencies and can be estimated using the sheet resistance data given in the CMOS process specifications. At high frequencies, however, the metal resistance becomes frequency-dependent due to the skin effect. The exact calculation of the frequency-dependent resistance of a metallic conductor

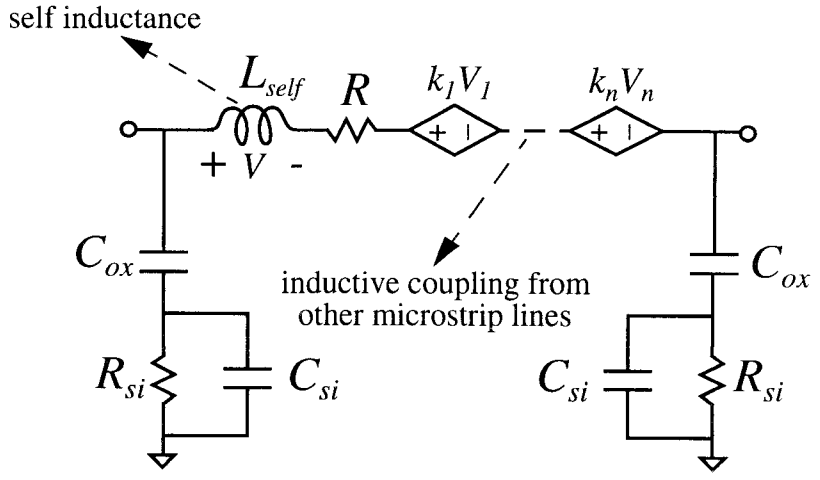


Figure 4.7: Modified lumped-element circuit model for a microstrip line.

with rectangular cross section is complicated and is usually done using numerical methods. However, empirical closed-form expressions have been developed by fitting to measurement results [74]. For a rectangular metal trace, assuming l is the length, w the line width, and t the metallization thickness, the frequency-dependent resistance can be expressed as

$$R = \frac{l}{\sigma w t}(1 + 0.0122X^{(3+0.01X^2)}) \tag{4.6}$$

for $X<2.5$; and

$$R = \frac{l}{\sigma w t}\left[\frac{0.43093X}{1+0.041\left(\frac{w}{t}\right)^{1.19}} + \frac{1.1147 + 1.2868X}{1.2296 + 1.287X^3} + 0.0035\left(\frac{w}{t} - 1\right)^{1.8}\right] \tag{4.7}$$

for $X \geq 2.5$; where

$$X = \sqrt{2f\sigma\mu wt}$$

In the above expressions, σ and μ stand for the conductivity and the permeability of the metal, with typical values of 2.4×10^7 S/m and $4\pi\times10^{-7}$ H/m for aluminum in CMOS, respectively. These formulas describe the metal resistance with an accuracy within 5% in the range $w/t < 12$ and $X < 20$ [74]. However, these conditions are not always satisfied for silicon-based spiral inductors. For example, in a typical CMOS process with a metallization thickness t of 1.2μm, the metal width would be limited to w<15μm to obtain an accuracy within 5%. Fortunately, *FastHenry* can also be used to compute the metal resistance including the skin effect with great accuracy.

There is another frequency-dependent effect, the proximity effect, which will increase the metal resistance when metal traces are placed closely. A metal trace carrying an alternating current has a changing magnetic field which will cause eddy current losses in nearby conductors. These losses are reflected in the form of increased resistance.

Fortunately, for the spiral structures with only a few turns, the influence of the proximity effect is relatively small compared to that of the skin effect and therefore can be ignored.

The resistive loss caused by the magnetically induced eddy current in the conductive substrate is difficult to evaluate. For a first-order approximation, the reflected resistance increase in the microstrip line due to this loss is proportional to the square of frequency and also to the substrate conductivity σ. This quadratic dependence on frequency has been confirmed experimentally and can be approximated by [75]

$$R_{loss} = \frac{(2\pi\mu_0)^2}{3} f^2 \cdot \sigma \cdot A \tag{4.8}$$

where A is a geometrical factor and μ_0 is the permeability of free space, $4\pi\times10^{-7}$H/m. It can be seen that a higher substrate conductivity and operating frequency result in a higher eddy current loss. The highly conductive substrate in CMOS is one of the reasons that spiral inductors in CMOS technology are inferior to those in bipolar or GaAs technologies.

Equation (4.8) gives only a qualitative description of the eddy current loss in the substrate. No explicit expression for the geometrical factor *A* is available. The only possible way to calculate the exact eddy current loss is using a 3-dimensional numerical simulator for the electromagnetic fields. Due to this complexity in computation, we turn to measurement data for a reasonable evaluation of the eddy current loss. As an example, for a substrate resistivity of 0.14Ω-cm and thickness of 250μm, the substrate resistive loss due to eddy current is about 0.28Ω/mm at 900MHz when reflected in a microstrip line with a width less than 80μm [75]. Here the presence of a lightly-doped epitaxial layer and the spiral geometry, which may result in a reduced eddy current loss, is ignored. The final value of the resistance *R* in the lumped-element circuit model of the microstrip segment (Fig. 4.7) is then obtained by summing the *FastHenry* results and the estimated eddy current loss.

4.2.4 Parasitic Capacitances

For each microstrip segment in the spiral inductor, there is segment-to-substrate capacitance C_{ox} and mutual interline capacitance C_c, as indicated in Fig. 4.4. The analysis and calculation of these capacitances has been reported using various techniques [76]-[80]. The even- and odd-mode analysis seems to be one simple, effective, yet accurate method to calculate the parasitic capacitances associated with a microstrip line [78]-[80].

The capacitance between non-adjacent microstrip lines can be neglected with little change in the final accuracy. Hence, we shall focus on the analysis of two adjacent (coupled) microstrip lines on the silicon substrate. As shown in Fig. 4.8, the total capacitance of the coupled microstrip lines is broken into parallel plate and fringing capacitances for even-mode and odd-mode, respectively. As shown, C_p is the plate capacitance per unit length of a microstrip line with width w on a silicon oxide of height h and is given by

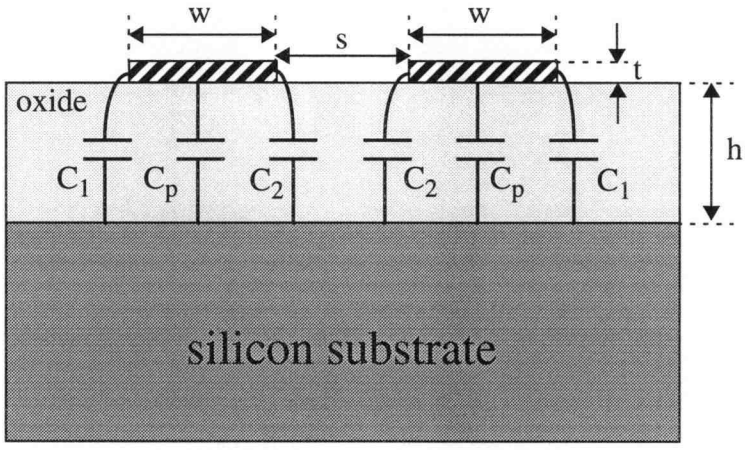
$$C_p = \frac{\varepsilon_0 \varepsilon_r w}{h} \tag{4.9}$$

where ε_0 is the permittivity of the free-space, 8.854 pF/m, and ε_r is the relative dielectric constant of silicon oxide, 3.9. C_1, C_2, C_{ga}, and C_{gd} represent various fringing capacitances in even-mode or odd-mode. They have been approximated using simple formulas given by Garg and Bahl [80]

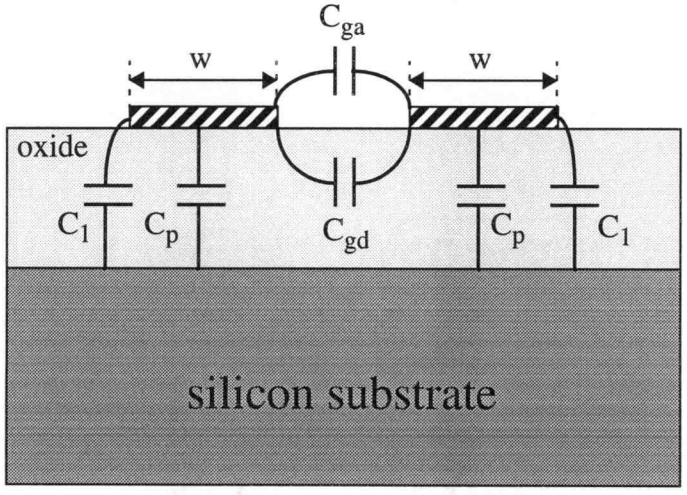
$$C_1 = 0.5\left(\frac{\sqrt{\varepsilon_{re}}}{C_0 Z_0} - \frac{\varepsilon_0 \varepsilon_r w}{h}\right) \tag{4.10}$$

where C_0 is the velocity of light in free space, 2.998×10^8 m/s. ε_{re} is the effective dielectric constant of the microstrip line and can be given approximately by [81]

$$\varepsilon_{re} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2} \frac{1}{\sqrt{1 + 12h/w}} \tag{4.11}$$



(a) Even-mode capacitances



(b) Odd-mode capacitances

Figure 4.8: Decomposition of the total capacitance of coupled microstrip lines. (a) Even-mode. (b) Odd-mode.

The characteristic impedance Z_0 of the microstrip line can then be calculated as [81]

$$Z_0 = \begin{cases} \dfrac{60}{\sqrt{\varepsilon_{re}}} \ln\left(\dfrac{8h}{w} + \dfrac{w}{4h}\right) & \text{for } w/h \le 1 \\ \dfrac{120\pi}{\sqrt{\varepsilon_{re}}[w/h + 1.393 + 0.667\ln(w/h + 1.444)]} & \text{for } w/h \ge 1 \end{cases} \tag{4.12}$$

The expression for fringing capacitance C_2 is obtained empirically as [80]

$$C_2 = \frac{C_1}{1 + A\dfrac{h}{s}\tanh(8s/h)} \tag{4.13}$$

where s is the spacing between the coupled microstrip lines and

$$A = \exp[-0.1\exp(2.33 - 2.53w/h)] \tag{4.14}$$

Odd-mode capacitance C_{ga} corresponds to the fringing field across the air gap and is approximately [80]

$$C_{ga} = \frac{\varepsilon_0 K(k')}{2\,K(k)} \tag{4.15}$$

where

$$k = \frac{s/h}{s/h + 2w/h} \tag{4.16}$$

$$k' = \sqrt{1 - k^2} \tag{4.17}$$

and the ratio of the complete elliptic function $K(k)$ and its complement $K(k')$ is given by

$$\frac{K(k')}{K(k)} = \begin{cases} \dfrac{1}{\pi}\ln\left(2\dfrac{1 + \sqrt{k'}}{1 - \sqrt{k'}}\right) & \text{for } 0 \le k^2 \le 0.5 \\ \dfrac{\pi}{\ln\left(2\dfrac{1 + \sqrt{k}}{1 - \sqrt{k}}\right)} & \text{for } 0.5 \le k^2 \le 1 \end{cases} \tag{4.18}$$

Capacitance C_{gd} in odd-mode corresponds to the fringing field across the gap in the dielectric region (silicon oxide) and is evaluated as [80]

$$C_{gd} = \frac{\varepsilon_0 \varepsilon_r}{\pi} \ln\left[\coth\left(\frac{\pi}{4}\frac{s}{h}\right)\right] + 0.65 C_1 \left(\frac{0.02}{s/h}\sqrt{\varepsilon_r} + 1 - \varepsilon_r^{-2}\right) \tag{4.19}$$

Upon obtaining the even-mode and odd-mode capacitances using Equations (4.9)-(4.19), the capacitance parameters C_{ox} (segment-to-substrate) and C_c (interline coupling) in the lumped-element circuit model of a microstrip segment can be simply calculated as

$$2C_{ox} = (C_p + 2C_2) \cdot Length \tag{4.20}$$

for the microstrip segments of in-between spiral turns, and

$$2C_{ox} = (C_p + C_1 + C_2) \cdot Length \tag{4.21}$$

for the microstrip segments of edge spiral turns (outermost or innermost turns), and

$$2C_c = [2(C_{ga} + C_{gd}) - C_2] \cdot Length \tag{4.22}$$

Since C_1 is always larger than C_2 (Equation (4.13)), the microstrip segments of edge spiral turns have a larger C_{ox} per unit length than in-between spiral turns.

The capacitance expressions given above are derived assuming zero strip thickness. For microstrip lines with finite thickness t, the capacitances can be evaluated using the concept of effective width [82][83], given by [83]

$$w_{eff} = w + \Delta w(1 - 0.5e^{-0.69\Delta w/\Delta t}) \tag{4.23}$$

where

$$\Delta t = \frac{th}{\varepsilon_r s} \tag{4.24}$$

and

$$\Delta w = \begin{cases} t[1 + \ln(2h/t)]/\pi & w > \dfrac{h}{2\pi} > 2t \\ t[1 + \ln(4\pi w/t)]/\pi & \dfrac{h}{2\pi} > w > 2t \end{cases} \tag{4.25}$$

Expression (4.23) is valid only for calculating even-mode capacitances when $s \gg 2t$. Unfortunately, for the spiral structures in silicon technology, this condition is often not satisfied. From Equation (4.23), we observe that $(w_{eff} - w)$ has a maximum value of Δw when s is infinite and decreases when s gets small. Considering Δw to be small compared with w, therefore, the influence of the effective width on closely coupled microstrip lines (as in the spiral inductors) is negligible. However, the additional odd-mode coupling capacitance arising from nonzero strip thickness should not be ignored when t is comparable to s. This excess coupling capacitance can be approximately modeled by a parallel-plate capacitance [83]. Hence, to account for the finite strip thickness effect, the interline coupling capacitance (Equation (4.22)) should be modified as

$$2C_c = [2(C_{ga} + C_{gd}) - C_2 + \varepsilon_0 t/s] \cdot Length \tag{4.26}$$

Given Equations (4.9)-(4.21) and (4.26), the capacitance parameters in the lumped-element circuit model of the microstrip segments can be easily obtained. One thing worthy of mention is that unequal effective microstrip lengths should be used for the calculation of various capacitances, due to the presence of the spiral corner as shown in Fig. 4.9. Specifically, *l1* should be used as *Length* in Equations (4.20) and (4.21), and *l2* as *Length* in Equation (4.26) where *l2=l1-w-s*.

4.2.5 Substrate Parasitics

Because the operating frequency is high (in GHz) in RF IC's, the influence of the substrate on inductor performance becomes significant and must be modeled carefully. However, modeling substrate effects is complicated and numerical analysis is usually used. In this subsection, we will discuss the substrate modeling using closed-form expressions, as a first-order approximation of the substrate influence on spiral inductor performance.

In addition to the substrate resistive loss caused by magnetically induced eddy current as discussed before, the silicon substrate forms a shunt path to ground for the signal present in the spiral structures. Fig. 4.10 illustrates a schematic cross-section of the substrate in CMOS technology, which consists of a lightly-doped P^- epitaxial layer grown on a heavily-doped P^+ bulk substrate. For each microstrip segment in the spiral structure, the underlying epitaxial layer and bulk substrate can be modeled as an RC network as indicated in Fig. 4.10.

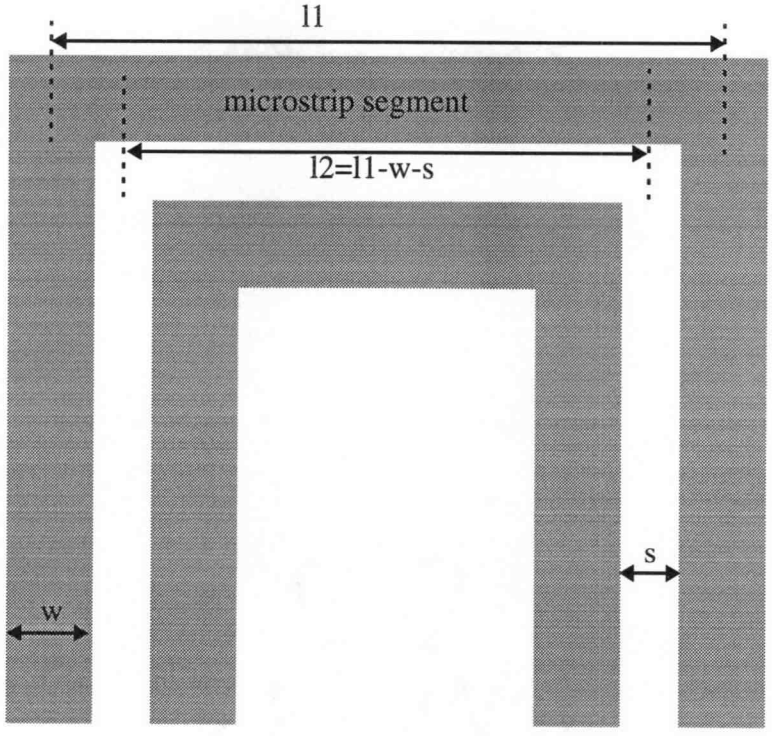


Figure 4.9: Effective microstrip lengths for the calculation of various capacitances.

The bulk capacitance C_{bulk} can be ignored here since the heavily-doped substrate resistivity is typically very small (about 0.1Ω-cm), which at 900MHz indicates a slow-wave mode of wave propagation as described in [75]. In other words, the bulk capacitance C_{bulk} is by-passed by the much smaller impedance R_{bulk}. Furthermore, even R_{bulk} is negligible due to its small value. Therefore, the heavily-doped bulk substrate can be regarded as a single node [84][85].

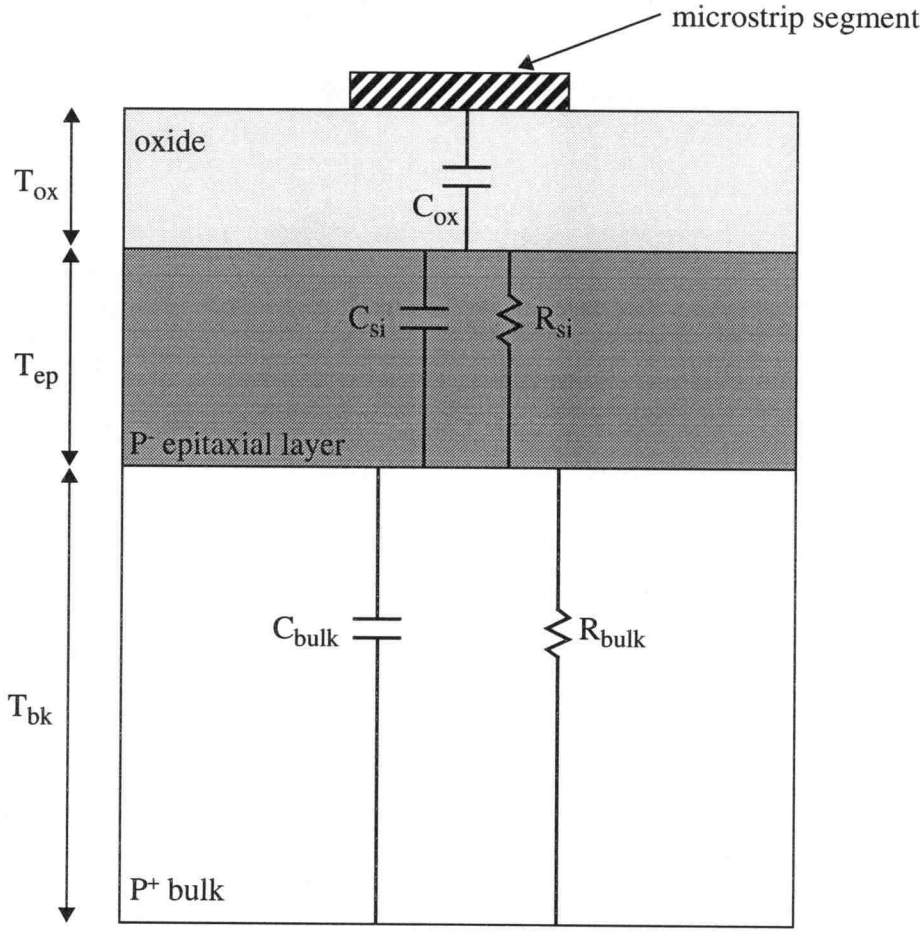


Figure 4.10: Schematic cross-section of the substrate in CMOS technology.

The lightly-doped epitaxial layer usually has a resistivity ρ_{ep} between 10Ω-cm to 20Ω-cm and a thickness T_{ep} of about 7μm. Resistance R_{si}, the loss caused by the transverse component of the conduction current in the epitaxial layer, can be evaluated approximately using closed-form expressions given in [75]. For a microstrip segment with width w and length l, R_{si} is given by

$$\frac{1}{2}R_{si} = \rho_{eff}\frac{T_{eff}}{wl} \tag{4.27}$$

where the effective resistivity ρ_{eff} is

$$\rho_{eff} = \frac{\rho_{ep}}{0.5 + 0.5/\sqrt{1 + 12T_{ep}/w}} \tag{4.28}$$

and the effective thickness T_{eff} is

$$T_{eff} = \begin{cases} \dfrac{w}{2\pi}\ln\left(\dfrac{8T_{ep}}{w} + \dfrac{w}{4T_{ep}}\right) & \text{for } w/T_{ep} \le 1 \\ \dfrac{w}{\left[\dfrac{w}{T_{ep}} + 1.393 + 0.667\ln\left(\dfrac{w}{T_{ep}} + 1.444\right)\right]} & \text{for } w/T_{ep} \ge 1 \end{cases} \tag{4.29}$$

Please note that Equations (4.28) and (4.29) are somewhat different than those given in [75] because we have adopted the improved formulas for the effective dielectric constant and characteristic impedance of the microstrip line as given in Equations (4.11) and (4.12) respectively [81].

Upon knowing R_{si}, the shunt capacitance C_{si} of the epitaxial layer can be obtained directly using Maxwell's Equations [86]-[87], from which we know both the normal (resistive) current density $\bar{J}$ and the displacement (capacitive) current density $\bar{D}$ as determined by the electric field $\bar{E}$. Recall that $\bar{J} = \bar{E}/\rho_{ep}$ (corresponding to $1/R_{si}$) and $\bar{D} = \varepsilon_0\varepsilon_r\bar{E}$ (corresponding to C_{si}). It is obvious that C_{si} and R_{si} are related by

$$R_{si}C_{si} = \frac{\rho_{ep}}{\bar{E}} \cdot \varepsilon_0 \varepsilon_r \bar{E} = \varepsilon_0 \varepsilon_r \rho_{ep} \tag{4.30}$$

where ε_r=11.9 is the relative dielectric constant of the epitaxial layer (Si).

Please note for ρ_{ep} about 10Ω-cm, the time constant $R_{si}C_{si}$ is about 10ps, indicating a cut-off frequency of about 15GHz (at which C_{si} has the same impedance as R_{si}). Therefore, for all practical silicon RF frequencies (below 2.4GHz), C_{si} can be ignored. Nevertheless, if the resistivity ρ_{ep} of the epitaxial layer increases so that its cut-off frequency is comparable to silicon RF frequencies, C_{si} should be included for accurate simulation of the substrate effects.

4.2.6 Summary

We have discussed the extraction of all the electrical parameters for the lumped-element circuit model of a spiral inductor using the classical microstrip line theory. With the aid of a computer program, these electrical parameters can be easily calculated for various inductor geometry dimensions and changes in fabrication technology. The scalable circuit model shown in Fig. 4.4 can then be built and used directly in standard circuit simulators (e.g., SPICE) along with other active and passive RF circuit elements to evaluate the complete circuit performance.

There are a number of other parasitics and higher-order effects which are usually negligible but should be taken into consideration when a very accurate inductor model is needed. For example, current crowding at the corners of the rectangular spiral adds parasitic inductance and capacitance which can be accounted for by a connection of lumped elements at each corner node. For frequencies in the low GHz range, this effect is small and

is often neglected [88]. Also, when applicable, temperature coefficients can be added to every resistive term in the lumped-element model to simulate the variations of the inductor loss with temperature.

4.3 Inductor Circuit Performance

The scalable circuit model of a spiral inductor derived in the previous section is most suitable for computer simulation but is inconvenient for hand analysis due to its complexity. To gain intuitive insight into the parameters' influence on inductor performance such as quality factor and self-resonant frequency, a compact circuit model should be developed for the inductor to approximate the fully scalable circuit model. It should posses great simplicity facilitating the derivation of inductor circuit performance and the optimization of more complex RF circuits.

4.3.1 Compact Circuit Model

A single π-model lumped-element circuit, as shown in Fig. 4.11, has been used by many researchers as a compact model for silicon monolithic inductors [52], [57][58], [62], [70]. This compact circuit model is similar to that of a microstrip segment we discussed before. Usually the electrical parameters in this compact model are estimated by fitting experimental measurements. However, the compact circuit model can be established directly from the electrical parameters of the fully scalable model using the technique described in [57].

Referring to the compact circuit model shown in Fig. 4.11, the series inductance L and resistance R are simply obtained by summing the series inductance and resistance of each individual microstrip segment in the spiral structure. As we discussed before, the

series inductance of each microstrip segment is frequency-dependent due to the phase shift along the spiral. This makes the summation difficult, if not impossible. However, the spiral inductor is usually used at an operating frequency well below its self-resonant frequency. Therefore, the model accuracy is acceptable for hand analysis even when the phase shift along the spiral is ignored.

Similarly, C_{ox}, R_{si}, and C_{si} are estimated as one-half of the summation of shunt parasitics in all the microstrip segments. This approach would result in symmetric shunt branches in the compact circuit model. From a strictly physical perspective, this is not accurate since the inductors are not symmetrical. As a practical matter due to fringing effects, the outermost microstrip segments usually have larger shunt parasitic capacitances

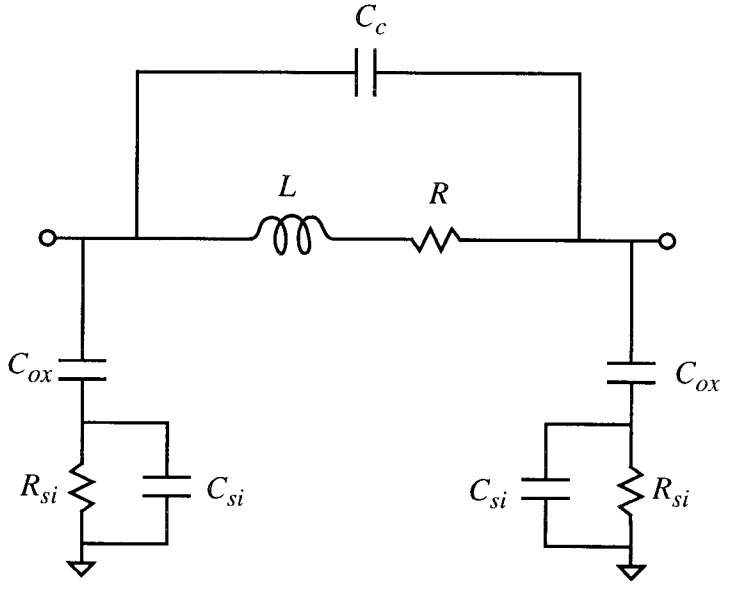


Figure 4.11: Compact lumped-element circuit model for a spiral inductor.

than the inner ones (Sec. 4.2.4). This difference, however, is small [89], and therefore by choosing equal shunt parasitics, we can simplify hand analysis with sufficient accuracy.

The estimation of side coupling capacitance C_c is difficult. One practical method of C_c estimation is by careful parameter fitting, using the fully scalable circuit model as a basis. Fortunately, this capacitance can also be ignored without great loss in model accuracy because of its typically small impact on inductor performance [52], [57].

The compact model is simple and adequate for hand analysis of inductor performance. With the aid of a computer optimizer, it is possible to closely match this compact model to the electrical characteristics of the fully scalable model by refinement of the electrical parameters [57]. Such a refined compact model may be used to replace the fully scalable model in a circuit simulator to reduce the simulation complexity. However, the parameters of the compact model can not be easily adjusted for slight changes in the inductor design because of the nonphysical nature of this simple model.

4.3.2 Quality Factor and Self-Resonant Frequency

In this subsection, we shall derive two figures of merit for the inductor circuit performance from the compact circuit model; i.e., the quality factor Q_l and the self-resonant frequency ω_r.

The quality factor Q_l of an inductor is defined by the ratio of the power stored in the inductive reactance to the total power dissipation in the parasitic resistances. For a first-order approximation, Q_l is given by

$$Q_l = \frac{\omega L}{R} \tag{4.31}$$

based on an equivalent circuit similar to the compact circuit model shown in Fig. 4.11. Equation (4.31) is valid only at low frequencies and a significant error is caused by the parasitic capacitances of a spiral inductor as the frequency increases. To investigate the influence of the parasitic capacitances, we will make use of the formulas for series/parallel impedance transformation as indicated in Fig. 4.12 [90]. We can easily obtain a parallel equivalent circuit for the compact circuit model of an inductor using these transformation formulas. To further simplify the analysis, the coupling capacitance C_c and shunt capacitance C_{si} are ignored. The resulting parallel equivalent circuit is shown in Fig. 4.13. Based on this equivalent circuit, the quality factor Q_l and self-resonant frequency ω_r of the inductor are easily computed. With one end of the inductor grounded, these two figures of merit are given by

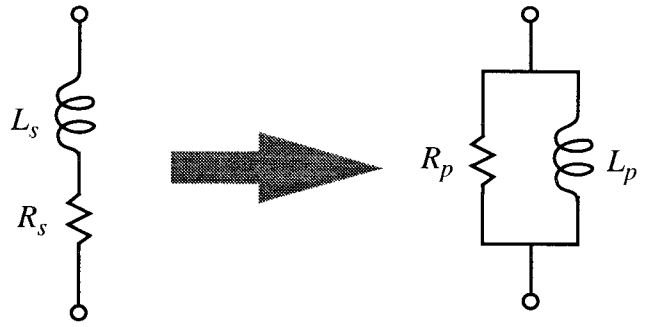
$$Q_l = \frac{\omega L(1-(\omega/\omega_r)^2)\cdot(1-R^2C_{ox}/L)}{R+R_{si}\omega^4L^2C_{ox}^2+(\omega R_{si}C_{ox})^2(R+R^2/R_{si})} \tag{4.32}$$

where the self-resonant frequency is

$$\omega_r = \frac{1}{\sqrt{LC_{ox}}}\left(\frac{1-R^2C_{ox}/L}{1-R_{si}^2C_{ox}/L}\right)^{0.5} \tag{4.33}$$

Note that the effect of the side coupling capacitance C_c can be included simply by replacing C_{ox} with $(C_{ox}+C_c)$ in the above equations since $\omega R_{si}C_{ox}$ is usually less than 0.1. However, C_{si} has little impact on the performance as discussed in Sec. 4.2.5 and thus can be ignored without significant loss of accuracy.

As can be seen from Equations (4.32) and (4.33), the quality factor Q_l decreases rapidly as the self-resonant frequency ω_r is approached. When $\omega=0.707\omega_r$, the quality factor of the inductor will be half that of an inductor without parasitic capacitances. Beyond the self-resonant frequency, the quality factor Q_l becomes negative, indicating that the inductor eventually becomes capacitive. The self-resonant frequency ω_r is limited mainly

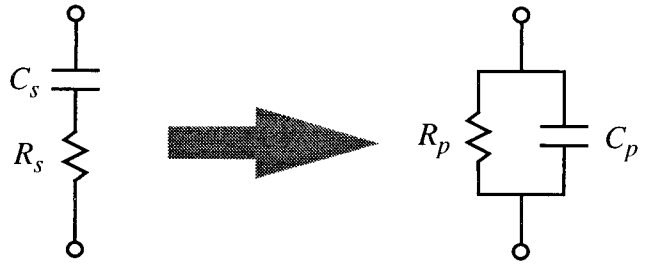


$$Q_{Ls} = \frac{\omega L_s}{R_s}$$

$$L_p = L_s(1 + Q_{Ls}^{-2})$$

$$R_p = R_s(1 + Q_{Ls}^{2})$$

(a) RL series/parallel transformation



$$Q_{Cs} = \frac{1}{\omega R_s C_s}$$

$$C_p = C_s/(1 + Q_{Cs}^{-2})$$

$$R_p = R_s(1 + Q_{Cs}^{2})$$

(b) RC series/parallel transformation

Figure 4.12: Series/parallel impedance transformation formulas. (a) RL network. (b) RC network.

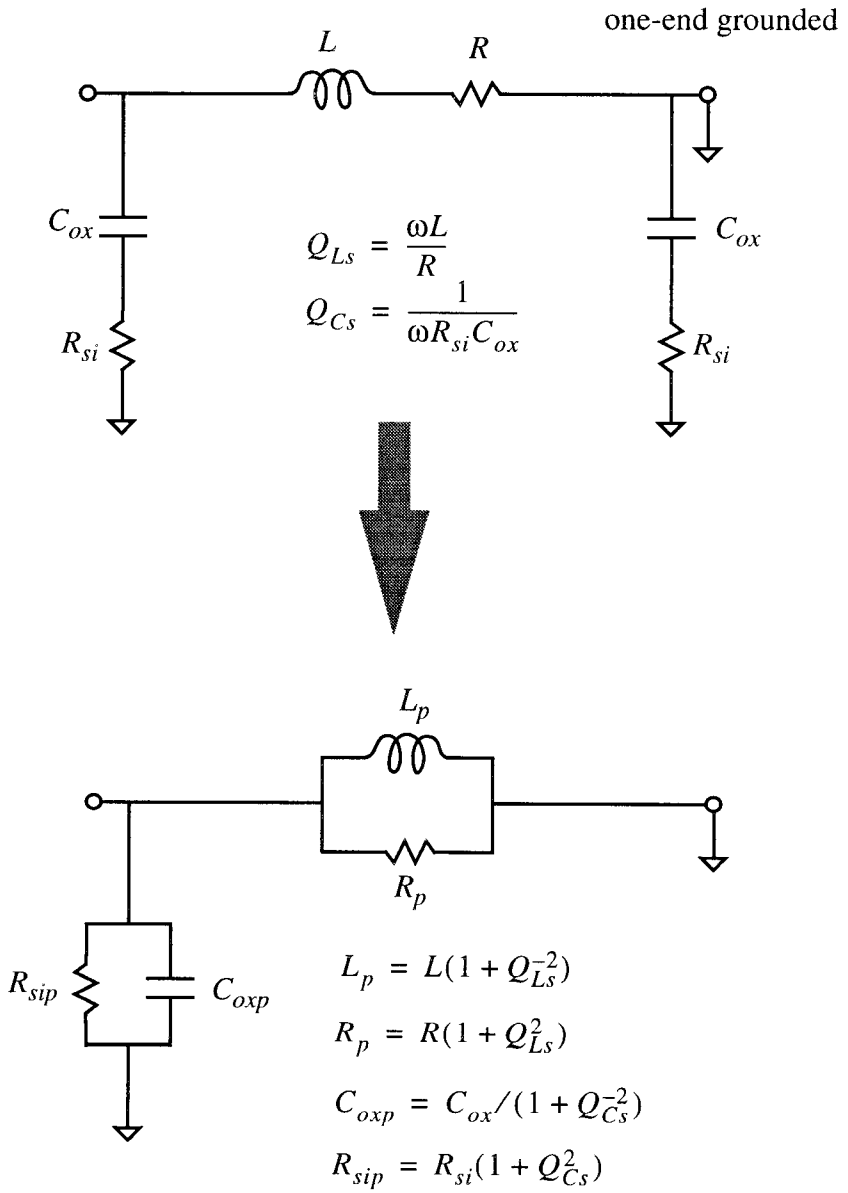


Figure 4.13: Equivalent circuit of an inductor's compact circuit model.

by C_{ox} which is inversely proportional to the oxide thickness. It is clear that decreasing R and C_{ox} increases both Q_l and ω_r. This observation confirms the well-known fact that lower metal resistivity and thicker oxide are desired for high quality monolithic inductors.

As a final remark on inductor circuit performance, it should be noted that if the spiral inductor is used as a floating inductor, the two shunt branches in the compact circuit model are effectively in series with one another. Hence, Equations (4.32) and (4.33) still hold provided that C_{ox} and R_{si} are replaced by $C_{ox}/2$ and $2R_{si}$, respectively. As a consequence, the quality factor Q_l is slightly improved and the self-resonant frequency ω_r is about 1.414 times higher than with one port grounded (The increase of ω_r is actually less due to the coupling capacitance C_c which remains unchanged in both cases).

4.4 Alternative Spiral Inductor Designs

Although rectangular spiral inductors have been widely used, circular spiral inductors provide somewhat higher performance. However, due to the layout limitations, octagonal spirals are often used to approximate circular spirals. It has been proved that the series resistance of a circular and octagonal shaped inductor is smaller by about 10% than that of a rectangular shaped inductor with the same inductance value [91]. A 1.8GHz CMOS VCO using optimized octagonal spiral inductors was described in [55].

There are many other novel spiral inductor designs attempting to achieve higher inductor performance. In this section, we shall investigate several interesting spiral inductors reported recently, with emphasis on the advantages and disadvantages associated with these novel designs.

4.4.1 Suspended Spiral Inductors

As we know, spiral inductor performance is limited by substrate parasitics. Specifically, the low self-resonant frequency of spiral inductors caused by substrate capacitances would limit their use at high frequencies, and the resistive loss in the highly conductive substrate would degrade their quality factor. These characteristic problems of a conductive silicon substrate may be overcome if the area underneath the spiral inductor is removed in a post-processing step using selective etching techniques.

Fig. 4.14 shows a schematic cross-section of a suspended spiral inductor and the substrate after selective etching. As can be seen, the selective etching of substrate leaves the spiral inductor encased in a suspended oxide layer, achieving a much lower capacitance to the substrate. With this selective etching technique, a large inductance value with a high self-resonant frequency is feasible. For example, a 100-nH suspended spiral inductor was

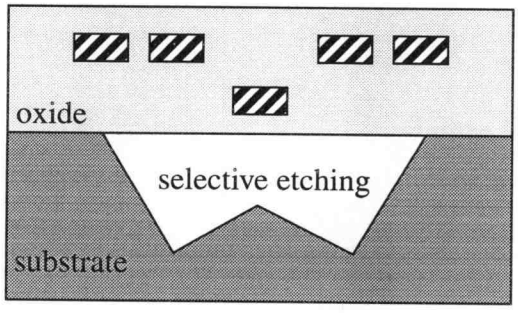


Figure 4.14: Cross-section of a suspended spiral inductor.

demonstrated in [59], achieving a self-resonant frequency of about 3GHz, compared with only 800MHz without the selective etching. [8] and [60] have demonstrated the application of such suspended spiral inductors in a CMOS RF LNA and an oscillator, respectively. This technique also eliminates the resistive losses in the substrate, effectively improving the quality factor. The performance of suspended spiral inductors would be limited mainly by the metal resistance, the minimized substrate capacitance, and the side coupling capacitance. Therefore, suspended spiral inductors in silicon technology would achieve performance comparable to those fabricated in GaAs technology (except that metallization with lower resistivity such as gold may be used in GaAs technology, while aluminum is the only option currently available in standard CMOS).

Though suspended spiral inductors look promising, they require extra non-standard processing steps which result in an increase of the fabrication cost. More importantly, the reliability of suspended spiral inductors is in doubt because of the lack of substrate support for the oxide and metallization.

4.4.2 Multilevel Spiral Inductors

To achieve a low series resistance and thus a high quality factor, multilevel spiral inductors have been proposed making use of several metal layers shunted with vias [61]-[63]. Fig. 4.15 shows the cross-section of such a multilevel spiral inductor with four levels of metallization. The multilevel sandwich-like structure simulates a thicker, hence more conductive, spiral inductor. By doing so, the series resistance of the spiral inductor is effectively reduced and a higher quality factor may be achieved. A 2nH multilevel spiral inductor in bipolar technology with measured quality factor approaching 10 at 2.4GHz and above 6 at 900MHz was recently reported [61], showing an improvement of up to 100% compared with conventional single-level spiral inductors.

Although there is no modification to the conventional wiring metallurgy and no need for extra processing steps using this technique, multilevel spiral inductors require at least three metal layers to achieve lower series resistance. In addition, the side coupling capacitance increases rapidly with the effective metal thickness and the parasitic capacitance to the substrate also increases due to the decrease of oxide thickness. All these effects lead to a lower self-resonant frequency for multilevel spiral inductors, limiting their use at higher frequencies.

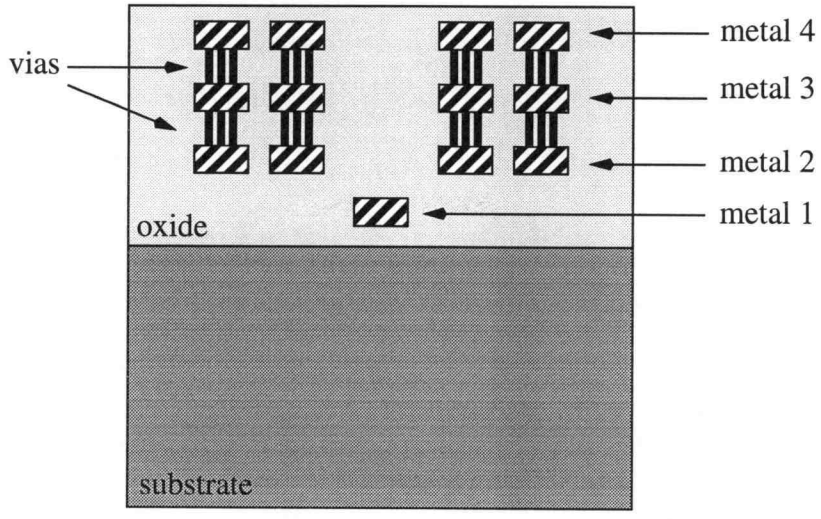


Figure 4.15: Cross-section of a multilevel spiral inductor in a 4-metal process.

4.4.3 Stacked Spiral Inductors

The Greenhouse formulas [67] predict that for any microstrip segment in the spiral structure, more adjacent microstrip segments with same current direction means more positive mutual inductances contributing to its total effective inductance. By stacking two planar spiral structures, it is possible to take advantage of this observation and improve the quality factor of the inductor.

A 1.8GHz CMOS voltage-controlled oscillator with inductors designed using this technique has been recently reported [92]. Fig. 4.16 illustrates a schematic view of such a stacked spiral inductor. Compared to one spiral of the same area, the stacked spiral inductor

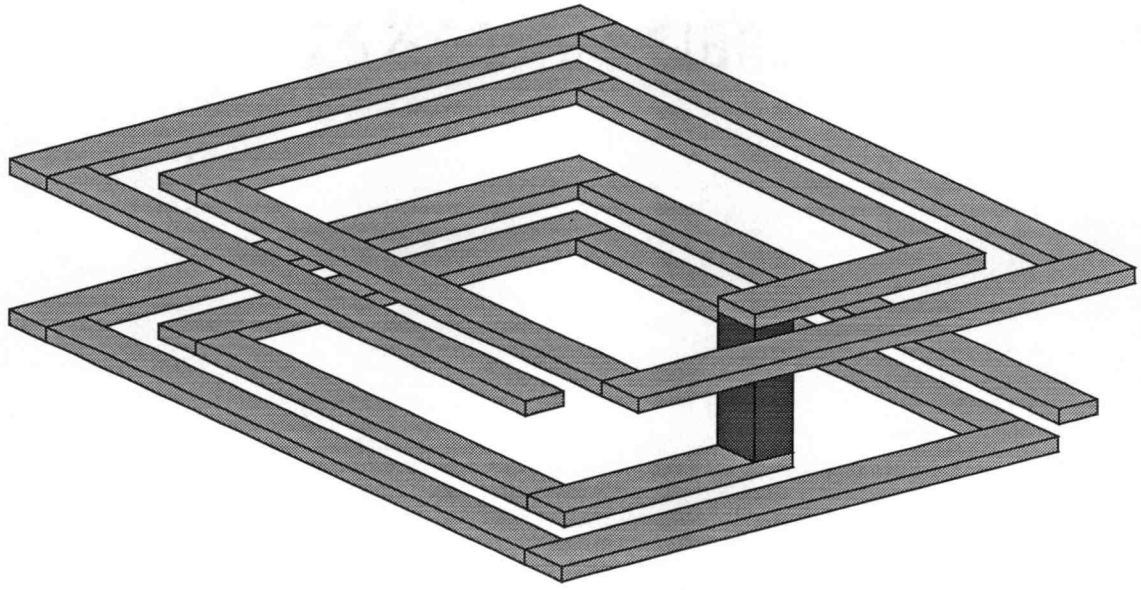


Figure 4.16: A stacked spiral inductor.

has a series parasitic resistance about two times higher but achieves an inductance about four times higher, considering the fact that the spiral inductance is proportional to the square of the number of turns. Therefore, a higher quality factor may be obtained.

The demerit of stacked spiral inductors is similar to that of multilevel spiral inductors. Their self-resonant frequency decreases drastically due to the increase of the coupling capacitance between metal segments and the increased parasitic capacitance to the substrate. The problem is so severe that stacked spiral inductors may be found to be useless for high-frequency applications.

4.4.4 Spiral Inductors with Active Compensation

The effective quality factor of a spiral inductor may be raised through active compensation [93]-[96]. The principle of this technique is illustrated in Fig. 4.17. As

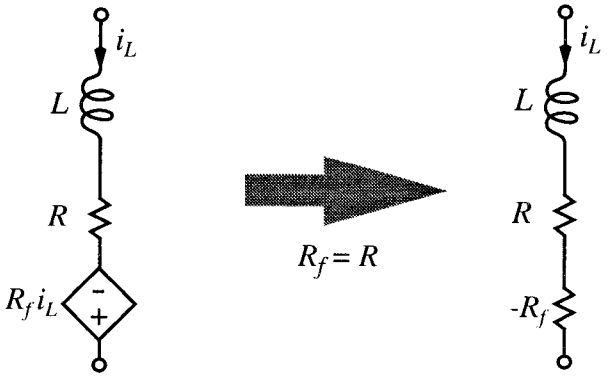


Figure 4.17: Spiral inductors with active compensation.

shown, a positive feedback circuit (current-dependent-voltage-source) is introduced to compensate the resistance loss which consequently enhances the quality factor of the spiral inductor. It is equivalent to inserting a negative resistor $-R_f$ in series with the inductor. By making R_f and the inductor series resistance R equal, the power loss in the spiral inductor will be ideally zero, leading to a much higher quality factor. This active compensation technique is widely used in bandpass filter designs as it also provides an electronically tunable quality factor.

One thing that should be noted is that although the power loss associated with the spiral inductor is reduced using active compensation, the noise actually increases due to the presence of the active components. Generally speaking, the use of spiral inductors with active compensation should be avoided in LNA designs because of the excessive noise associated with them. However, they may find use in the output stage of an LNA where the noise contribution from the inductors is suppressed by the LNA gain.

4.4.5 Spiral Inductors with Ground Shields

In addition to various interesting spiral inductors discussed before, spiral inductors with an appropriate ground shielding have also been reported [97], attempting to suppress the coupling between spiral inductors and the substrate. An improvement up to 25% in the quality factor of spiral inductors with patterned ground shields, which are orthogonal to the direction of current flow in the spiral, has been measured [97]. These have also been extensively used in a CMOS GPS receiver design [98]. However, the use of ground shields increases the inductor's parasitic capacitances, resulting in a lower self-resonant frequency. The trade-off may not be worth for high-frequency applications.

4.5 Conclusions

As can be seen from Equations (4.32) and (4.33), lower metal resistivity, lower substrate conductivity, and thicker oxide are desired for high quality monolithic spiral inductors. Unfortunately, for a chosen CMOS technology, these factors are fixed. For example, although substrate losses can be considerably reduced by using a high-resistivity substrate or silicon-on-insulator (SOI) substrates [99], they are not available in standard digital CMOS technologies. Recently, several novel spiral inductors have been proposed using various design tricks. However, while these new designs achieve higher inductor performance in one aspect, they usually make a trade-off in other aspects such as the self-resonant frequency, noise performance, reliability, or cost.

Inductor performance is also significantly affected by the geometric design. Within the constraints imposed by the conventional process technology, we should concentrate on the optimized geometry, including metal width, metal spacing, number of turns, center hole spacing, and so on, with the aid of computer simulations. Furthermore, for some special applications, it is possible to make use of the mutual inductive coupling between inductors to achieve higher inductor performance. We shall see in the next chapter, that a transformer (two coupled inductors) outperforms two independent inductors in differential applications.

CHAPTER 5. SILICON-BASED MONOLITHIC TRANSFORMERS

The optimization of monolithic inductors is crucially important to obtaining a high performance in LNA circuits. The analysis and modeling of monolithic spiral inductors on silicon substrates have been discussed in the previous chapter. A number of methods to improve and predict their performance are also investigated. It is found within the constraints imposed by conventional process technology that the inductor performance is limited due to the metal resistive loss and substrate parasitics. However, as we will see, two identical monolithic inductors in differential applications can be replaced by a transformer with 1:1 turns ratio for better circuit performance, taking advantage of the strong mutual coupling effect between the primary and the secondary inductors.

In this chapter, a brief introduction to monolithic spiral transformers on a silicon substrate is given first, followed by a detailed description of the characterization and modeling of a transformer consisting of two identical spiral inductors. Design guidelines for transformer optimization are then described based on simulation results. Computer simulation is also conducted to compare the circuit performance between various transformers and inductors.

5.1 Introduction to Monolithic Transformers

Transformers have been widely used in low-power electronic circuits for impedance matching to achieve maximum power transfer, for voltage/current step-up or step-down conversions, and for dc isolation. Monolithic spiral transformers also have a wide variety of potential applications as components to perform impedance matching, signal coupling, and phase splitting functions, in MMIC and silicon RFIC designs [100]-[103].

Spiral transformers implemented in a production BiCMOS process with standard metallization have been exploited as elements for low-loss feedback and single-ended to differential signal conversion in a 1.9GHz receiver front-end [100]. They have also been used as matching and coupling elements in an image rejection mixer [101] and a balanced amplifier topology [102][103]. However, little has been investigated on the use of a transformer as two coupled inductors to achieve better performance in fully-differential circuits [13]. As will be demonstrated in the following sections, compared to using two independent inductors, there are three advantages to utilizing a transformer in a fully-differential LNA circuit. First, a transformer with the same equivalent differential-mode inductance occupies less die area and thus has less series metal resistance and substrate parasitics. Better circuit performance can be achieved especially as the required equivalent inductance increases. Second, a transformer provides additional common-mode rejection for the fully-differential LNA circuit. Unlike two independent inductors, a transformer has a reduced equivalent inductance in the common mode. Hence, the common-mode circuit is effectively de-tuned which significantly reduces the common-mode gain. Finally, because of the symmetric inter-winding layout of the transformer, substrate noise coupling through the parasitic capacitance is more likely to be seen as a common-mode signal by the transformer, leading to a higher substrate noise rejection.

A monolithic spiral transformer can be formed by a series of turns of thin metallized coupled microstrip lines [104]. This structure can be easily changed to realize transformers with various turns ratio. It has been analyzed and modeled in [104] and an improved layout has been proposed for an amplifier design in [103]. However, this transformer structure is not perfectly symmetrical even for 1:1 turns ratio. Since the transformer described herein is intended to be used as a substitute for two identical inductors in differential applications, a symmetrical structure with identical primary and secondary windings should be chosen. By inter-winding two identical spiral inductors, such a transformer can be formed with

inherent symmetry. One possible layout of such a transformer is shown in Fig. 5.1, in which the primary and secondary windings comprise two identical spiral inductors.

As shown in Fig. 5.1, the transformer can simply be viewed as two spiral inductors coupled together. Therefore, all the design tricks and analysis methods for the spiral inductor can be directly applied to the transformer. Correspondingly, as discussed in Chapter 4, suspended spiral transformers, multilevel spiral transformers, and active-compensated spiral transformers are all feasible. Even a novel stacked spiral transformer is physically possible, as illustrated in Fig. 5.2.

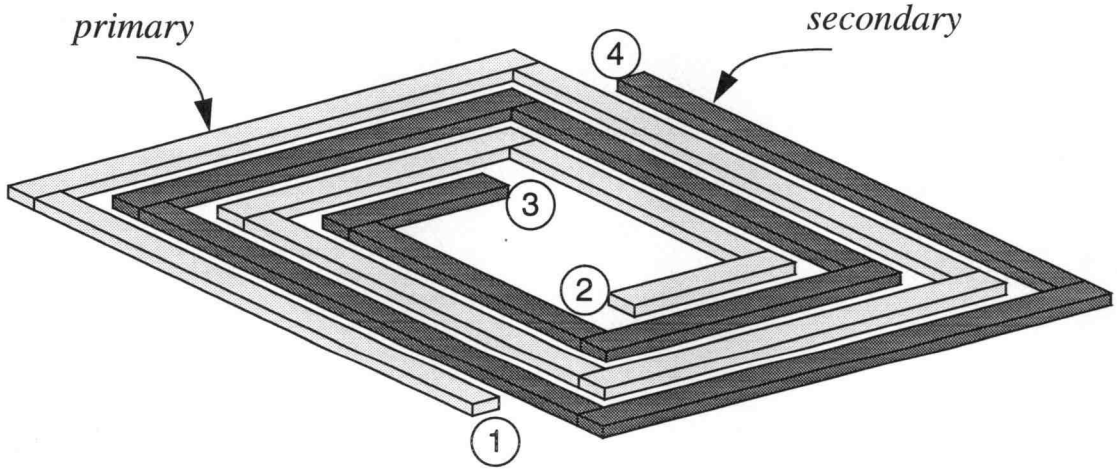


Figure 5.1: Layout of a transformer consisting of two identical spiral inductors.

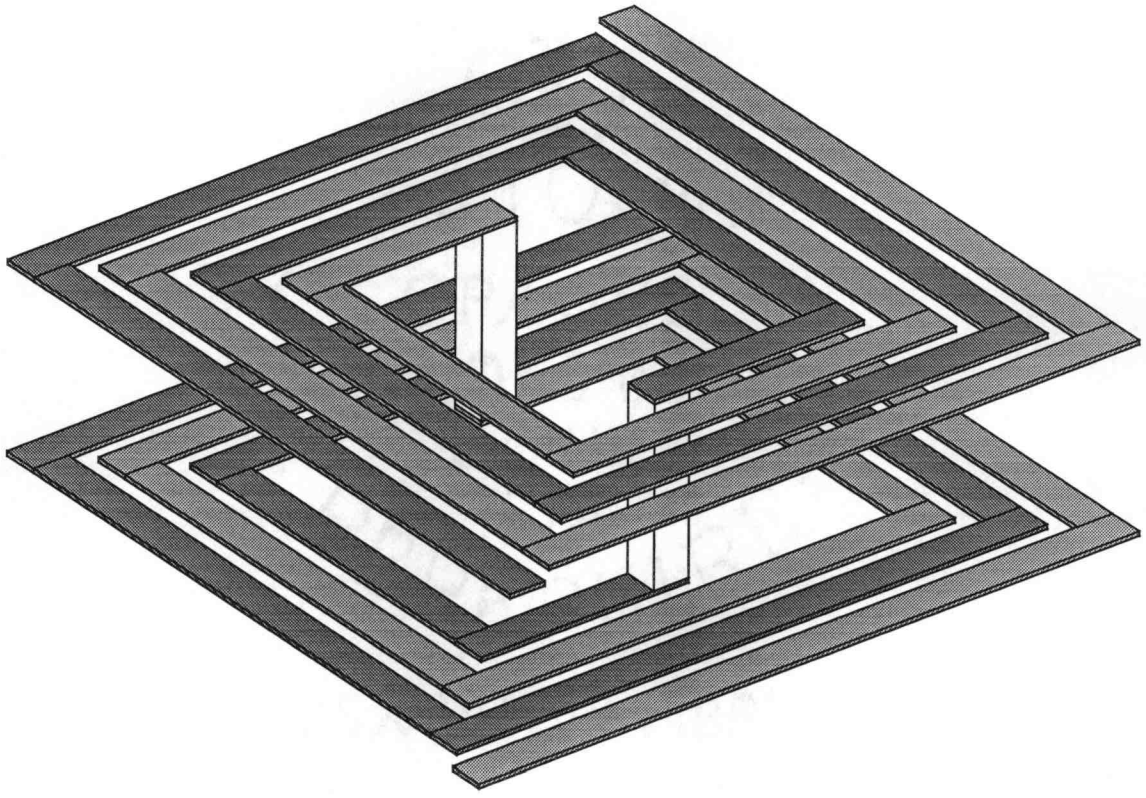


Figure 5.2: A stacked spiral transformer.

5.2 Modeling of Spiral Transformers

5.2.1 Scalable Circuit Model

Modeling of the transformer follows the inductor modeling approach described in Chapter 4. The primary and the secondary windings are first partitioned into a series of segments of coupled microstrip lines, as suggested in [100], [105]-[106]. A lumped-

element circuit model for each segment which is essentially a combination of the π-model of two microstrip lines plus mutual inductive and capacitive coupling effects is then constructed. Because of the inter-winding structure of the transformer and the conductive substrate, there is also a substrate resistive coupling effect between the two coupled microstrip lines which should be considered.

Fig. 5.3 shows a lumped-element circuit model for one segment of the coupled microstrip lines. As discussed in Sec. 4.2.2, the inductances L_1 and L_2 and mutual coupling coefficient k can be computed using the three-dimensional inductance extraction program,

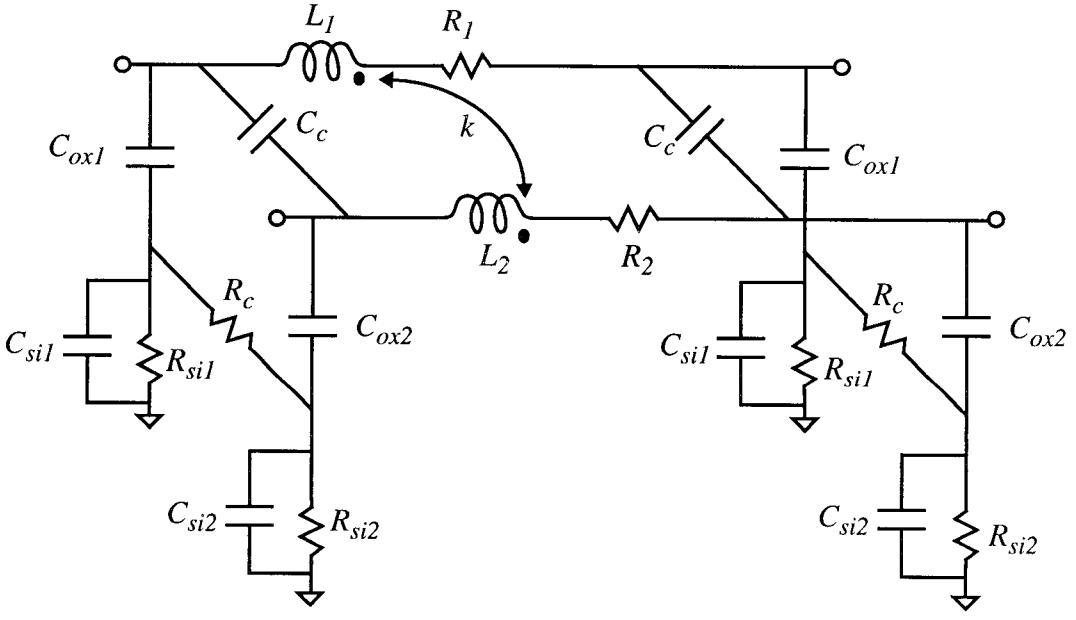


Figure 5.3: Lumped-element circuit model for two coupled microstrip lines.

FastHenry. Please note the inductances L_1 and L_2 include the mutual coupling effects from other microstrip lines in parallel (the vertical coupling effects can be ignored). The frequency-dependent resistances *R1* and *R2* represent the metal trace resistances including the skin effect, computed using *FastHenry*, and the resistive losses caused by the induced eddy current in the heavily-doped silicon substrate (proportional to the square of frequency and the substrate conductivity), as discussed in Sec. 4.2.3. The shunt capacitance C_{ox} and interline coupling capacitance C_c are estimated from the closed-form expressions given in Sec. 4.2.4. Substrate parasitic resistance R_{si} and capacitance C_{si} can be obtained using the same techniques described in Sec. 4.2.5. Resistance R_c represents the resistive coupling between two microstrip lines in the substrate, which can be approximated using the techniques described in [85]. However, this resistance is typically very small because the spacing between two coupled microstrip lines is much smaller than the width and length of the microstrip lines. Hence, R_c can be treated as a short, resulting in a simplification of the lumped-element circuit model for spiral transformers.

A series connection of the lumped-element circuits for two coupled microstrip lines can model a rectangular spiral transformer as shown in Fig. 5.1 and can be employed directly in a standard circuit simulator, such as SPICE. As shown in Fig. 5.4, the scalable circuit model for a 2-turn spiral transformer consists of four lumped-element sub-circuits, corresponding to four segments of coupled microstrip lines in the spiral structure. Please note that substrate resistive coupling R_c is considered as a short circuit. Such a scalable lumped-element circuit model of a spiral transformer can be easily scaled to reflect changes in the geometry dimensions and the fabrication technology, providing valuable design benefits.

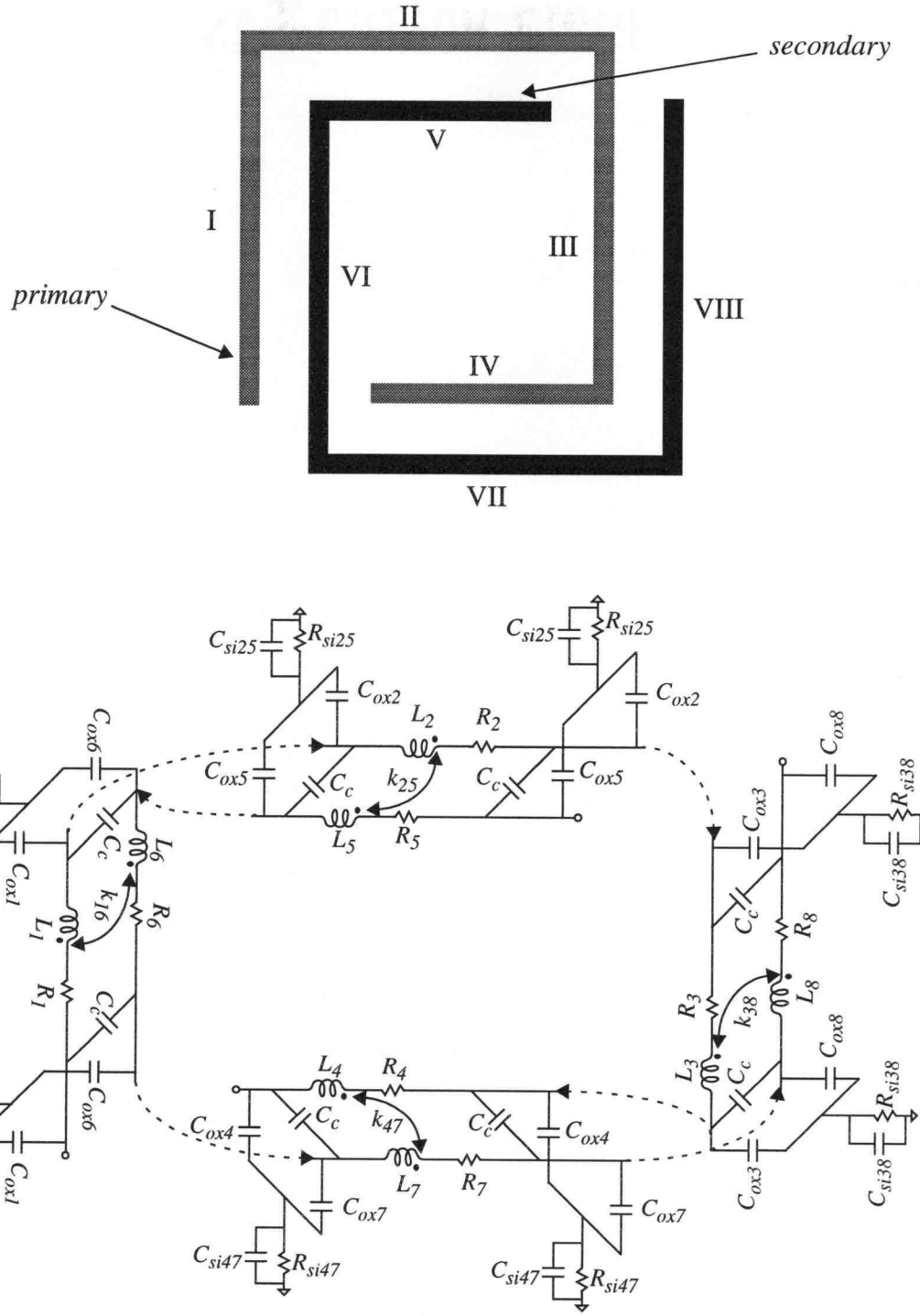


Figure 5.4: Scalable lumped-element circuit model for a spiral transformer.

5.2.2 Compact Circuit Model

Similar to that of spiral inductors, a compact circuit model for spiral transformers is important for RF designers to gain an insight into the parameters influence on the transformer performance, and to facilitate the hand analysis of transformer's circuit performance.

A combination of either π-model lumped-element circuits [106]-[107], or T-model lumped-element circuits [105], has been used as a compact circuit model for silicon monolithic spiral transformers. To make use of the modeling and characterization of spiral inductors as described in the previous chapter, a π-model circuit is chosen for our purposes. As illustrated in Fig. 5.5, the compact circuit model for a spiral transformer is symmetrical because the primary and the secondary are identical spirals. All electrical parameters of this compact model can be obtained using the technique described in Sec. 4.3.1. The small value of substrate coupling resistance R_c is considered a short path between the oxide capacitances of the primary and the secondary in the compact model. Please note that such a compact circuit model for spiral transformers should mainly be used for hand analysis, while the fully scalable circuit model should be used in a circuit simulator for accurate simulations.

5.2.3 Quality Factor and Self-Resonant Frequency

Having established the compact circuit model for spiral transformers, we are ready to derive the quality factor Q_l and self-resonant frequency ω_r for the primary and secondary of a transformer. Since we intend to use the transformer as two identical inductors in differential circuits, two special cases, i.e., differential-mode and common-mode, shall be investigated. For the transformer compact model shown in Fig. 5.5, if equal and opposite

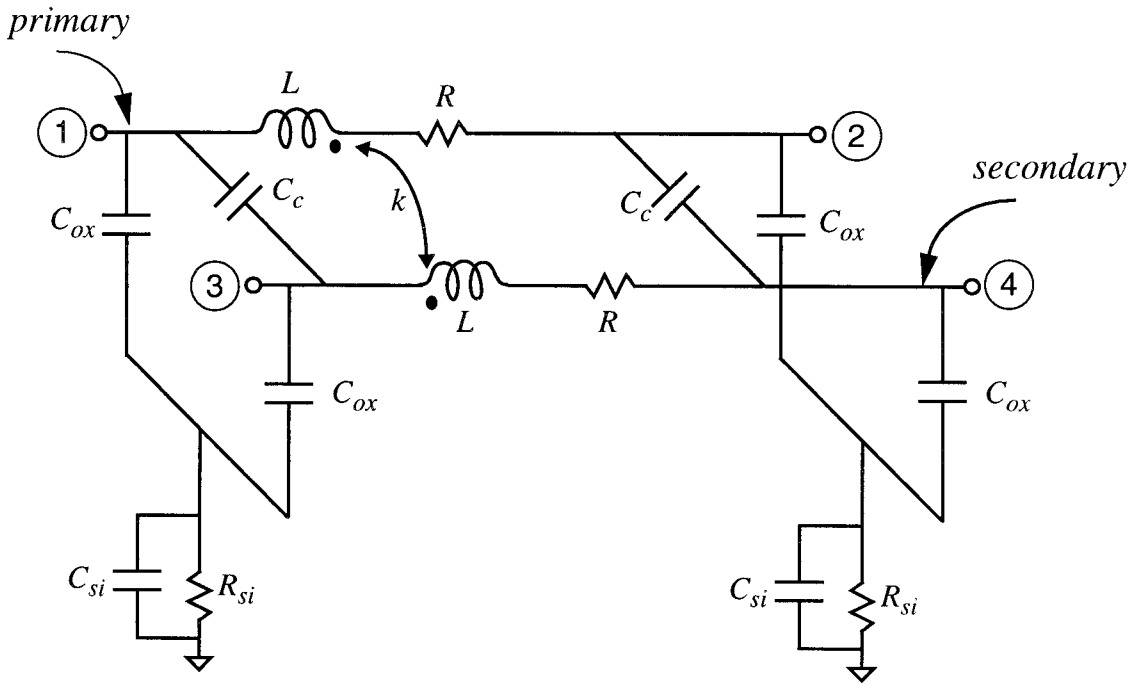
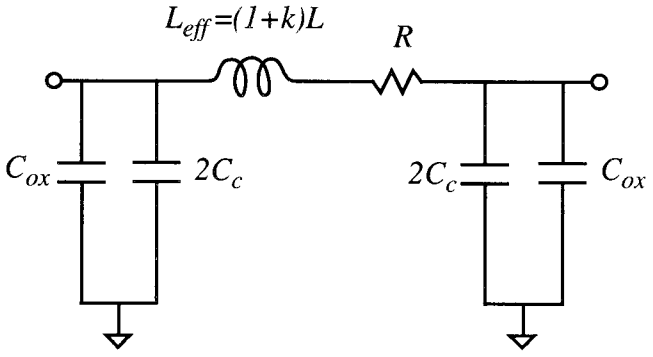
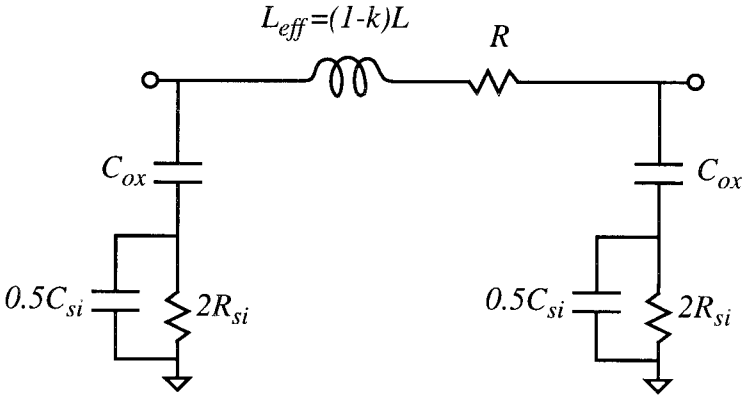


Figure 5.5: Compact circuit model for a spiral transformer.

currents flow through the primary and secondary spiral windings (differential-mode) as in a fully-differential circuit, the effective inductance of the primary and the secondary coils is increased to $L_{eff} = (1+k)L$. On the other hand, if equal currents flow in the same direction (common-mode), the effective inductance becomes $L_{eff} = (1-k)L$. In each mode, the transformer can be partitioned into two independent inductors using the concept of a half-circuit. Fig. 5.6 (a) and (b) show the equivalent half-circuits for a spiral transformer in differential-mode and common-mode, respectively. Based on the half-circuits, the quality



(a) differential-mode



(b) common-mode

Figure 5.6: Equivalent half-circuits for a spiral transformer in (a) differential-mode; (b) common-mode.

factor Q_l and self-resonant frequency ω_r of the primary or the secondary are easily computed as

$$Q_l = \frac{\omega(1+k)L(1-(\omega/\omega_r)^2)}{R}\left(1-\frac{R^2C_{ox}}{(1+k)L}\right) \quad \text{in differential-mode} \qquad (5.1)$$

where

$$\omega_r = \frac{1}{\sqrt{(1+k)LC_{ox}}}\left(1-\frac{R^2C_{ox}}{(1+k)L}\right)^{0.5} \qquad (5.2)$$

and in common-mode,

$$Q_l = \frac{\omega(1-k)L(1-(\omega/\omega_r)^2)\cdot\left(1-\frac{R^2C_{ox}}{(1-k)L}\right)}{R+2R_{si}\omega^4(1-k)^2L^2C_{ox}^2+(2\omega R_{si}C_{ox})^2\left(R+\frac{R^2}{2R_{si}}\right)} \qquad (5.3)$$

where

$$\omega_r = \frac{1}{\sqrt{(1-k)LC_{ox}}}\left(\frac{1-R^2C_{ox}/[(1-k)L]}{1-4R_{si}^2C_{ox}/[(1-k)L]}\right)^{0.5} \qquad (5.4)$$

In the above derivations, the primary or the secondary is assumed to be grounded at one port. When used as a floating transformer, the quality factor Q_l will be slightly improved because the self-resonant frequency ω_r is about 1.414 times higher. Although the effect of the interline coupling capacitance C_c is ignored in the above derivations, it can be included simply by replacing C_{ox} with $(C_{ox}+2C_c)$ in Equations (5.1)-(5.2). In addition, the shunt capacitance C_{si} has little impact on the circuit performance and is neglected here to further simplify the derivations.

From Equations (5.1)-(5.4), it is clear that decreasing R and C_{ox} would increase both the quality factor Q_l and self-resonant frequency ω_r of a spiral transformer. Therefore, we can choose a fabrication technology with lower metal resistivity and thicker oxide to

achieve high quality transformers. However, for a given technology, the optimized spiral geometry, including metal width, metal spacing, number of spiral turns, and center hole spacing, should be carefully investigated with the aid of computer simulation.

5.3 Optimization of Spiral Transformers and Inductors

The design of spiral inductors and transformers for a particular application would require extensive simulation work in order to determine the optimized geometry layout to achieve the best performance within a given technology. In this section, the effects of the physical layout upon the inductors/transformers performance and the complex trade-off between various layout parameters are investigated using computer simulation. Based on the simulation results, a set of guidelines for design optimization of spiral inductors and transformers is summarized.

To conduct the simulation, Matlab programs (see Appendix A) were written which included closed-form expressions for the computation of shunt parasitics C_{ox}, C_c, R_{si}, and C_{si}, derived in Chapter 4. *FastHenry* was used for the computation of the self-inductance, mutual coupling coefficient, and metal resistance including the skin effect. A standard digital process, 3-metal 0.6μm CMOS, was chosen as the fabrication technology. It is assumed that only the topmost third-layer metallization is used to implement both the transformers and inductors, because it provides the lowest metal resistance and oxide capacitance. Table 5-1 lists the MOSIS process parameters used for the computation of the electrical parameters of the inductors and transformers. Note that the resistivity and thickness of the epitaxial layer and the bulk substrate were estimated based on the available process information.

To include the induced substrate resistive loss due to eddy current in the simulation, an empirical closed-form expression for its estimation is derived based on the experimental data given in [75]. For a substrate thickness of 250μm and resistivity of 0.14Ω-cm, the equivalent resistance increase reflected in the metal trace due to eddy current loss is approximately given by

$$R_{loss} = 2.8 \times 10^{-22} \times f^2 \qquad (\Omega/\mu m) \tag{5.5}$$

where f is the operating frequency.

Unless otherwise stated, the following simulations were conducted on various geometry layouts of spiral transformers in differential-mode. Since the primary and the secondary are identical, only layout effects on the circuit performance of the primary inductor are investigated.

Table 5-1: Process parameters of a 3-metal CMOS technology

Parameter	Value
Metal 3 resistivity	0.042 Ω-μm
Metal 3 thickness	1.2 μm
Oxide thickness (Metal 3 to substrate)	4 μm
Epitaxial layer resistivity	10 Ω-cm
Epitaxial layer thickness	7 μm
Substrate resistivity	0.14 Ω-cm
Substrate thickness	250 μm
Oxide dielectric constant	3.9
Silicon dielectric constant	11.9

5.3.1 Circuit Performance v/s Frequency

A spiral transformer with 6 spiral turns (3 turns for each inductor with total length about 2.2mm), metal width of 20μm, trace spacing of 2μm, and center spacing of 100μm was simulated and its element values were computed. Fig. 5.7 shows the compact circuit model at 900MHz for such a transformer. The self-inductance of the primary is about 1.92nH. The mutual coupling coefficient is 0.73, indicating an effective inductance about 3.32nH in differential-mode. The coupling capacitance C_c between the primary and the secondary is about 0.07pF. The oxide capacitance C_{ox} is about 0.21pF, while the substrate shunt capacitance C_{si} is 0.87pF and resistance R_{si} is 12Ω. Please note that C_{si} and R_{si} is the combination of substrate parasitics of both the primary and the secondary. Because the epitaxial layer is only 7μm thick and silicon has a higher dielectric constant than that of oxide, C_{si} is larger than C_{ox}. However, C_{si} can still be ignored due to the small value of R_{si} ($R_{si}C_{si}$ is a constant independent of the geometries).

Fig. 5.8 shows the simulated results of the circuit performance of the primary inductor. As shown, the parasitic series resistance R has a value of 3.87Ω at low frequencies and increases to 4.65Ω at 900MHz due to the skin effect and the eddy current loss in substrate. At higher frequencies, where the eddy current loss dominates, the series resistance is proportional to the square of frequency, as indicated in Fig. 5.8 where the resistance curve shows a slope of 2. When one-end of the spiral is connected to ground, the real part of the one-port impedance of the primary inductor is equal to the parasitic series resistance R at low frequencies and increases rapidly as the primary inductor approaches self-resonance. The imaginary part, on the other hand, first increases and then goes to zero at the self-resonant frequency. Beyond the self-resonance, the primary eventually becomes capacitive. At 900MHz, the imaginary impedance is equivalent to an inductance of 3.42nH, a little bit higher than the series inductance of the primary which is equal to 3.32nH. The

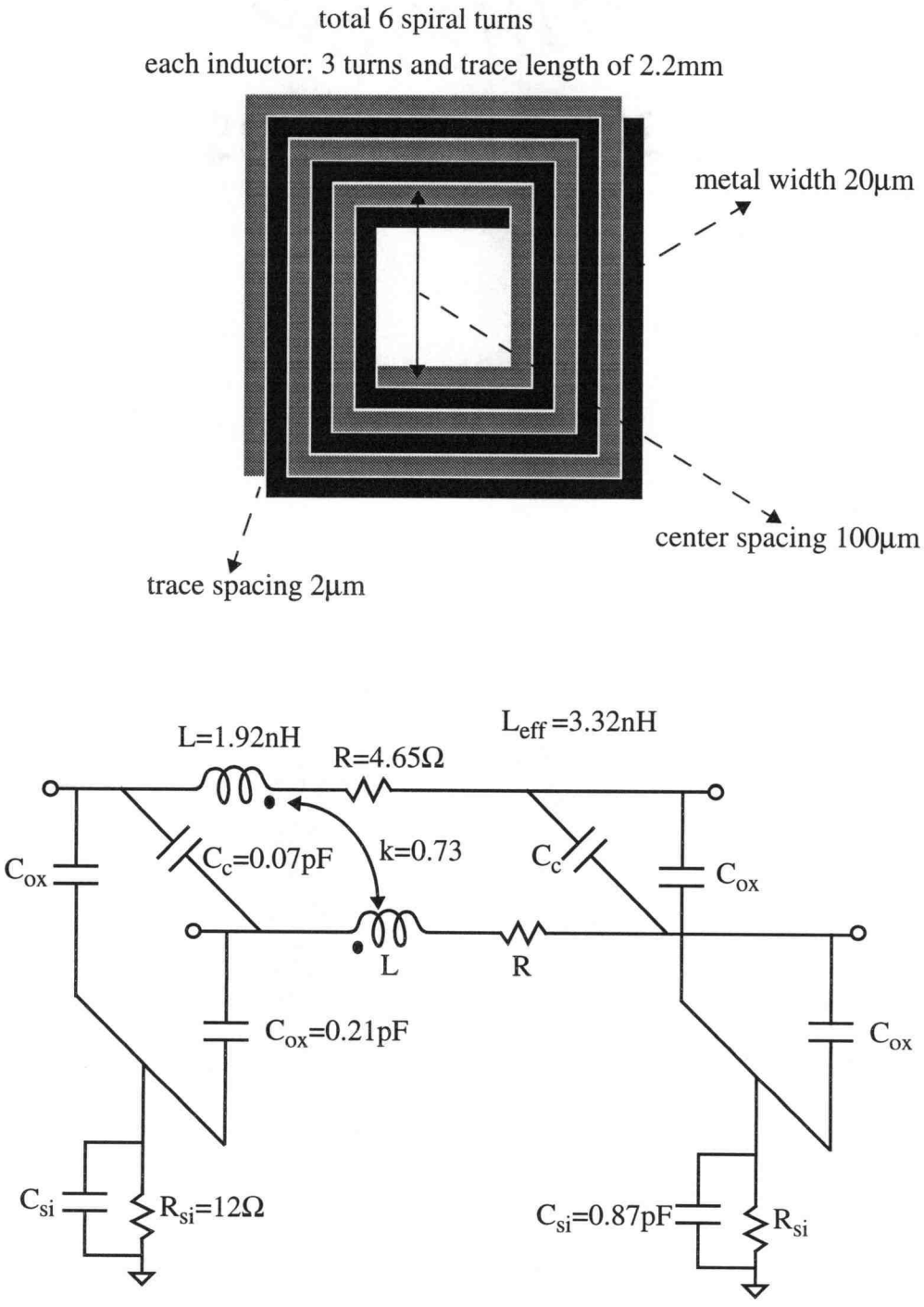


Figure 5.7: A six-turn spiral transformer at 900MHz.

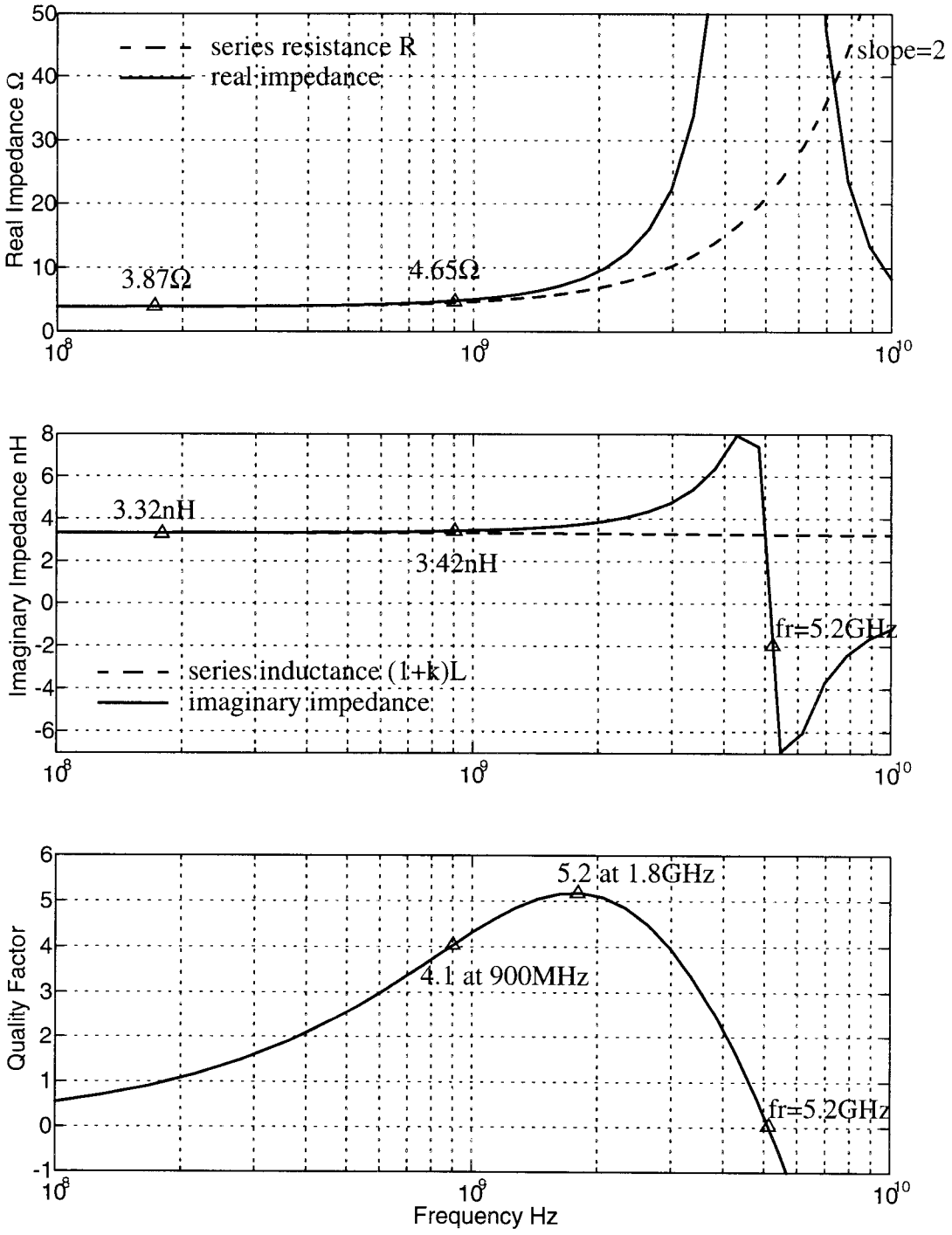


Figure 5.8: Simulated circuit performance of the primary.

quality factor of the primary shows a peak value of 5.2 around 1.8GHz and is equal to 4.1 at 900MHz. The self-resonant frequency, at which the quality factor is equal to zero, is about 5.2GHz.

From the above simulation results, it is clear that the spiral transformer is equivalent to two identical inductors in differential applications, each demonstrating a series inductance of 3.32nH and a differential-mode quality factor of about 4.1 at 900MHz.

5.3.2 Effects of Number of Spiral Turns

A variation in the number of spiral turns has a significant effect on the transformer performance. In order to determine this effect, four different transformers, with two, three, four, and five spiral turns for the primary and the secondary are computed. Each transformer has the same geometry layout except the number of spiral turns. Specifically, the metal width is 20μm, the trace spacing is 2μm, and the spiral center spacing is 100μm.

Fig. 5.9 shows the simulation results for the four transformers. It is observed that the series inductance and resistance of the primary increases with the number of spiral turns. However, the quality factor does not have such a simple relationship with the number of turns. As shown, the primary inductors with three and four spiral turns demonstrate higher peak values of the quality factor than those with two and five spiral turns. At 900MHz, the 4-turn primary inductor has the highest quality factor of about 4.7. In addition, the substrate parasitics have more significant effects on the circuit performance when the number of turns increases. As a direct consequence, the 5-turn primary inductor has a self-resonance frequency only about 1.65GHz, while the 2-turn one achieves a self-resonant frequency higher than 10GHz.

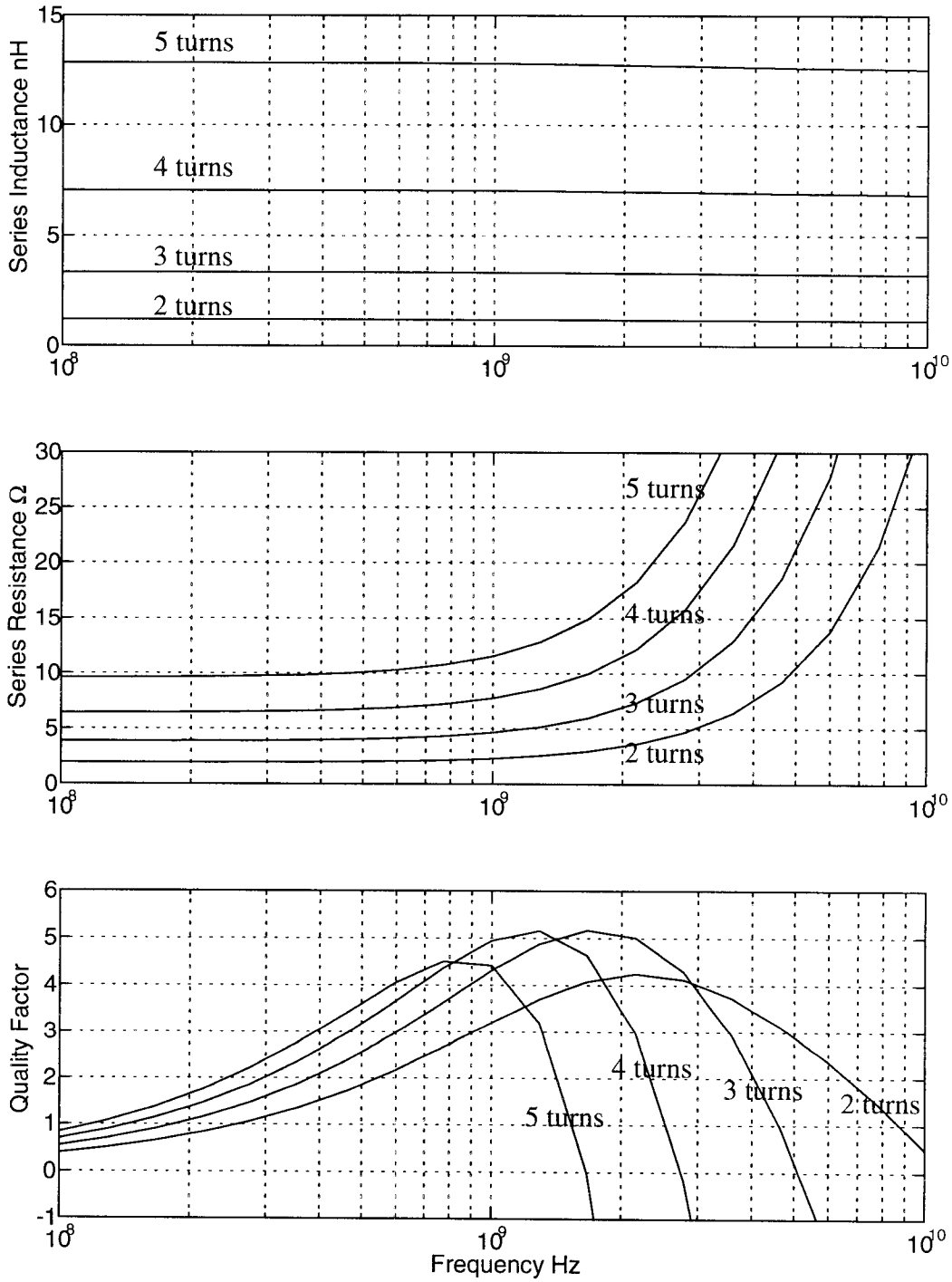


Figure 5.9: Effects of the number of spiral turns.

5.3.3 Effects of Metal Width

The simulated circuit performance of rectangular spiral transformers with metal widths of 10μm, 20μm, 30μm, and 40μm is shown in Fig. 5.10. All transformers have a 3-turn spiral, a trace spacing of 2μm, and a spiral center spacing of 100μm for both the primary and the secondary.

To maintain the same number of spiral turns, the wider metal trace is also longer, which results in a little bit higher series inductance. The inductance dependence on frequency is also more significant for the primary inductor with the wider metal trace, as can be seen in Fig. 5.10. As expected, the series resistance of wider metal trace is smaller at low frequencies where the skin effect and substrate effect are negligible. As frequency continues to increase, the eddy current loss in the heavily-doped silicon substrate and the skin effect begin to increase rapidly. For a wider metal trace, the increase of these effects is faster and eventually its series resistance becomes larger than that of a narrower metal trace. As shown in Fig. 5.10, although the series resistance of 40μm-wide metal trace is much smaller than that of 10μm-wide trace at 900MHz, it is almost identical to that of 30μm-wide metal trace due to the skin effect. Beyond 2.5GHz, 40μm-wide metal trace has the largest series resistance. In addition, the larger area associated with wider metal trace results in a higher parasitic capacitance and thus lower self-resonant frequency. Although a wider trace generates a higher peak value of the quality factor, there is no major difference between a 30μm-wide trace and a 40μm-wide trace, indicating that further increase of the metal width is not an effective way to improve the quality factor. Moreover, the peak quality factor occurs at a lower frequency for a wider metal trace. For example, the peak quality factor of a 40μm-wide trace occurs around 1GHz while the peak quality factor of a 10μm-wide trace occurs at around 2.8GHz. For 900MHz applications, it seems that a 40μm-wide trace provides the optimized design due to its highest available quality factor.

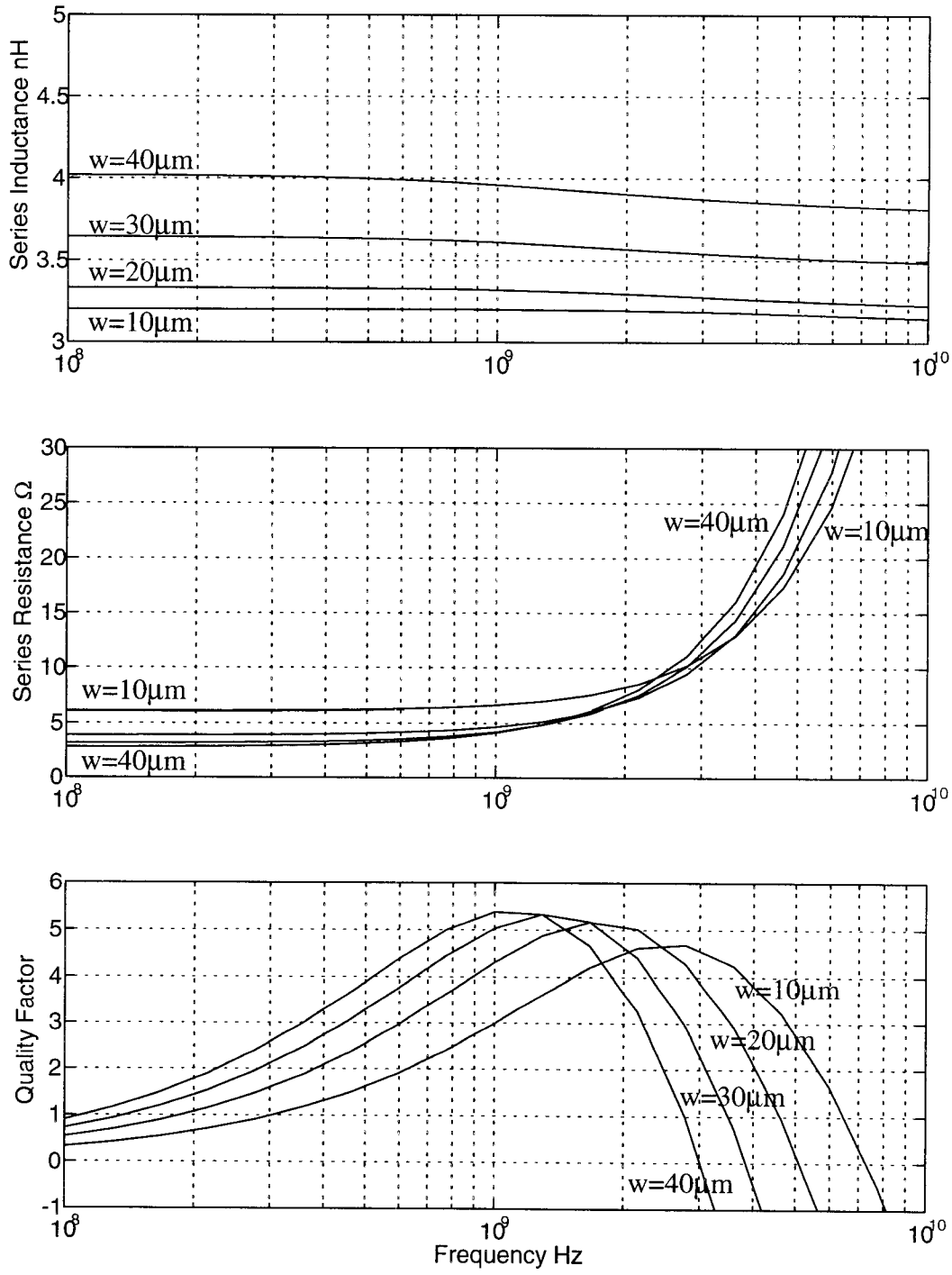


Figure 5.10: Effects of the metal width.

This is true if the required inductance is only about 4nH. The frequency where the peak quality factor occurs decreases as the series inductance increases, as can be seen in Fig. 5.9. If a higher inductance is required, a metal trace with width between 20μm-30μm would provide the best circuit performance around 900MHz.

5.3.4 Effects of Trace Spacing

The effects of trace spacing on the transformer circuit performance have also been investigated. The simulation results of rectangular spiral transformers with trace spacings of 1μm, 3μm, 5μm, and 7μm are shown in Fig. 5.11. Again, all other layout parameters are kept the same, i.e., spiral turns of 3, metal width of 20μm, and spiral center spacing of 100μm, for the primary and the secondary of the four transformers.

As can be seen, the variation of trace spacing has little effect on the transformer circuit performance. Generally speaking, small trace spacing increases the magnetic coupling between each microstrip line and also between the primary and the secondary. This will cause an increase of the available inductance for a given metal trace length. However, reducing trace spacing also results in an increase of the interline coupling capacitance. As a consequence, the self-resonant frequency is decreased. Nevertheless, these effects are found to be insignificant. Probably the biggest advantage of reducing trace spacing is the conservation of die real estate.

5.3.5 Effects of Spiral Center Spacing

To investigate the effects of spiral center spacing on the transformer circuit performance, two transformers with a given primary inductance of about 8nH in

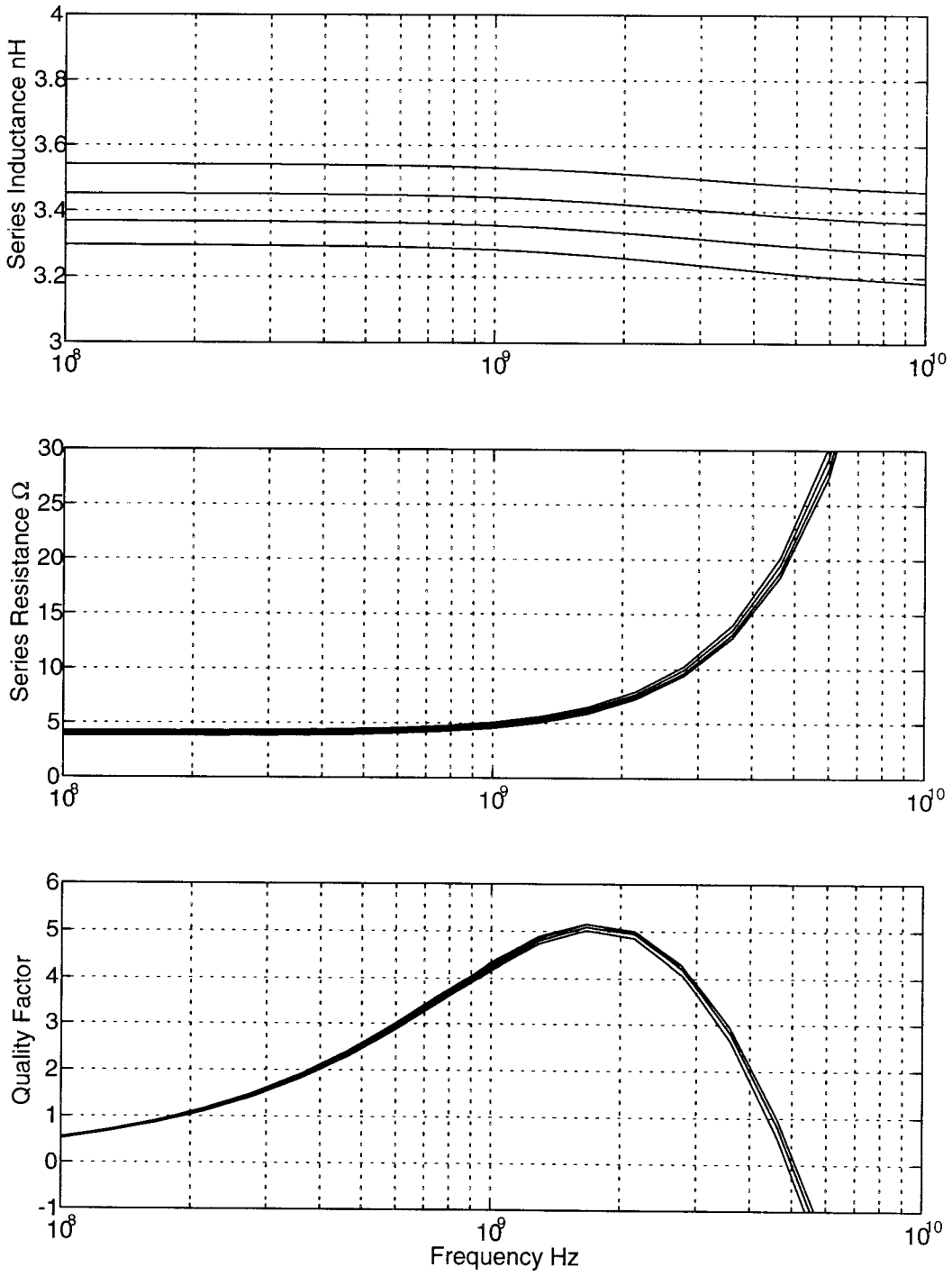


Figure 5.11: Effects of the trace spacing.

differential-mode are simulated with center spacings of 50μm and 120μm, respectively. The simulation results are shown in Fig. 5.12.

It is found that larger center spacing results in lower series resistance and higher quality factor and self-resonant frequency. As we know, the increase of center spacing decreases the negative mutual coupling between opposite sides in the spiral structure. Therefore, to achieve the same inductance, the spiral with larger center spacing needs a shorter metal trace. As a consequence, the parasitic series resistance and shunt capacitance are smaller, which causes an increase in both quality factor and self-resonant frequency. However, there is also a down-side that larger center spacing consumes more die real estate.

5.3.6 Simulations v/s Measurements

In order to verify the accuracy of the simulation, the simulated results are compared with the measurement results of two 5nH spiral inductors fabricated in a 0.8-μm BiCMOS process [57]. The technological parameters for the BiCMOS process used for simulations are: metal resistivity 0.03Ω-μm, metal thickness 1μm, oxide thickness 5μm, substrate thickness 380μm, and substrate resistivity 10Ω-cm. Since the substrate resistivity is about 70 times higher than 0.14Ω-cm, the equivalent resistance increase reflected in the metal trace due to eddy current loss in the substrate is approximately given by

$$R_{loss} = 4 \times 10^{-24} \times f^2 \qquad (\Omega/\mu m) \tag{5.6}$$

where f is the operating frequency.

The simulated quality factors of the two 5nH inductors, with metal widths of 5μm and 15μm respectively, are shown in Fig. 5.13. Compared with the measurement results as shown in Fig. 6 in [57], there is an excellent agreement. For the 5μm inductor, the measured

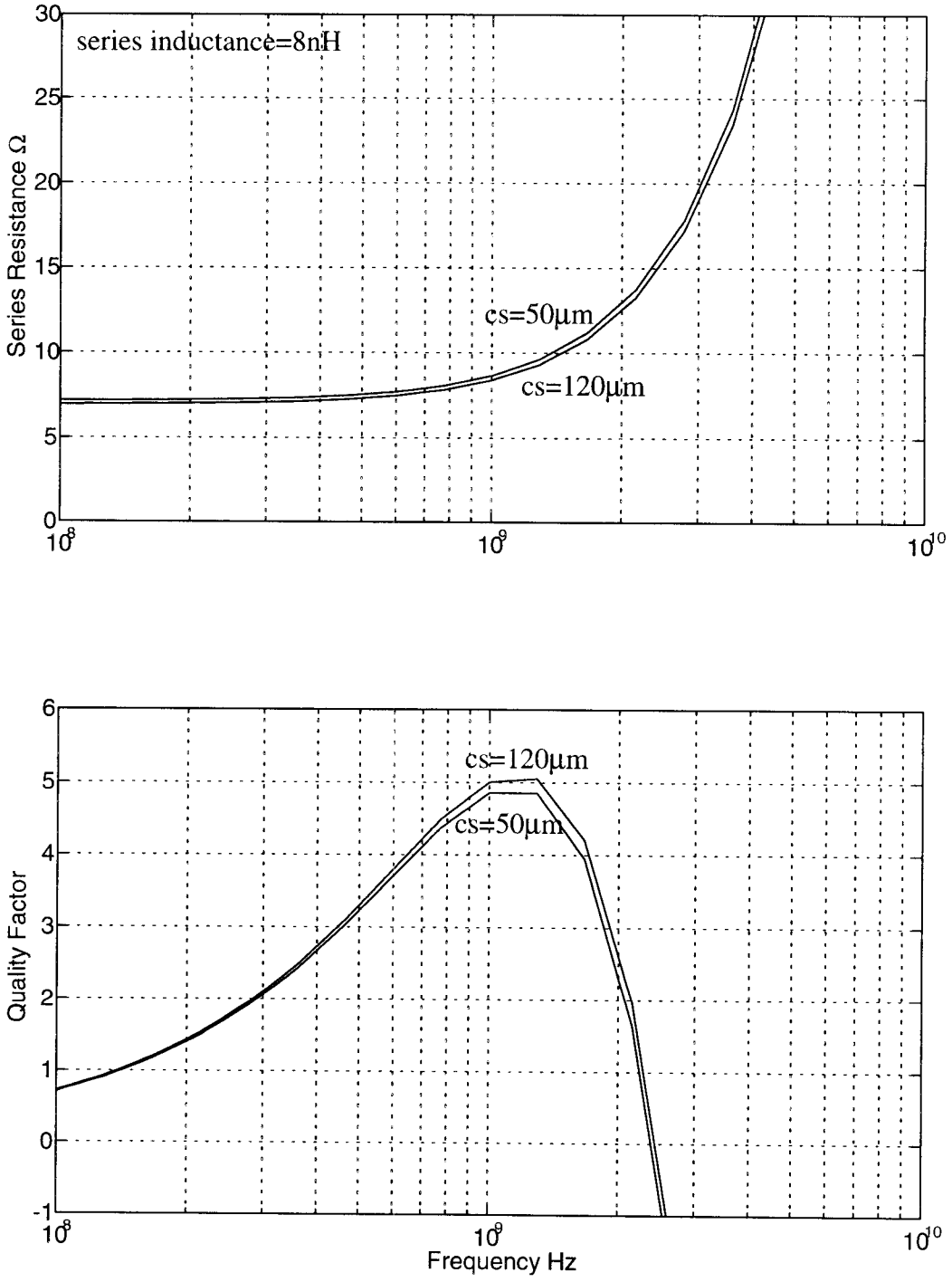


Figure 5.12: Effects of the spiral center spacing.

peak value of the quality factor is coincident with the simulated result, about 5.2 around 3.5GHz. At 1GHz, the measured quality factor is about 2.4 while the simulation shows a quality factor of about 2.2. Around 5GHz, the difference is still within 10%. For the 15μm inductor, the measurements are coincident with the simulation results until the frequency is higher than that where the peak quality factor occurs.

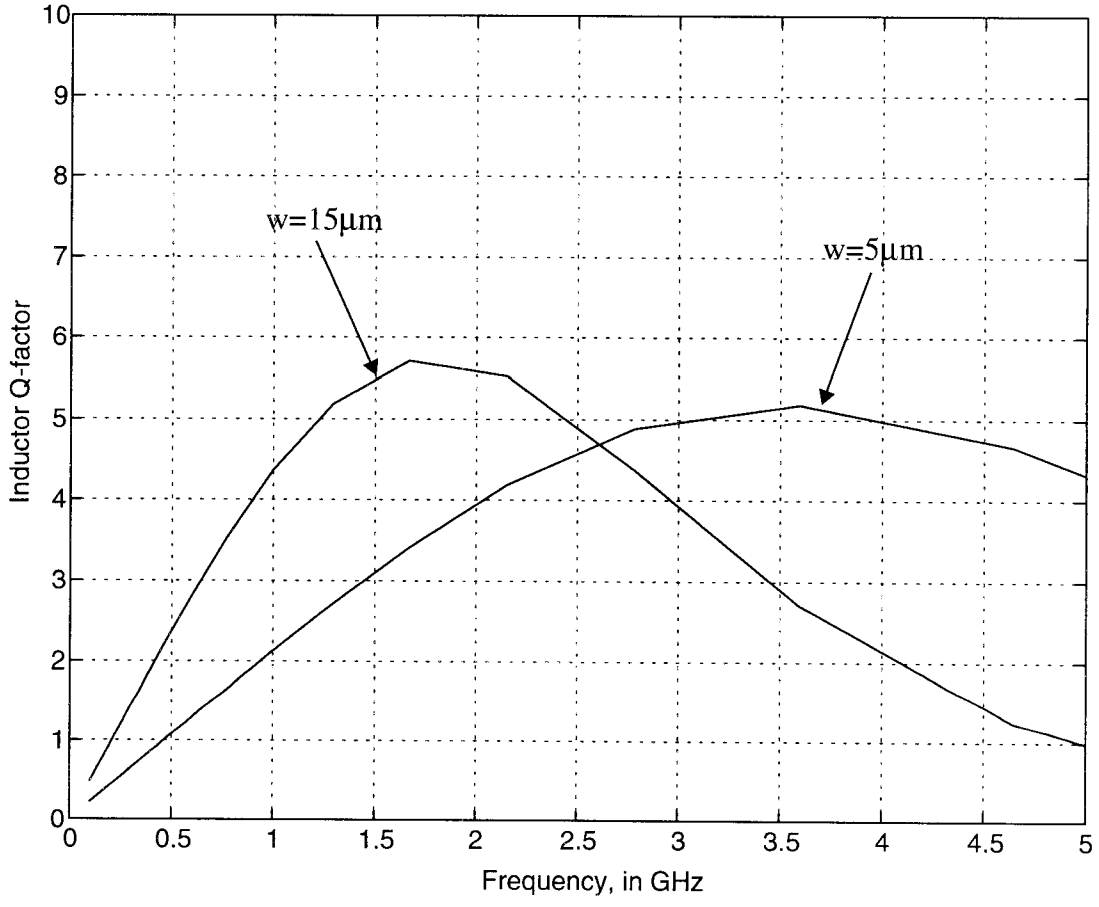


Figure 5.13: Simulated Q-factor for two inductors fabricated in a 0.8-μm BiCMOS process.

5.3.7 Summary

As demonstrated in the above simulations, the design of spiral transformers in silicon technologies involves a complex trade-off between various geometry layout parameters. To facilitate the first-phase design, some qualitative design guidelines for the optimization of silicon-based transformers are summarized below.

i) Large inductance values can be achieved by increasing the number of spiral turns. However, this results in a low self-resonant frequency limiting the applications at high frequencies.

ii) Increasing the metal width will reduce the parasitic series resistance and may improve the quality factor. However, the improvement is limited due to the skin effect and eddy current loss in the substrate. More importantly, a wider metal width results in a lower self-resonant frequency. For applications around 1GHz, a metal width of between 20-40μm is found to be best for most transformer designs fabricated with the technological parameters listed in Table 5-1.

iii) Minimizing the metal trace spacing would maximize the magnetic coupling and the capacitive coupling between microstrip segments (also the primary and the secondary). Tight coupling reduces the die area for a given inductance and thus the parasitics, except for the interline coupling capacitance which would be increased. It is found that this has only a slight impact upon the circuit performance of transformers when operated around 1GHz.

iv) Increasing the spiral center spacing will improve both the quality factor and self-resonant frequency slightly. However, the cost is the die real estate.

The above design guidelines can also be applied to the optimization of silicon-based inductors because there is no difference between the transformers and the inductors in

terms of the layout effects on the circuit performance. These optimization design guidelines are useful in the first-phase of transformer and inductor design. However, for best results, extensive computer simulations are required. In addition, these optimization guidelines should be modified accordingly if the fabrication technological parameters are different from those listed in Table 5-1.

5.4 Comparison Between Transformers and Inductors

A center-tapped spiral inductor (two coupled inductors with one end tied together) has been recently reported to have advantages versus two independent inductors [108]-[109]. In this section, we shall demonstrate that a silicon-based monolithic transformer viewed as two coupled inductors is advantageous in fully-differential applications compared to two independent inductors with the same total differential-mode inductance. More specifically, the transformer occupies less die area, exhibits smaller parasitics and higher quality factor and self-resonant frequency, and thus affords better circuit performance.

5.4.1 Simulation Results

Various monolithic inductors and transformers, fabricated in a three-metal 0.6μm digital CMOS technology as listed in Table 5-1, are modeled in order to compare their circuit performance. All geometric layout parameters except the number of spiral turns are kept unchanged for each implementation of the transformers and inductors. The operating frequency is assumed to be 900MHz. Table 5-2 shows the common geometric layout parameters used in the simulation.

Table 5-3 and Table 5-4 summarize the simulation results for three different transformers and inductors, respectively. The listed electrical parameters are corresponding to the compact circuit models of the spiral transformer and inductor, as shown in Fig. 5.5 and Fig. 4.11 respectively. Please note that only the electrical parameters for the primary inductors of the transformers are listed here. The last column of Table 5-3 and Table 5-4 shows the total trace length of the primary spirals and the inductor spirals, which is an indication of the die area.

5.4.2 Differential-Mode Circuit Performance

Traditionally, transformers are used for signal coupling or impedance transformation, while independent inductors are utilized if only inductance is required. As we shall see, compared to two independent inductors, there are several advantages to utilizing a transformer in differential circuits. As shown in Fig. 5.6, the primary or the secondary of a spiral transformer is equivalent to an inductor with effective inductance of $(1+k)L$ in differential-mode. From Table 5-3 and Table 5-4, we can see that compared to a spiral inductor with the same effective differential-mode inductance, the primary or the secondary of a spiral transformer demonstrates less parasitic series resistance and shunt capacitance and occupies less die area. Table 5-5 shows the comparison of differential-mode circuit performance between the primary inductors and the spiral inductors fabricated in the same technology. As can be seen, the primary inductors outperform the spiral inductors in differential-mode in terms of the quality factor and the self-resonant frequency, for a given effective inductance. For instance, the improvement in quality factor is about 45%, from 3.5 to more than 5, and about 12% in self-resonant frequency, for an equivalent inductance of about 9nH. In addition, the advantage of transformers increases further as the required equivalent inductance increases. This occurs because the transformer takes advantage of the strong mutual coupling effect between two coupled inductors in

Table 5-2: Common geometric layout parameters for transformers and inductors

Parameter	Value
Metal trace width	30 μm
Metal trace spacing	3 μm
Spiral center spacing	120 μm
Frequency	900 MHz

Table 5-3: Simulation results of three transformers (primary only)

L (nH)	k	R (Ω)	Cox (pF)	Cc (fF)	Rsi (Ω)	Csi (pF)	trace length (μm)
0.89	0.59	2.11	0.20	35.3	27.05	0.39	1428
2.44	0.71	4.40	0.42	79.5	13.03	0.81	2964
5.18	0.77	7.48	0.71	140.3	7.68	1.37	5028

Table 5-4: Simulation results of three inductors

L (nH)	R (Ω)	Cox (pF)	Cc (fF)	Rsi (Ω)	Csi (pF)	trace length (μm)
1.41	2.58	0.25	19.0	21.90	0.48	1764
4.17	5.49	0.53	45.9	10.48	1.01	3688
9.17	9.39	0.89	82.2	6.17	1.71	6258

differential-mode. Therefore, to achieve the same equivalent inductance in differential-mode, less metal trace length may be used in a transformer. Consequently, there are less parasitics and better circuit performance is achieved. Furthermore, due to the differential signals, the substrate under the transformer can be considered as a virtual ground as shown in Fig. 5.6. As a consequence, the substrate parasitics R_{si} and C_{si} can be ignored (note the eddy current loss in the substrate is not affected and is included as part of the series resistance R). This further improves the differential-mode circuit performance of the transformer.

5.4.3 Common-Mode Circuit Performance

A transformer provides additional common-mode rejection for the differential applications. As shown in Fig. 5.14, for instance, the transformer provides the required equivalent inductance of about 9nH for an LC tuning network of the LNA circuit in

Table 5-5: Differential-mode circuit performance

		effective inductance (nH)	quality factor	self-resonant frequency (GHz)
I	transformer	1.41	3.74	8.65
	inductor		3.03	8.25
II	transformer	4.17	4.99	3.48
	inductor		3.87	3.15
III	transformer	9.17	5.16	1.80
	inductor		3.53	1.61

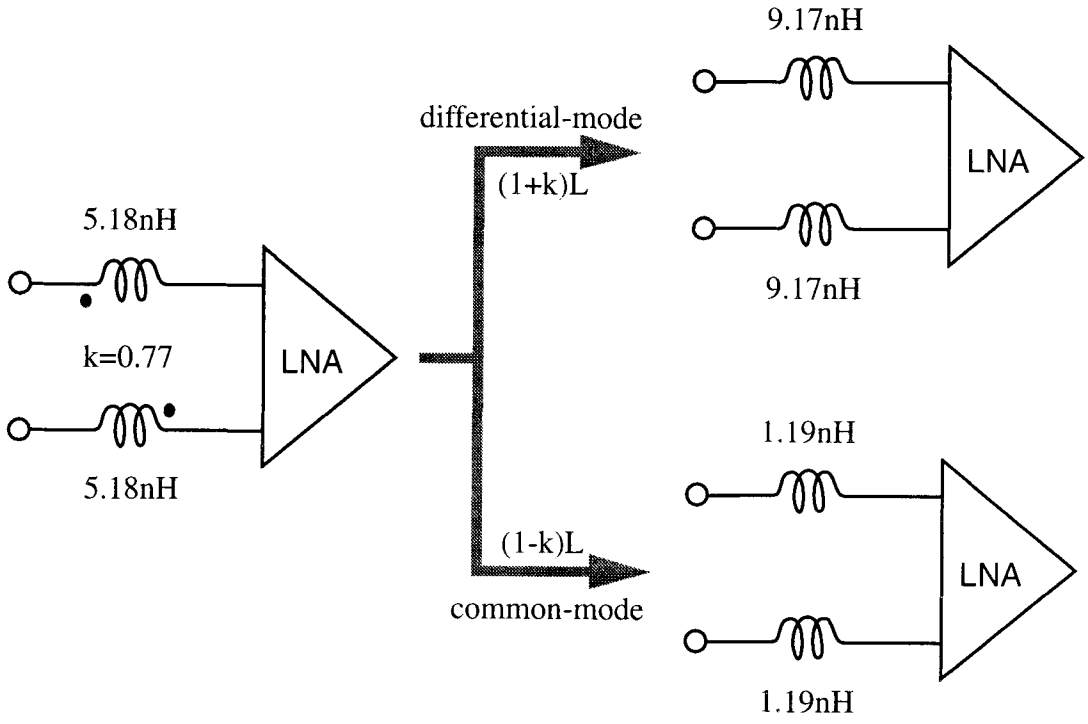


Figure 5.14: A spiral transformer in a differential LNA circuit.

differential-mode, achieving a higher quality factor and self-resonant frequency compared to two independent inductors. While in common-mode, the equivalent inductance of the transformer is decreased to *(1-k)L*, which is about only 1nH. Hence, the common-mode LC tuning network is effectively de-tuned at the frequency of interest, which significantly reduces the common-mode gain of the LNA circuit. This leads to a higher common-mode rejection. Note that two independent inductors are unable to provide this advantage because their inductance remains the same in both differential-mode and common-mode.

In addition, because of the symmetric inter-winding layout of the transformer, substrate noise coupling through the parasitic oxide capacitance is more likely to be seen

as a common-mode signal by the transformer. This would result in higher substrate noise rejection for the differential circuits. Nevertheless, additional experimental investigation is required to provide further verification of this feature.

5.5 Conclusions

We have described the characterization and modeling of spiral transformers in silicon technology. Extensive simulations have been conducted to investigate the circuit performance of the transformers with layout parameter variations. Design guidelines for transformer optimization are also summarized based on the simulation results. Comparisons between transformers and inductors are given to demonstrate several advantageous features provided by transformers.

When utilized in a differential application, it is found that a transformer occupies less die area and thus has less parasitic series resistance and shunt capacitance, compared to two independent inductors with the same equivalent differential-mode inductance. As a consequence, the quality factor and the self-resonant frequency are improved. The improvements become more significant as the required equivalent inductance increases. In addition, a transformer provides additional common-mode rejection for the differential circuits.

As we know, high quality monolithic inductors present a great challenge to the full integration of RF circuits. For differential RF IC's, a spiral transformer provides a promising solution because of its proven advantages. Used as a substitute for two separate inductors wherever applicable, e.g., in differential amplifiers, mixers, or band-pass filters, a spiral transformer improves the inductance quality and thus the overall circuit performance.

CHAPTER 6. CIRCUIT IMPLEMENTATION AND EXPERIMENTAL RESULTS

Taking advantage of the transformers described in last chapter, a 900MHz fully-differential LC tuned LNA has been implemented in a standard digital 0.6μm CMOS technology available through MOSIS. In this chapter, the complete circuit implementation and layout description of the CMOS LNA are illustrated. The LNA utilizes three monolithic transformers in on-chip tuning and matching networks. Bias current re-use is used to reduce the power dissipation and process-, voltage-, and temperature-tracking biasing techniques are employed. The experimental results are presented as a conclusion[1].

6.1 Complete Circuit Implementation

6.1.1 Basic Input Circuit

Fig. 6.1 illustrates the input circuit of the LC tuned CMOS LNA. As discussed in Chapter 3, inductances are required to form series resonant networks with the gate-source capacitances of the input transistors M1-M2 so that a minimum noise figure can be achieved. Spiral transformer T1 is used to provide the required inductances at the differential input gates, taking advantage of the transformer's smaller series resistance and shunt parasitics as compared to a conventional design using two independent spiral inductors. As has been demonstrated in Chapter 5, a spiral transformer has a higher quality factor in differential applications and thus better circuit performance of the LC tuned CMOS LNA can be achieved.

[1] Portions of this chapter are extracted from: J. J. Zhou and D. J. Allstot, "A fully-integrated CMOS 900MHz LNA utilizing monolithic transformers," in *ISSCC Digest of Technical Papers*, Feb. 1998, pp. 132-133. Copyright 1998 IEEE. Reprinted with permission.

Transistors M1-M4 form a cascode input stage which improves the reverse isolation of the LNA. As shown, the reverse signal path in the cascode stage contains the drain-source capacitance C_{ds} of M3 (or M4) and the gate-drain capacitance C_{gd} of M1 (or M2). Since C_{ds} is usually much smaller than C_{gd}, higher reverse isolation is achieved as compared to an input circuit without cascoded transistors in which the reverse signal path contains only C_{gd}. Another benefit of the cascode configuration is the reduced Miller effect on the input capacitance. In the cascode configuration, M1(or M2) is a common-source (CS) stage which has a large current gain and a small voltage gain while M3 (or M4) is a common-gate (CG) stage which has a unity current gain and a relatively large voltage gain.

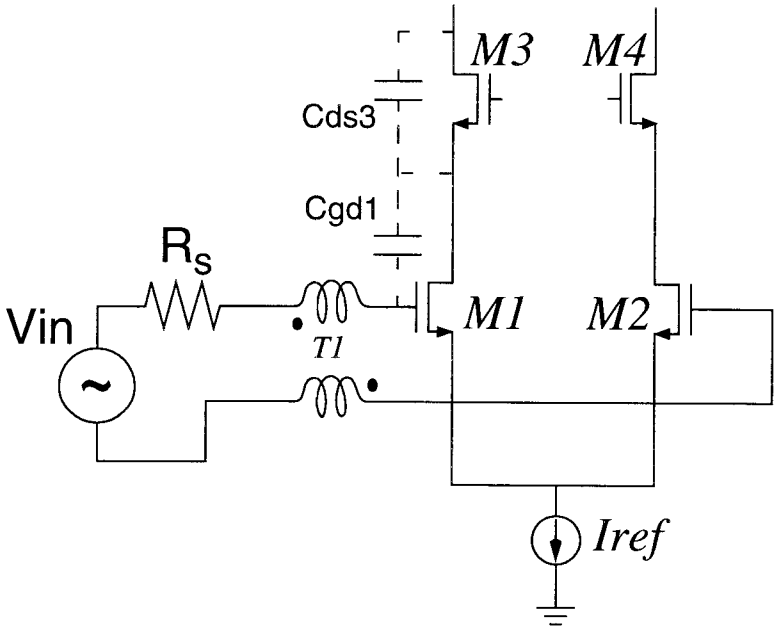


Figure 6.1: Cascode input circuit of the LC tuned CMOS LNA.

Assuming that the total voltage gain of the input circuit is designed to be 20dB, it is not difficult to show that the voltage gain of M1 (or M2) is approximately $-g_{m1}/g_{m3}$. Therefore, the input Miller capacitance is about $(1 + g_{m1}/g_{m3})C_{gd1}$, compared to $11C_{gd1}$ if the input circuit comprises only CS stage M1 (or M2). This advantage is significant because Miller capacitance shunts the input RF signal and degrades circuit performance.

The cascoding transistor M3 (or M4) contributes additional noise to the circuit. However, since the impedance seen at the drain of M1 (or M2) is relatively high, about $1/g_{ds1}$ at low frequencies and $1/j\omega C_{gs3}$ at high frequencies, the channel thermal noise contribution from M3 (or M4) is small compared to that of M1 (or M2). In addition, the gate of M3 (or M4) is at ac ground and thus the induced gate current noise of M3 (or M4) is negligible.

6.1.2 Low Power Techniques

An LNA must provide power gain to the incoming small signal without over-driving the following down-conversion circuits. If directly driving a 50Ω resistive load, the input circuit, as shown in Fig. 6.1, can only achieve a voltage gain of about $25g_{m1}Q$, where Q is roughly 4.5 for an optimum design (refer to Chapter 3). Even if M1 is biased at a large current so that g_{m1} can be as large as $0.03\Omega^{-1}$, the voltage gain will merely be about 10dB. Therefore, a driving stage is needed for the resistive load (50Ω) to achieve a moderate power gain (typically 10-20dB). Fig. 6.2 shows a two-stage differential CMOS LNA design. It comprises an input stage formed by transformer T1 and M1-M4 (identical to that shown in Fig. 6.1), an interstage transformer T2, and a driver stage formed by M5-M6 and transformer T3. Transformer T3 provides a dc path to the supply and tunes out the total output capacitance so that the LNA is capable of driving an off-chip 50Ω load.

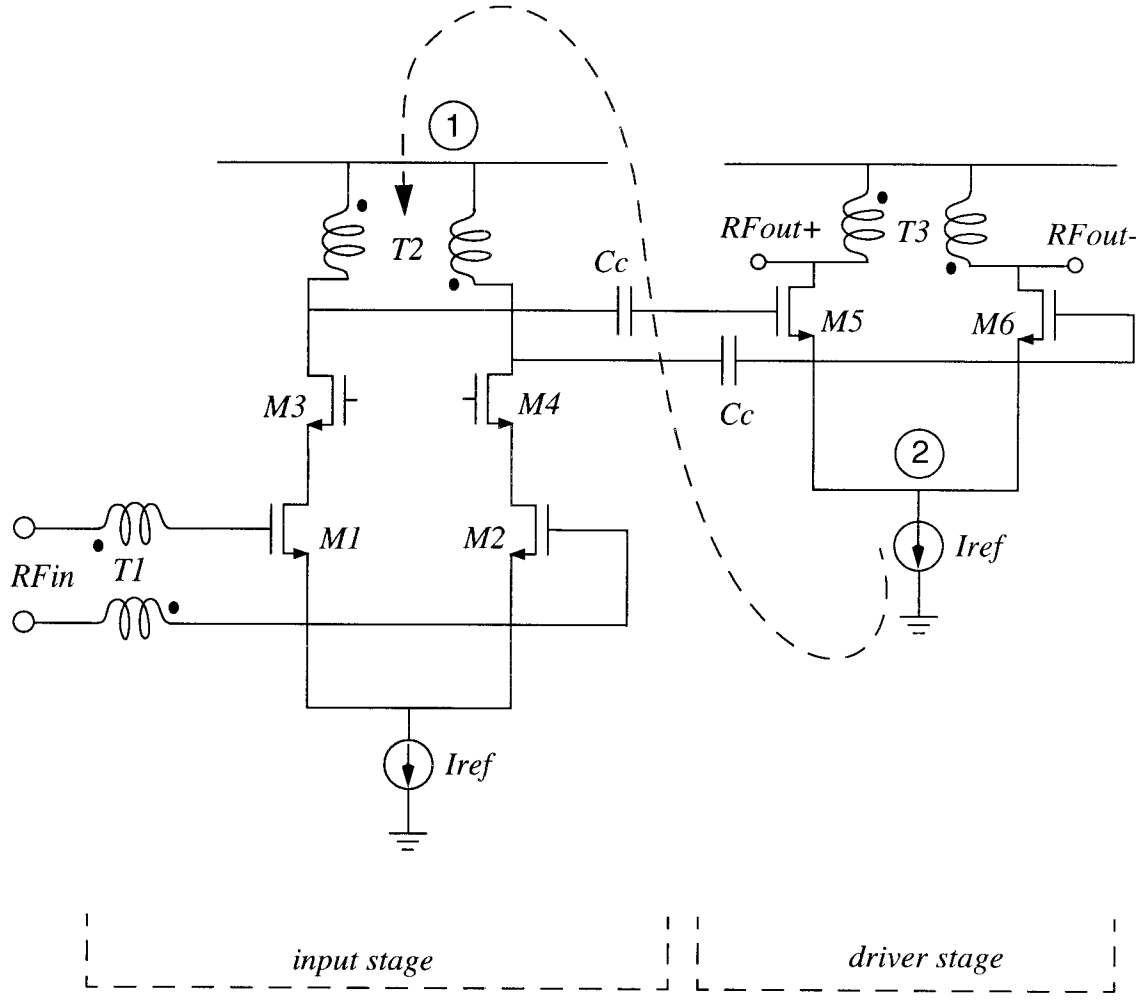


Figure 6.2: Two-stage LC tuned CMOS LNA.

As discussed in Chapter 3, an LNA usually consumes a large amount of power in a receiving system because a large bias current is required to achieve low noise performance. This not only increases the system cost but also causes excessive heat which consequently reduces the effective g_m and increases the device noise temperature. To reduce power consumption, a bias current re-use technique may be employed at a cost of reduced voltage headroom [10], [110]. As can be seen from Fig. 6.2, both nodes 1 and 2 are ac grounds. By stacking the driver stage upon the input stage, the two stages share the bias current I_{ref}, effectively reducing the total power consumption while still maintaining the large bias current needed for low noise and high power gain.

The stacked circuit schematic of the CMOS LNA is shown in Fig. 6.3. The output driver is a PMOS source-follower pair M5-M6, changed from the NMOS common-source pair as shown in Fig. 6.2, with transformer T3. Though PMOS has lower g_m than NMOS with the same bias current, and a source-follower does not achieve as high of a voltage gain as a common-source amplifier, this implementation reduces circuit complexity by allowing direct dc coupling between the input and output stages. It eliminates the need for on-chip coupling capacitors C_c as shown in Fig. 6.2, which saves die area and avoids potential signal losses through the capacitive substrate parasitics. It also eliminates the need for a biasing circuit for M5 and M6. The gain loss of the PMOS source-follower pair can be compensated by increasing the gain of the input stage. Since the load of the input stage is a parallel resonant LC circuit formed by transformer T2, the gate capacitances of M5-M6, and the drain capacitances of M3-M4, large voltage gain is easily obtained.

The interstage transformer T2 serves two purposes in the circuit: First, it forms the parallel resonant LC circuit to develop the necessary voltage gain for the LNA. Second, it also acts as a high impedance for ac and a very low impedance for dc signals which makes the re-use of bias current feasible.

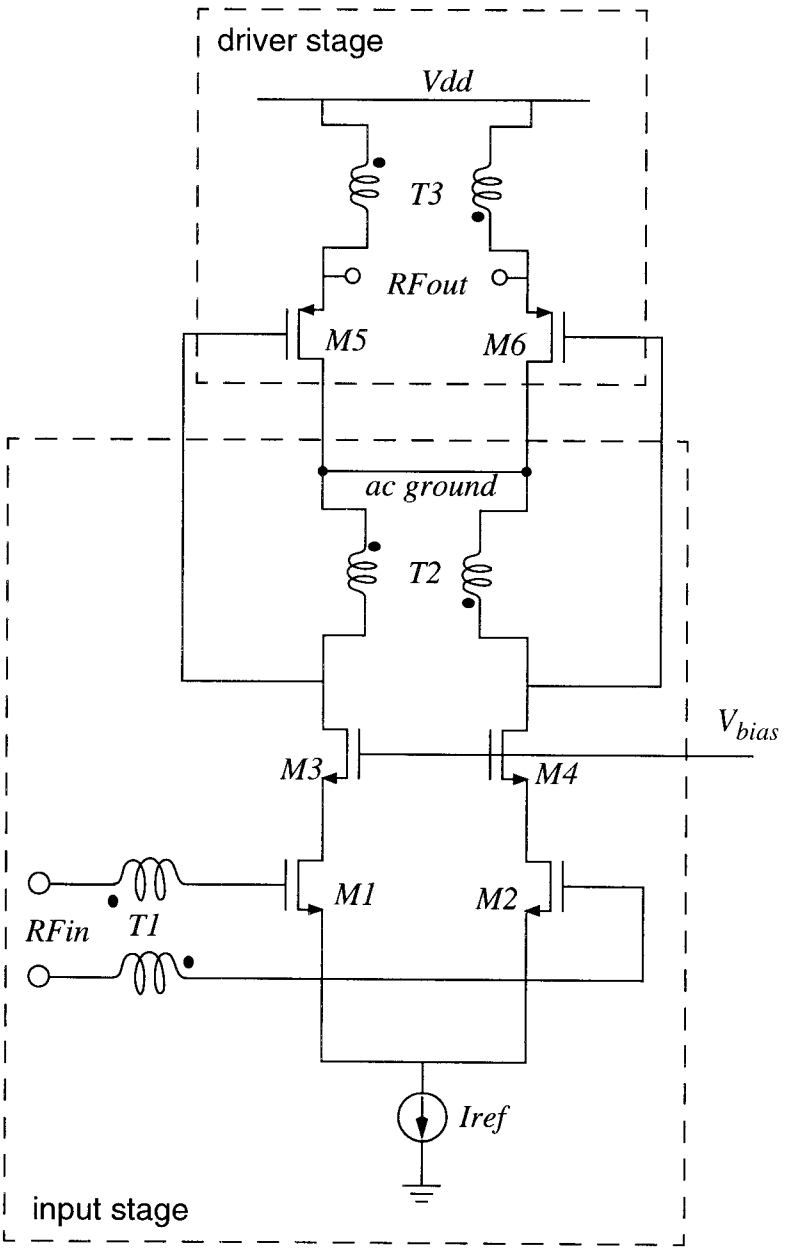


Figure 6.3: Stacked LC tuned CMOS LNA.

6.1.3 Biasing Circuit

The complete circuit schematic of the LC tuned CMOS LNA, including the biasing circuit, is shown in Fig. 6.4. The drains of transistors M5 and M6 are connected to one port of transformer T2 which is at ac ground with a dc voltage of V_{dd}-$V_{sg}(M6)$ (dc drops in T2 and T3 are almost the same and yet very small). This dc potential serves as the gate bias voltage for M1 and M2 through resistors R_{b1} and R_{b2}. Resistors R_{b1} and R_{b2} are chosen to be large enough (e.g., 40kΩ) so that they block the incoming RF signal from going to the ac ground and contribute negligible resistive thermal current noise to the circuit.

The biasing circuit consisting of transistors M22, M44 and M66 is designed to track process, voltage, and temperature variations in generating the gate bias voltage V_{bias} for cascoding transistors M3 and M4. To accomplish this goal, the size of the transistors is chosen to be

$$(W/L)_{66} = \frac{1}{n}(W/L)_6,\ (W/L)_{22} = \frac{1}{n}(W/L)_2 \tag{6.1}$$

where n is an arbitrary integer number. Also note that $V_{sg}(M66)=V_{sg}(M6)$ and $V_{gs}(M22)=V_{gs}(M2)$. Therefore, the biasing circuit consumes approximately $1/2n$ of the total bias current and by selecting a large value for n, the power consumed in the biasing circuit is negligible. Setting $(W/L)_{44} = \frac{1}{n}(W/L)_4$, we have $V_{gs}(M4)=V_{gs}(M44)$ and $V_{ds}(M2)=V_{ds}(M22)$. The bias voltage V_{bias} for M3 and M4 is then given by

$$\begin{aligned} V_{bias} &= V_{ds}(M7) + V_{ds}(M22) + V_{gs}(M44) \\ &= V_{ds}(M7) + V_{ds}(M2) + V_{gs}(M4) \end{aligned} \tag{6.2}$$

where

$$V_{ds}(M7) = Vdd - V_{sg}(M6) - V_{gs}(M2) \tag{6.3}$$

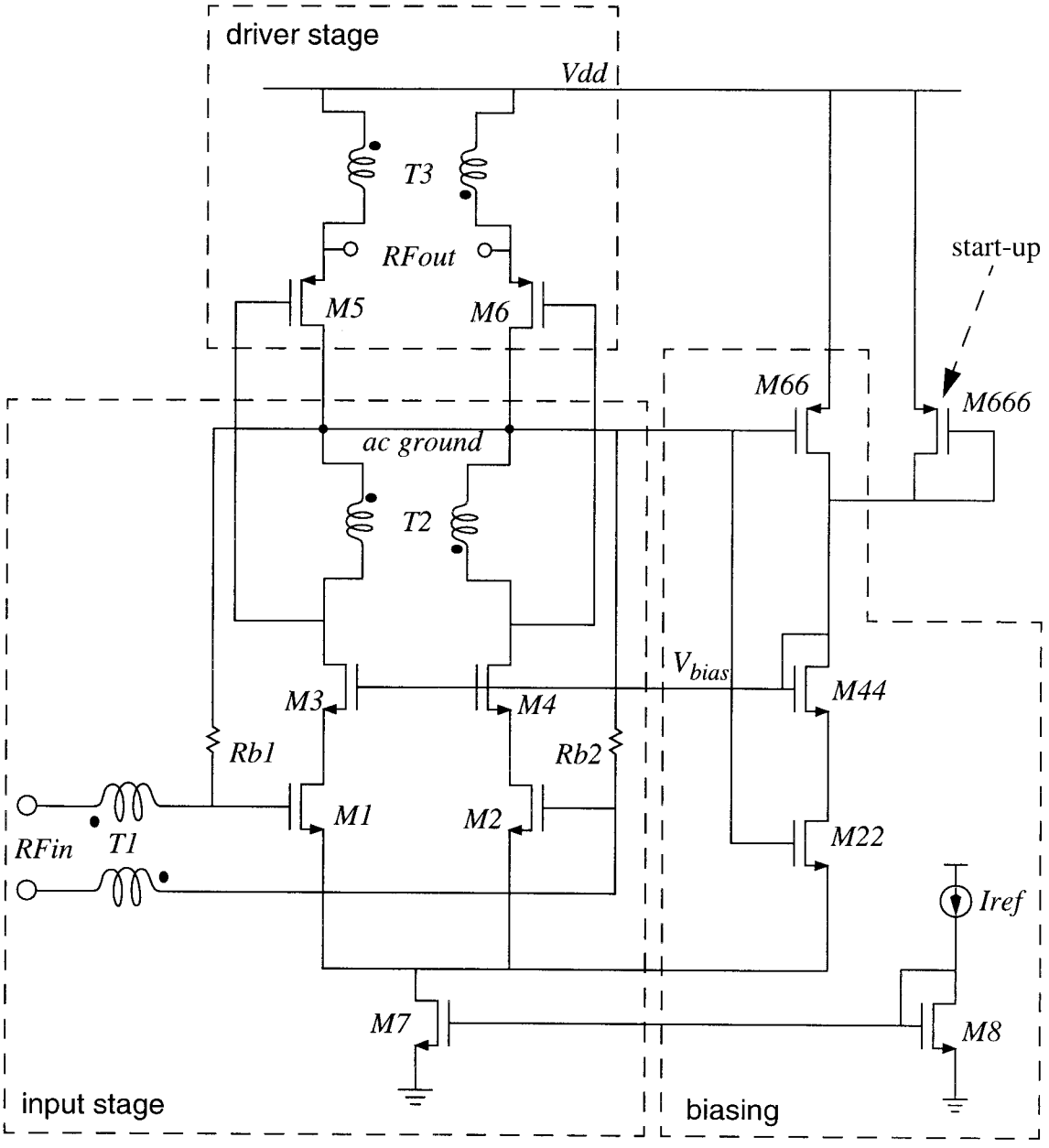


Figure 6.4: Complete circuit schematic of the LC tuned CMOS LNA.

It can be seen from Equations (6.2) and (6.3) that the bias voltage V_{bias} tracks power supply and transistor threshold voltage (process and temperature) variations with $V_{ds}(M7)$ and $V_{gs}(M4)$.

It is likely that V_{bias} is near ground and M66 is off during start-up. Transistor M666 is used to boost V_{bias} during start-up and thus guarantee a reliable turn-on of the circuit. After start-up, M666 is turned off and thus has no impact on the circuit performance.

6.1.4 Device Parameters

The CMOS LNA is designed to operate at 900MHz with a power supply voltage of 3V. To determine the device parameters, we assume the total power consumption is about 18mW which implies a total bias current of about 6mA, ignoring the power consumed in the biasing circuit.

Based on the design procedures summarized in Chapter 3, we can easily determine the optimal values for device parameters. Each MOSFET should have the minimum 0.6μm drawn channel length, with 0.51μm and 0.53μm effective channel length for NMOS and PMOS device respectively (data from MOSIS). The optimal device width for M1 and M2 can be chosen based on Equation (3.39). Please note that the effective source impedance for the differential inputs is $R_s/2$=25Ω. Assuming the gate inductance has a parasitic series resistance of about 6Ω, the optimum device width is then given by

$$W_{opt} \approx \frac{1}{3\omega R' C_{ox} L} = \frac{1}{3 \times 2\pi \times 0.9 \times 10^{9} \times (25+6) \times 3.8 \times 10^{-15} \times 0.51} \approx 981(\mu m) \tag{6.4}$$

where C_{ox} is 3.8fF/μm^2 (data from MOSIS).

Having determined the device width for M1 (or M2), we can select the inductance value for transformer T1 based on Equation (3.21) (assuming $\beta/\gamma=2$ and $c=0.35$j):

$$\begin{aligned} L_g &\approx \frac{0.6628}{(\omega^2 C_{gs})} \\ &= \frac{0.6628}{(2\pi \times 0.9\times 10^9)^2 \times 2/3 \times 981 \times 0.51 \times 3.8\times 10^{-15}} \\ &= 16.3(\text{nH}) \end{aligned} \tag{6.5}$$

This inductance is unrealistic for silicon-based spiral transformers, much less spiral inductors. To make a compromise, we may increase the width of M1 (or M2) since the noise figure is fortunately not very sensitive to small variations of the optimal device width. If the width of M1 (or M2) is chosen to be 1080μm, a bit larger than the optimal value of 981μm, the inductance value would be about 14.8nH, still too large to be realized using spiral transformers. However, taking the parasitic capacitance of the spiral transformer into consideration, the input capacitance of the LNA would be larger than C_{gs} and thus a smaller inductance is required to form the resonance specified in Equation (3.21). The input Miller capacitance further increases the input capacitance. Simulation in HSPICE using the transformer model established in Chapter 5 has shown that a transformer with series inductance 9.17nH (and series resistance of 7.48Ω) achieves a minimum noise figure at 900MHz when the input device is 1080μm wide. However, although the parasitic capacitance of the transformer lowers the requirement for a large inductance value L_g, it degrades the noise performance of the amplifier. As a consequence, the simulated minimum noise figure, which is about 3dB, is significantly higher than the theoretical minimum noise figure of about 2.4dB based on Equation (3.43).

The cascoding transistors M3 and M4 are chosen to have widths of 420μm. Larger width would cause an increase in the noise contribution from M3 and M4 due to the increase in C_{gs3} and C_{gs4}, which consequently decreases the impedance seen at the drains of M1 and M2. However, smaller width increases the voltage gain of M1 (or M2) and thus

the input Miller capacitance. M5, M6 and M7 are chosen to have the same width as M1 and M2. This choice is somewhat arbitrary but their large widths help for low voltage design. Refer to the LNA circuit shown in Fig. 6.4. We observe

$$Vdd \approx V_{gs}(M5) + V_{gs}(M1) + V_{ds}(M7) \tag{6.6}$$

It is clear the voltage headroom is improved by increasing the device size. Biasing transistors M22, M44, and M66 are chosen to be 1/40 of M2, M4 and M6, respectively. Therefore the current consumed in the biasing circuit is only about 3mA/40. The size of start-up transistor M666 is arbitrarily chosen to be 36/0.6. The resistors R_{b1} and R_{b2} are chosen to have a large value of 40kΩ, as discussed before.

Table 6-1 lists the design parameters of the LNA components. All three transformers are laid out on the topmost metal layer (metal 3). Geometry parameters for the transformers are listed in Table 5-2, except for the center hole dimension of transformers T2 and T3, which is larger (180μm). Note that the inductances of transformers T2 and T3 are limited by the self-resonant frequency and parasitics, although theoretically larger inductance are required for resonance.

6.1.5 Layout Issues

Special cautions need to be taken for layout in the design of the CMOS LNA. Three transformers should be separated as much as possible to minimize the potential interactions between them. Finger-gate structures are used for the layout of the large-size transistors M1-M7 to minimize the noise contributed by the gate resistance (see Sec. 2.4). The wide gates of transistors M1-M7 are partitioned into 40 fingers with each finger width listed in Table 6-1. For transistors M1-M2 (W/L=1080/0.6) with 40 finger gates, the total equivalent gate resistance would be 0.375Ω if the poly-gate sheet resistance is 4Ω/square,

Table 6-1: Design parameters of the LNA components

Parameters	Value
$(W/L)_1$, $(W/L)_2$	40 X (27/0.6)
$(W/L)_3$, $(W/L)_4$	40 X (10.5/0.6)
$(W/L)_5$, $(W/L)_6$	40 X (27/0.6)
$(W/L)_7$	40 X (27/0.6)
$(W/L)_{22}$	(27/0.6)
$(W/L)_{44}$	(10.5/0.6)
$(W/L)_{66}$	(27/0.6)
$(W/L)_{666}$	(36/0.6)
Rb1, Rb2	40kΩ
transformer T1	Leff=9.17nH
transformers T2, T3	Leff=11.83nH

compared to 600Ω with only one finger gate. Therefore, the thermal noise associated with the gate resistance is negligible.

To avoid substrate noise coupling to the RF circuits through the bonding pads, usually a grounded metallic plate underlying the pad oxide is used to short the substrate noise to ground [111]. In our case, the differential input pads have N+ diffusions below them to form a virtual ground so that any substrate noise coupling into the pads is presented as a common-mode signal that is rejected by the differential circuit.

6.2 Simulation Results and Sensitivity Analysis

HSPICE simulation results of the transformer-tuned CMOS LNA shown in Fig. 6.4 are summarized in Table 6-2. To demonstrate the advantages provided by the transformers, the simulation results of an inductor-tuned LNA with the same device parameters as listed in Table 6-1 are also listed. The transformer and inductor models used in HSPICE simulations were computed using the approach described in Chapters 4 and 5. It is seen that the transformer-tuned LNA outperforms the inductor-tuned LNA, especially in the noise figure and the power gain.

We have shown that transformers are advantageous in the LC-tuned CMOS LNA design compared to inductors because transformers have less parasitics (R, C_{ox}, C_c, etc.) for a given effective inductance as demonstrated in Chapter 5. This fact indicates that the

Table 6-2: Simulation Results of Transformer- and Inductor-Tuned CMOS LNA's

	Transformer-Tuned CMOS LNA	Inductor-Tuned CMOS LNA
Supply voltage	3 V	3 V
Power dissipation	18 mW	18 mW
Frequency	900 MHz	900 MHz
Noise figure	3.0 dB	3.2 dB
S21	15.0 dB	13.5 dB
S12	-39.0 dB	-37.4 dB
S11	-7.6 dB	-8.1 dB
1-dB compression (input)	-17.3 dBm	-15.7 dBm

LNA performance is sensitive to the variations of the transformer and the inductor parasitics. For example, the series resistances of transformers T1-T3 have significant effect on the power gain (S21) of the LNA. For a first-order approximation, the LNA gain is proportional to the quality factor of the input gate capacitance which is given by

$$Q = \frac{1}{\omega(R_s + R_{L1})C_{gs}} \tag{6.7}$$

where R_s is the source impedance (25Ω for the half circuit of the differential LNA) and R_{L1} is the series resistance of the primary or the secondary of transformer T1. The LNA gain is also proportional to the output impedance which is significantly affected by the parasitic resistances and capacitances of transformers T2 and T3. The sensitivity of the LNA gain to the parasitics R_L, C_{ox}, and C_c of transformers T1-T3 are simulated and demonstrated in Fig. 6.5, where the effects of positive and negative variations of 20% from the nominal values of R_L, C_{ox}, and C_c of transformers T1-T3 are shown. Note that for transformer T1, the nominal values of R_L, C_{ox}, and C_c are 7.48Ω, 0.71pF, and 140.3fF, respectively.

It is seen from Fig. 6.5 that the LNA power gain is most sensitive to the variation of parasitic resistances of T1-T3 and has an 8% deviation from the nominal 15dB power gain when parasitic resistances of T1-T3 are varied by 20%. Oxide capacitances of T1-T3 also have a significant effect on the LNA performance. However, the effect of the interline coupling capacitances of T1-T3 is relatively small. From these observations, it is concluded that accurate modeling of monolithic transformers and inductors, especially the estimation of the series resistances and oxide capacitances, is crucially important to predict the performance of the LC-tuned CMOS LNA. Since the circuit parameters of monolithic transformers and inductors are sensitive to some process variations, such as metal resistivity and oxide thickness variations, it is expected that the performance of the LC tuned CMOS LNA would suffer severely from these process variations.

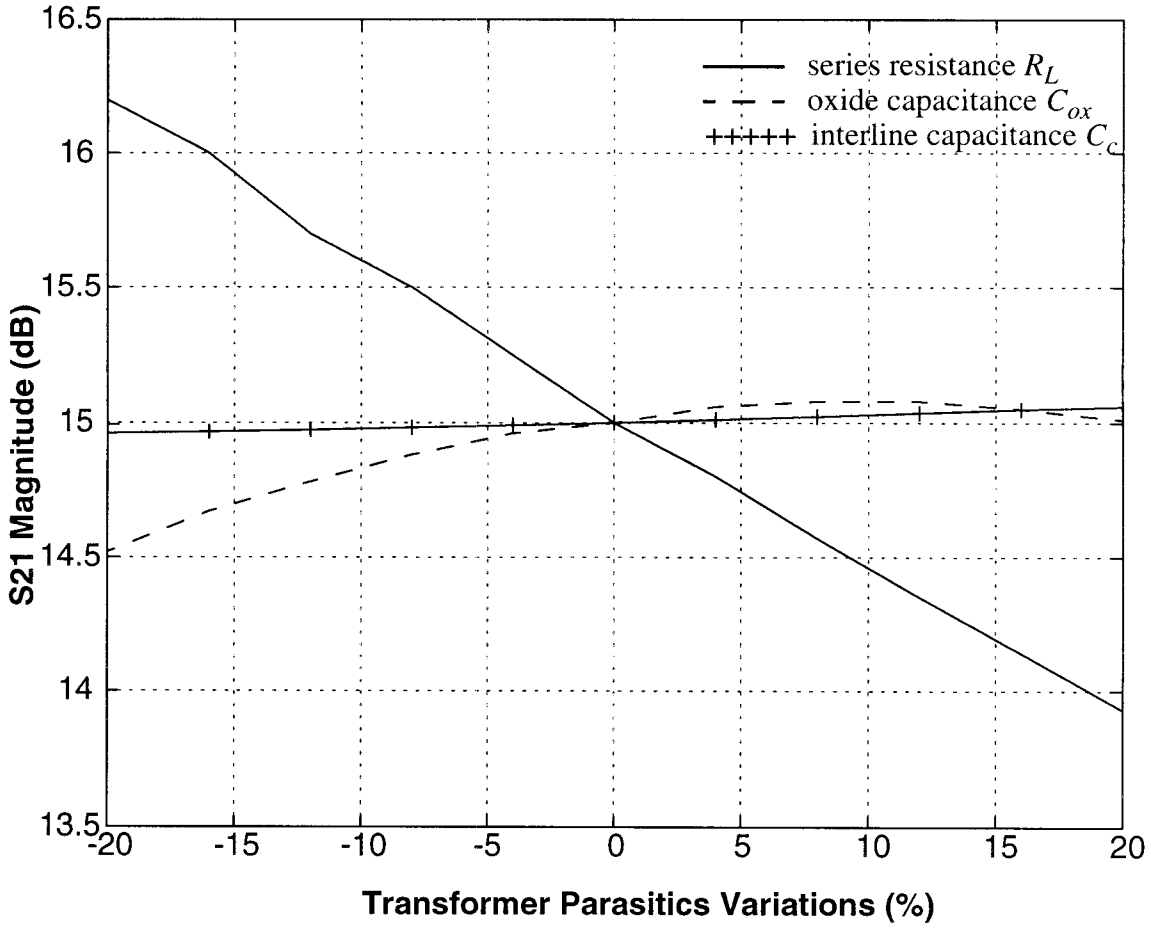


Figure 6.5: Sensitivity simulations of the LC tuned CMOS LNA.

6.3 Experimental Results

Fig. 6.6 shows the chip micrograph of the 900MHz LNA integrated in a standard digital 0.6μm CMOS process available through MOSIS. To exclude the package effects on performance, the tests were conducted with the die directly attached to a test board using

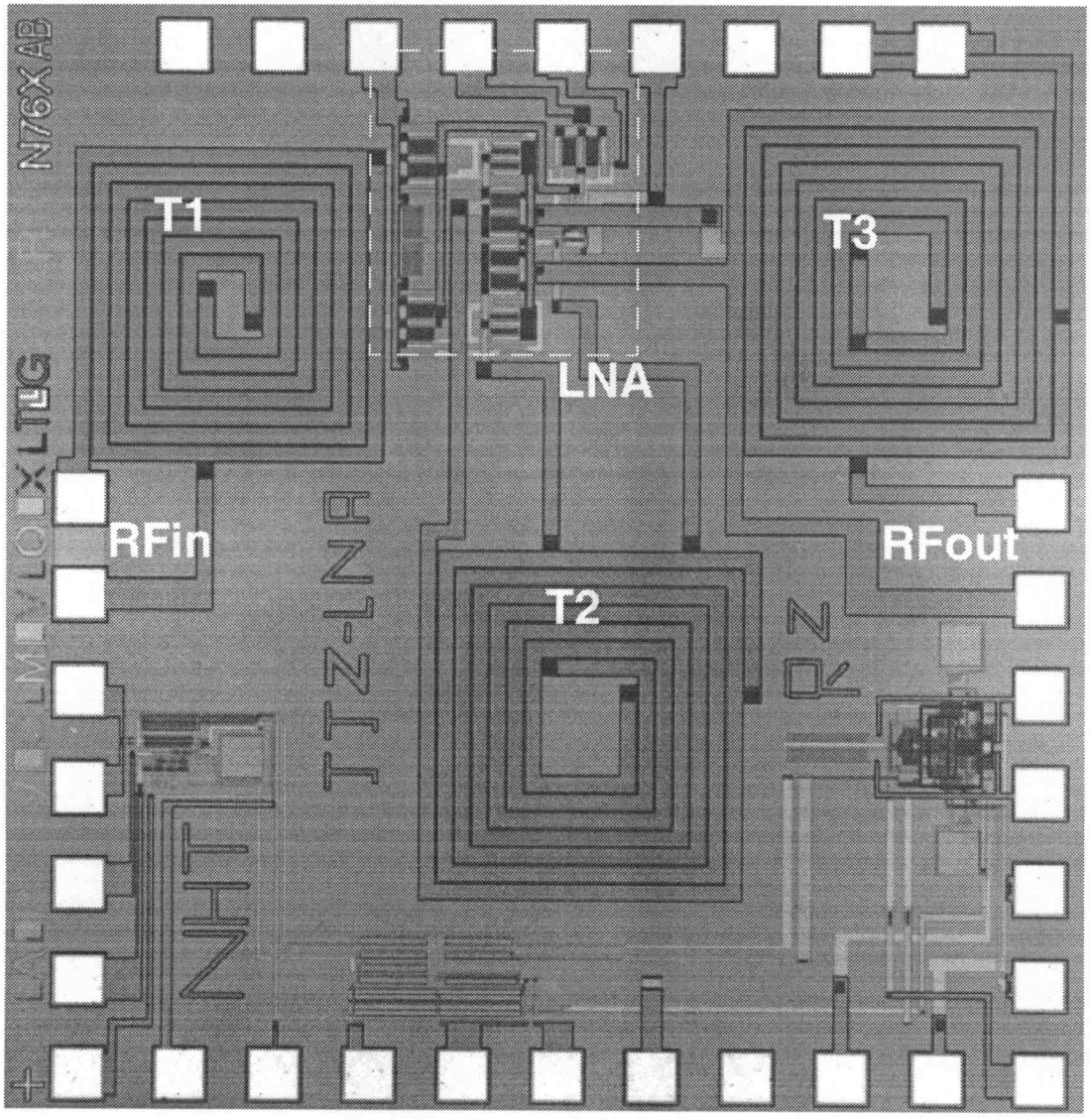


Figure 6.6: Chip micrograph of the CMOS LNA.

pad-to-board wire bonding. External RF baluns were used at the LNA input and output to perform the necessary single-ended/differential conversions.

The measured noise figure of the LNA is 4.1dB at 900MHz, higher than the simulated results in HSPICE, which is about 3dB. This is partly explained by the fact that the actual third-layer metal resistance, measured at 50mΩ/square (0.06Ω-μm), is 43% higher than the data (0.042Ω-μm) we obtained through MOSIS and used for the simulation, as listed in Table 5-1. The series resistance in transformer T1 is measured at about 10.7Ω, compared to 7.48Ω as simulated. This corresponds to an increase of about 0.13 in the amplifier's noise factor, even not considering the consequent deviation from the optimum condition. With this metal resistance, HSPICE simulation shows the LNA has a noise figure of 3.3dB. The hot carrier effects and other short-channel effects, which accordingly increase the noise coefficients, γ and β, could be attributed to the remaining difference of 0.8dB between the measured and the simulated noise figure. Also other higher-order effects such as substrate distributed resistance and balun losses may further degrade the measured noise performance. Because the noise model implemented in HSPICE does not include any of these effects, the simulated result is expected to be a bit too good. Some simple calculations based on Equation (3.43) show how the noise performance suffers from these effects. For example, taking the transformer's series resistance of 10.7Ω into consideration, the theoretical minimum noise figure of the CMOS LNA increases from 2.7dB to 4.4dB if γ increased from 2/3 to 2, and to 3.9dB if γ increased from 2/3 to 1.5. Unfortunately, we do not have enough knowledge to accurately predict γ, other than some previous experimental studies [23]. Further exploits in depth on the physical nature of short-channel devices need to be performed for CMOS RF applications as the current technology continues scaling down.

The measured forward power gain (S21) and reverse isolation (S12) of the CMOS LNA are shown in Fig. 6.7. The forward power gain achieves 12.3dB at 900MHz while the reverse isolation has a value of -33.0dB. The curve of the forward power gain clearly shows a bandpass characteristic with a peak value of 13.5dB at the center frequency around 880MHz. Worthy of mention is that the bandpass curve demonstrates a quality factor of about 6.3 while the transformer we designed achieves a simulated quality factor of 5.2. This

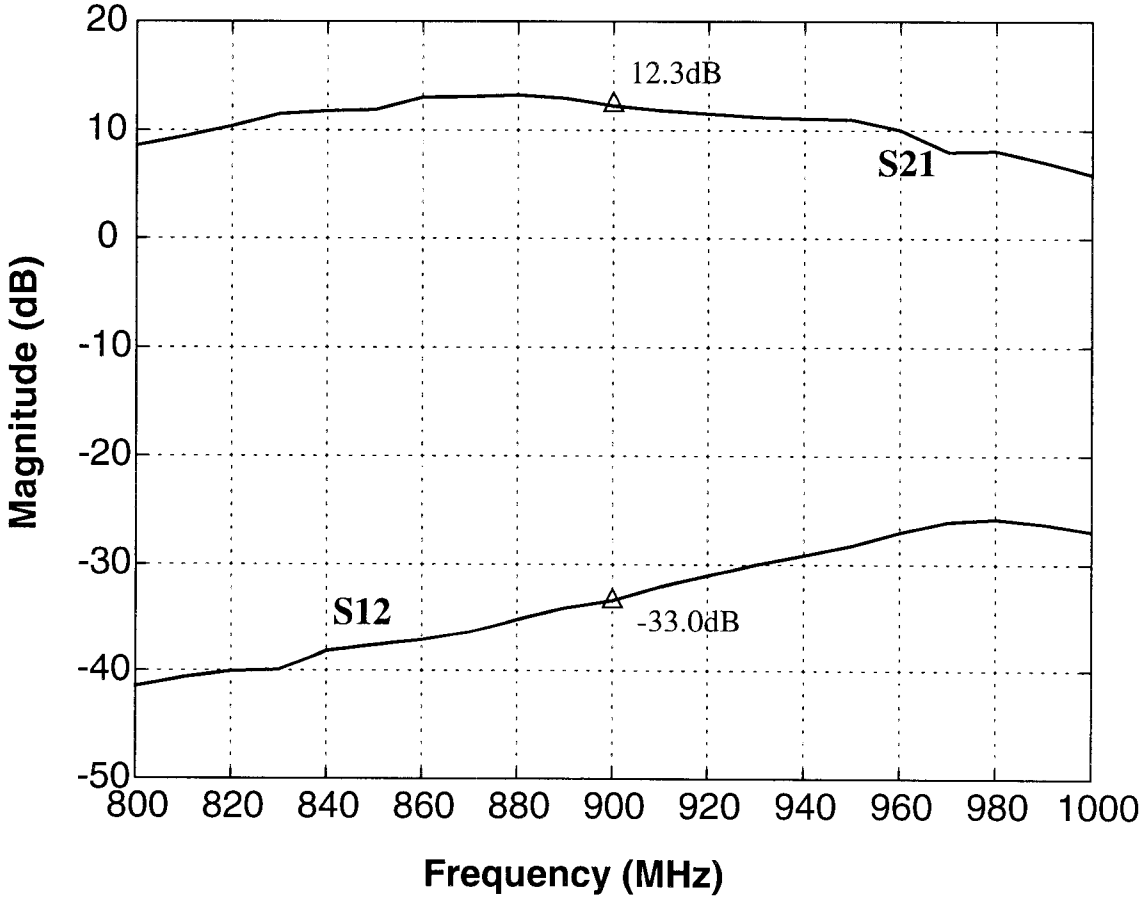


Figure 6.7: S21 and S12 measurements.

indicates the three transformer-tuning networks, which can be viewed as in cascade, increase the selectivity of the LNA circuit.

The measured input reflection coefficient S11 is -6dB at 900MHz. As discussed in Chapter 3, the input impedance of the LC tuned CMOS LNA contains a resistive term, about $R_L+1/5g_m=17\Omega$, and a capacitive term. For perfect input matching, inductive source

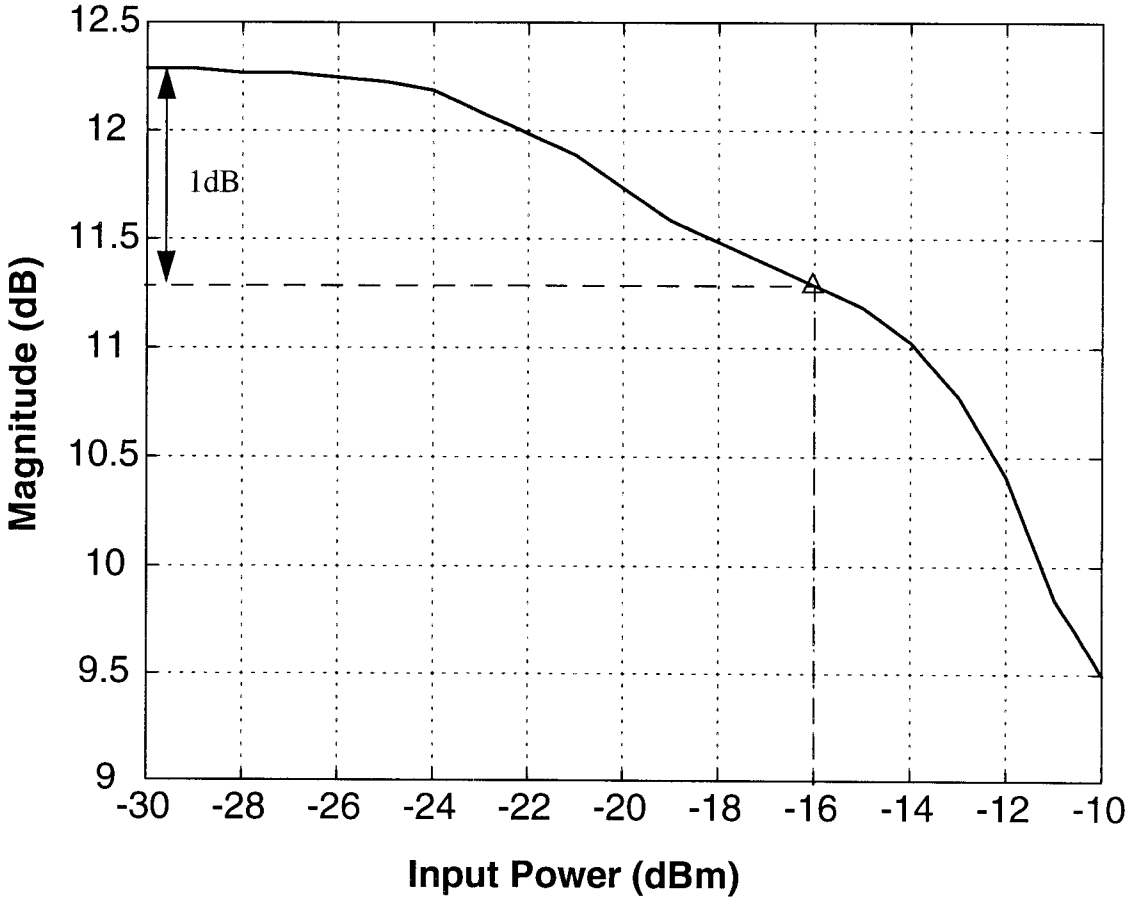


Figure 6.8: 1-dB compression point measurements.

degeneration may be employed to increase the resistive term to 25Ω (the effective source impedance for half circuit of the differential LNA). However, this will cause a power gain loss of about 6dB. Also it is difficult to eliminate the capacitive term of the input impedance due to the presence of inductor parasitics.

Fig. 6.8 shows the measured 1-dB compression point at 900MHz. The power gain of the LNA drops by 1 dB to 11.3dB at the input power of about -16dBm. As discussed in Chapter 3, this indicates the input-referred third-order intercept point (IIP3) of the LNA is about -6dBm or higher.

The LNA dissipates only 18mW from a single 3V supply, thanks to the re-use of the bias current. It occupies 2.88mm^2 in a 3-metal 0.6μm CMOS technology and almost 90% of the die area is used by the three transformers T1-T3. It is clear that not only the transformer quality is critical to the circuit performance, but also its size is a significant factor for the system cost.

The experimental results for the CMOS LNA in a 50Ω test environment are summarized in Table 6-3. The simulation results of the CMOS LNA using the new models of transformer T1-T3 computed using the measured metal resistivity (0.06Ω-μm) are also listed for comparison.

Table 6-3: Measured and Simulated LNA performance

	Measurements	Simulations
Supply voltage	3 V	3 V
Power dissipation	18 mW	18 mW
Frequency	900 MHz	900 MHz
Noise figure	4.1 dB	3.3 dB
S21	12.3 dB	12.8 dB
S12	-33.0 dB	-39.0 dB
S11	-6.0 dB	-8.3 dB
1-dB compression (input)	-16.0 dBm	-15.1 dBm
Technology	3-metal 0.6μm CMOS	
Die area	2.88 mm^2	

CHAPTER 7. CONCLUSIONS

Current implementations of wireless communication systems usually use a mix of technologies. GaAs devices are commonly used in the RF front-end, i.e., low noise amplifiers, mixers, and power amplifiers, because of their high intrinsic mobility and high f_T. Also the semi-insulating substrate of GaAs allows for integration of high quality passive components. Bipolar or BiCMOS is mainly used for IF applications and possibly for applications up to RF with modern processes having f_T up to around 30GHz. CMOS is a natural choice for mixed-signal applications and has been mainly used for the baseband signal processing.

Implementations of future generation wireless transceivers will likely be highly integrated for low cost, low power, and small size while still meeting increased performance demands. The CMOS solution is attractive because of its potentially lower cost, driven by the digital VLSI industry. However, innovative circuit techniques are required for high RF performance because CMOS devices, though achieving greater and greater f_T's in recent years, are still inferior to the Bipolar or GaAs counterparts.

In this thesis, we have explored the possibility of CMOS implementations of a critical RF front-end circuit, the low noise amplifier (LNA). A critical problem faced in the design of RF CMOS LNA's is the inaccurate high-frequency noise model of MOSFETs implemented in circuit simulators such as SPICE. To address this problem, we have investigated various noise sources associated with a MOSFET. It is found that two noise sources, i.e., the channel thermal noise and the induced gate current noise, are of significant importance at RF frequencies. A small-signal noise circuit model that includes both of these two noise sources was then developed for circuit analysis and simulations. Having been in such a good position, we then turn to the study of the basic principles of CMOS

LNA design. Theoretical analysis of various amplifier architectures using the improved noise circuit model has demonstrated that a CMOS LNA utilizing a series LC resonant network at its inputs has the best possible noise performance. We have also described optimization techniques and summarized design guidelines and procedures for the LC tuned CMOS LNA.

However, full integration of CMOS LNAs still presents a challenge. As the study led to the conclusion that inductors are critically important in achieving low noise performance, they turn out to be one bottleneck in fully-integrated RF CMOS designs due to the poor quality factor of silicon-based monolithic inductors. Beginning with a review of various implementations of monolithic inductors, we described the detailed analysis and modeling of square spiral inductors. The inductor performance was also analyzed and formulated to facilitate hand analysis. It was concluded that the quality factor and self-resonant frequency of a monolithic spiral inductor are mainly limited by the parasitic series resistance and shunt capacitance as constrained by the standard digital CMOS process. Further study has demonstrated some advantages provided by monolithic transformers consisting of two identical spiral inductors. Analysis and modeling of silicon-based monolithic transformers were presented and it was shown that in fully-differential applications, a monolithic transformer occupies less die area and achieves higher quality factor compared to two independent inductors with the same total effective inductance. It was also shown that monolithic transformers improve the common-mode rejection of the differential circuits. Design guidelines for transformer optimization were also described based on extensive simulation results.

Taking advantage of the transformer, a fully-integrated 900MHz LNA in 0.6μm CMOS, utilizing three monolithic transformers for input and output tuning, has been demonstrated. The complete circuit and layout description were presented. A bias current

re-use technique was used to reduce the power dissipation, and process-, voltage-, and temperature-tracking biasing techniques were discussed. Experimental results show that at 900MHz, the LNA dissipates 18mW from a single 3V power supply and provides a 4.1dB noise figure, 12.3dB power gain, -33.0dB reverse isolation with a 1-dB compression point at -16dBm, while requiring no off-chip components, other than baluns for testing.

As this work has demonstrated, CMOS technology is a promising and feasible solution to full integration of RF front-end circuits, which will eventually lead to a fully-integrated wireless communication system in future. As mainstream digital CMOS technology continues scaling down (0.1 micron CMOS devices with f_T's of around 100GHz have been reported [112]), the performance of RF CMOS circuits such as the LNA will continue to improve. It is also expected that monolithic transformers can be employed in other RF differential circuits such as bandpass filters and oscillators, to achieve higher quality factor and thus better circuit performance.

Nevertheless, several critical issues remain to be addressed. First, much more work needs to be carried out on the short-channel effects on MOSFET performance; particularly, the noise characteristics. Experiments have shown that the noise performance of a MOSFET is significantly influenced by hot carrier effects but the mechanism is not yet completely understood.

Second, circuit simulators should be developed to accurately predict the MOSFET noise performance at radio frequencies so that RF CMOS circuit designs can be optimized with maximal accuracy and minimal endeavor.

Third, limitations on the RF performance imposed by the CMOS processes should be explored in order to make improvements by adapting new architectures or modifying the

process itself. Specially, techniques should be investigated for realizing high quality monolithic inductors and transformers which are mainly limited by current digital CMOS processes and only achieve quality factors of about 3 and 5 respectively, as been demonstrated. Though many novel structures and processes have been proposed, thicker oxide and higher-conductivity metallization (gold or copper) seem to be more promising and practical in future CMOS technologies.

Finally, the effects of finite conductivity of the silicon substrate on the RF performance, of both the active and passive devices, should be investigated in more depth. Accurate substrate and package modeling and simulation methodology are needed to predict their effects upon circuit and system performance.

Given the strong motivation and intense interest in CMOS RF IC's, we believe that continued progress and improvements will be made. Prospects for a single-chip transceiver in a low-cost, low-power, and high-integration CMOS technology are excellent.

BIBLIOGRAPHY

[1] U. L. Rhode and T. T. N. Bucher, *Communications Receivers: Principles and Design*, McGraw-Hill Book Company, 1988, ch1.

[2] B. P. Lathi, *Modern Digital and Analog Communication Systems*, 2nd ed., Holt, Rinehart, and Winston, Philadelphia, PA, 1989, ch4.

[3] A. A. Abidi, "Low-power radio-frequency IC's for portable communications," *Proc. IEEE*, vol. 83, no. 4, pp. 544-569, April 1995.

[4] P. R. Gray and R. G. Meyer, "Future directions in silicon ICs for RF personal communications," in *Proc. of the IEEE 1995 Custom Integrated Circuits Conference*, May 1995, pp. 83-90.

[5] L. E. Larson, editor, *RF and Microwave Circuit Design for Wireless Communications*, Artech House Publishers, Massachusetts, 1996.

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

[7] A.N. Karanicolas, "A 2.7V 900MHz CMOS LNA and mixer," in *ISSCC Digest of Technical Papers*, Feb. 1996, pp. 50-51, 416.

[8] A. Rofougaram, J. Y. Chang, M. Rofougaram, and A. A. Abidi, "A 1 GHz CMOS RF front-end IC for a direct-conversion wireless receiver," *IEEE J. Solid-State Circuits*, vol. 31, pp. 880-889, July 1996.

[9] C. Y. Wu and S. Y. Hsiao, "The design of a 3-V 900-MHz CMOS bandpass amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 159-168, Feb. 1997.

[10] A. R. Shahani, D. K. Shaeffer, and T. H. Lee, "A 12mW wide dynamic range CMOS front-end for a portable GPS receiver," in *ISSCC Digest of Technical Papers*, Feb. 1997, pp. 368-369, 487.

[11] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May 1997.

[12] J. Janssens, M. Steyaert, and H. Miyakawa, "A 2.7 volt CMOS broadband low noise amplifier," in *Symposium on VLSI Circuits Digest of Technical Papers*, June 1997, pp. 87-88.

[13] J. J. Zhou and D. J. Allstot, "A fully-integrated CMOS 900MHz LNA utilizing monolithic transformers," in *ISSCC Digest of Technical Papers*, Feb. 1998, pp. 132-133.

[14] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., John Wiley & Sons, 1993, ch11.

[15] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, 1987, ch8.

[16] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physics Review*, 32, 1928, pp. 110-113.

[17] R. Pettai, *Noise in Receiving Systems*, John Wiley & Sons, 1984, ch3.

[18] C. T. Sah, S. Y. Wu, and F. H. Hielsher, "The effects of fixed bulk charge on the thermal noise in metal-oxide-semiconductor transistor," *IEEE Trans. on Electron Devices*, vol. 13, no. 4, pp. 410-414, April 1966.

[19] S. Tedja, J. van der Spiegel, and H. H. Williams, "Analytical and experimental studies of thermal noise in MOSFET's," *IEEE Trans. on Electron Devices*, vol. 41, no. 11, pp. 2069-2074, Nov. 1994.

[20] R. M. Fox, "Comments on circuit models for MOSFET thermal noise," *IEEE J. Solid-State Circuits*, vol. 28, no. 2, pp. 184-185, Feb. 1993.

[21] Z. Y. Chang and W. M. C. Sansen, *Low-Noise Wide-Band Amplifiers in Bipolar and CMOS Technologies*, Kluwer Academic Publishers, 1991, ch2.

[22] B. Wang, J. R. Hellums, and C. G. Sodini, "MOSFET thermal noise modeling for analog integrated circuits," *IEEE J. Solid-State Circuits*, vol. 29, no. 7, pp. 833-835, July 1994.

[23] A. A. Abidi, "High-frequency noise measurements on FET's with small dimensions," *IEEE Trans. on Electron Devices*, vol. ED-33, no. 11, pp. 1801-1805, Nov. 1986.

[24] R. P. Jindal, "Hot electron effects on channel thermal noise in fine-line NMOS field effect transistors," *IEEE Trans. on Electron Devices*, vol. ED-33, no. 9, pp. 1395-1397, Sept. 1986.

[25] A. van der Ziel, "Gate noise in field effect transistors at moderately high frequencies," *Proc. IEEE*, vol. 51, pp. 461-467, March 1963.

[26] H. E. Halladay and A. van der Ziel, "On the high frequency excess noise and equivalent circuit representation of the MOS-FET with n-type channel," *Solid-State Electronics*, vol. 12, pp. 161-176, 1969.

[27] A. van der Ziel, *Noise in Solid State Devices and Circuits*, John Wiley & Sons, 1986, ch5.

[28] F. M. Klaassen, "High frequency noise of the junction field-effect transistor," *IEEE Trans. on Electron Devices*, vol. ED-14, no. 7, pp. 368-373, July 1967.

[29] R. P. Jindal, "Noise associated with distributed resistance of MOSFET gate structures in integrated circuits," *IEEE Trans. on Electron Devices*, vol. ED-31, no. 10, pp. 1505-1509, Oct. 1984.

[30] K. K. Thornber, "Resistive-gate-induced thermal noise in IGFETs," *IEEE J. Solid-State Circuits*, vol. SC-16, no. 4, pp. 414-415, Aug. 1981.

[31] B. Razavi, R. H. Yan, and K. F. Lee, "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Trans. on Circuits and Systems I*, vol. 41, pp. 750-754, Nov. 1994.

[32] J. B. Johson, "The schottky effect in low frequency circuits," *Physics Review*, 26, 1925, pp. 71-85.

[33] A. van der Ziel, "Unified presentation of 1/f noise in electronic devices: fundamental 1/f noise sources," *Proc. IEEE*, vol. 76, pp. 233-258, March 1988.

[34] F. M. Klaassen, "Characterization of low 1/f noise in MOS transistors," *IEEE Trans. on Electron Devices*, vol. ED-18, no. 10, pp. 887-891, Oct. 1971.

[35] G. Reimbold, "Modified 1/f trapping noise theory and experiments in MOS transistors biased from weak to strong inversion-influence of interface states," *IEEE Trans. on Electron Devices*, vol. ED-31, no. 9, pp. 1190-1198, Sept. 1984.

[36] A. van der Ziel, "Theory of shot noise in junction diodes and junction transistors," *Proc. IRE*, pp. 1639-1646, Nov. 1955.

[37] R. P. Jindal, "Distributed substrate resistance noise in fine-line NMOS field-effect transistors," *IEEE Trans. on Electron Devices*, vol. ED-32, no. 11, pp. 2450-2453, Nov. 1985.

[38] L. M. Rucker and A. van der Ziel, "Noise associated with JFET gate current resulting from avalanching in the channel," *Solid-State Electronics*, vol. 21, pp. 798-799, 1978.

[39] R. P. Jindal, "Noise associated with substrate current in fine-line NMOS field-effect transistors," *IEEE Trans. on Electron Devices*, vol. ED-32, no. 6, pp. 1047-1052, June 1985.

[40] N. K. Verghese and D. J. Allstot, "Computer-aided design considerations for mixed-signal coupling in RF integrated circuits," *IEEE J. Solid-State Circuits*, vol. 33, pp. 314-323, March 1998.

[41] H. T. Friis, "Noise figures of radio receivers," *Proc. IRE*, pp. 419-422, July 1944.

[42] "Description of the noise performance of amplifiers and receiving systems," sponsored by IRE subcommittee 7.9 on Noise, *Proc. IEEE*, pp. 436-442, March 1963.

[43] K. A. Simons, "The decibel relationships between amplifier distortion products," *Proc. IEEE*, pp. 1071-1086, July 1970.

[44] B. Razavi, *RF Microelectronics*, Prentice Hall, New Jersey, 1998, ch2.

[45] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998, ch11.

[46] Y. Netzer, "The design of low-noise amplifier," *Proc. IEEE*, vol. 69, pp. 728-741, June 1981.

[47] B. K. Ko and K. Lee, "A comparative study on the various monolithic low noise amplifier circuit topologies for RF and microwave applications," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1220-1225, August 1996.

[48] Y. Netzer, "A new interpretation of noise reduction by matching," *Proc. IEEE*, vol. 62, pp. 404-406, Mar. 1974.

[49] B. Razavi, *RF Microelectronics*, Prentice Hall, New Jersey, 1998, ch6.

[50] R. E. Lehmann and D. D. Heston, "X-band monolithic series feedback LNA," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-33, no. 12, pp. 1560-1566, Dec. 1985.

[51] R. G. Meyer and W. D. Mack, "A 1-GHz BiCMOS RF front-end integrated circuit," *IEEE J. Solid-State Circuits*, vol. 29, pp. 350-355, Mar. 1994.

[52] N. M. Nguyen and R. G. Meyer, "Si IC-compatible inductors and LC passive filters," *IEEE J. Solid-State Circuits*, vol. 25, pp. 1028-1031, Aug. 1990.

[53] N. M. Nguyen and R. G. Meyer, "A Si bipolar monolithic RF bandpass amplifier," *IEEE J. Solid-State Circuits*, vol. 27, pp. 123-127, Jan. 1992.

[54] N. M. Nguyen and R. G. Meyer, "A 1.8-GHz monolithic LC voltage-controlled oscillator," *IEEE J. Solid-State Circuits*, vol. 27, pp. 444-450, Mar. 1990.

[55] J. Craninckx and M. Steyaert, "A 1.8-GHz low-phase-noise CMOS VCO using optimized hollow spiral inductors," *IEEE J. Solid-State Circuits*, vol. 32, pp. 736-744, May 1997.

[56] J. Crols, P. Kinget, J. Craninckx, and M. Steyaert, "An analytical model of planar inductors on lowly doped silicon substrates for high frequency analog design up to 3GHz," in *Symposium on VLSI Circuits Digest of Technical Papers*, June 1996, pp. 28-29.

[57] J. R. Long and M. A. Copeland, "The modeling, characterization, and design of monolithic inductors for silicon RF IC's," *IEEE J. Solid-State Circuits*, vol. 32, pp. 357-369, March 1997.

[58] D. Lovelace, N. Camilleri, and G. Kannell, "Silicon MMIC inductor modeling for high volume, low cost applications," *Microwave Journal*, pp. 60-71, August 1994.

[59] J. Y. Chang, A. A. Abidi, and M. Gaitan, "Large suspended inductors on silicon and their use in a 2-μm CMOS RF amplifier," *IEEE Electron Device Lett.*, vol. 14, no. 5, pp. 246-248, May 1993.

[60] A. Rofougaran, J. Rael, M. Rofougaran, and A. Abidi, "A 900MHz CMOS LC-oscillator with quadrature outputs," in *ISSCC Digest of Technical Papers*, Feb. 1996, pp. 392-393.

[61] M. Soyuer, J. N. Burghartz, K. A. Jenkins, S. Ponnapalli, J. F. Ewen, and W. E. Pence, "Multi-level monolithic inductors in silicon technology," *Electron. Let.*, vol. 31, no. 5, pp. 359-360, Mar. 2, 1995.

[62] J. N. Burghartz, M. Soyuer, and K. A. Jenkins, "Microwave inductors and capacitors in standard multilevel interconnect silicon technology," *IEEE Trans. on Microwave Theory and Techniques*, vol. 44, pp. 100-104, Jan. 1996.

[63] M. Soyuer, K. A. Jenkins, J. N. Burghartz, H. A. Ainspan, F. J. Canora, S. Ponnapalli, J. F. Ewen, and W. E. Pence, "A 2.4-GHz silicon bipolar oscillator with integrated resonator," *IEEE J. Solid-State Circuits,* vol. 31, pp. 268-270, Feb. 1996.

[64] S. Hara, T. Tokumitsu, and M. Aikawa, "Lossless broad-band monolithic microwave active inductor," *IEEE Trans. on Microwave Theory and Techniques*, vol. 36, pp. 1920-1924, Dec. 1988.

[65] G. F. Zhang and J. L. Gautier, "Broad-band, lossless monolithic microwave active floating inductor," *IEEE Microwave and Guided Wave Lett.*, vol. 3, pp. 98-100, April 1993.

[66] J. Craninckx and M. Steyaert, "Low-noise voltage controlled oscillators using enhanced LC-tanks," *IEEE Trans. on Circuits and Systems II*, vol. 42, pp. 794-804, Dec. 1995.

[67] H. M. Greenhouse, "Design of planar rectangular microelectronic inductors," *IEEE Trans. on Parts, Hybrids and Packaging*, vol. PHP-10, pp. 101-109, June 1974.

[68] J. Craninckx and M. Steyaert, "A 1.8-GHz CMOS low-phase-noise voltage-controlled oscillator with prescaler," *IEEE J. Solid-State Circuits*, vol. 30, pp. 1474-1482, Dec. 1995.

[69] F. W. Grover, *Inductance Calculations*, Van Nostrand, Princeton, NJ, 1946; reprinted by Dover Publications, New York, NY, 1962.

[70] K. B. Ashby, I. A. Koullias, W. C. Finley, J. J. Bastek, and S. Moinian, "High Q inductors for wireless applications in a complementary silicon bipolar process," *IEEE J. Solid-State Circuits*, vol. 31, pp. 4-9, Jan. 1996.

[71] E. Pettenpaul, H. Dapusta, A. Weisgerber, H. Mampe, J. Luginsland, and I. Wolff, "CAD models of lumped elements on GaAs up to 18 GHz," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-36, no. 2, pp. 294-304, Feb. 1988.

[72] D. M. Krafesik and D. E. Dawson, "A closed-form expression for representing the distributed nature of the spiral inductor", in *IEEE 1986 Microwave and Millimeter-Wave Monolithic Circuits Symp. Digest of Papers*, 1986, pp. 87-92.

[73] *FastHenry USER'S GUIDE (version 3.0)*, Research Laboratory of Electronics, MIT, Cambridge, MA, Nov. 1996.

[74] S. J. Haefner, "Alternating current of rectangular conductors," *Proc. IRE*, vol. 25, pp. 434-447, 1937.

[75] H. Hasegawa, M. Furukawa, and H. Tanai, "Properties of microstrip line on Si-SiO_2 system," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-19, no. 11, pp. 869-881, November 1971.

[76] T. Sakurai and K. Tamaru, "Simple formulas for two- and three-dimensional capacitances," *IEEE Trans. on Electron Devices*, pp. 183-185, Feb. 1983.

[77] R. L. M. Dang and N. Shigyo, "Coupling capacitances for two-dimensional wires," *IEEE Electron Device Lett.*, vol. EDL-2, no. 8, pp. 196-197, August 1981.

[78] D. W. Kammler, "Calculation of characteristic admittances and coupling coefficients for strip transmission lines," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-16, no. 11, pp. 925-937, November 1968.

[79] J. I. Smith, "The even- and odd-mode capacitance parameters for coupled lines in suspended substrate," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-19, no. 5, pp. 424-431, May 1971.

[80] R. Garg and I. J. Bahl, "Characteristics of coupled microstriplines," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-27, no. 7, pp. 700-705, July 1979.

[81] D. M. Pozar, *Microwave Engineering*, Addison-Wesley, 1990, reprinted in 1993, pp. 183-186.

[82] H. A. Wheeler, "Transmission-line properties of a strip on a dielectric sheet on a plane," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-25, no. 8, pp. 631-647, Aug. 1977.

[83] R. H. Jansen, "High-speed computation of single and coupled microstrip parameters including dispersion, high-order modes, loss and finite strip thickness," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-26, no. 2, pp. 75-82, Feb. 1978.

[84] N. K. Verghese, T. J. Schmerbeck, and D. J. Allstot, *Simulation Techniques and Solutions for Mixed-Signal Coupling in Integrated Circuits*, Kluwer Academic Publishers, 1995, ch6.

[85] D. K. Su, M. J. Loinaz, S. Masui, and B. A. Wooley, "Modeling techniques and experimental results for substrate noise in mixed-signal integrated circuits," *IEEE J. Solid-State Circuits*, vol. 28, pp. 420-430, April 1993.

[86] M. Pfost, H. M. Rein, and T. Holzwarth, "Modeling substrate effects in the design of high-speed Si-bipolar IC's," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1493-1501, Oct. 1996.

[87] E. Tuncer and D. P. Neikirk, "Highly accurate quasi-static modeling of microstrip lines over lossy substrate," *IEEE Microwave and Guided Wave Lett.*, vol. 2, pp. 409-411, Oct. 1992.

[88] R. J. P. Douville and D. S. James, "Experimental study of symmetric microstrip bends and their compensation," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-26, no. 3, pp. 175-181, March 1978.

[89] M. Parisot, Y. Archambault, D. Pavlidis, and J. Magarshack, "Highly accurate design of spiral inductors for MMIC's with small size and high cut-off frequency characteristics," in *1984 IEEE MTT-S Digest*, 1984, pp. 106-110.

[90] H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, John Wiley & Sons, 1980, ch3.

[91] S. Chaki, S. Aono, N. Andoh, Y. Sasaki, N. Tanino, and O. Ishihara, "Experimental study on spiral inductors," in *1995 IEEE MTT-S Digest*, 1995, pp. 753-756.

[92] B. Razavi, "A 1.8GHz CMOS voltage-controlled oscillator," in *ISSCC Digest of Technical Papers*, Feb. 1997, pp. 388-389.

[93] Y. P. Tsividis, "Integrated continuous-time filter design," in *Proc. Custom Int. Circuits Conference,* May 1993, pp. 6.4.1-6.4.7.

[94] R. A. Duncan, K. W. Martin, and A. S. Sedra, "A Q-enhanced active-RLC bandpass filter," in *IEEE Int. Symp. on Circuits and Systems*, 1993, pp. 1416-1419.

[95] S. Pipilos and Y. Tsividis, "Design of active RLC integrated filters with application in the GHz range," in *IEEE Int. Symp. on Circuits and Systems*, 1994, pp. 645-648.

[96] W. B. Kuhn, F. W. Stephenson, and A. Elshabini-Riad, "A 200 MHz CMOS Q-enhanced LC bandpass filter," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1112-1122, Aug. 1996.

[97] C. Yue and S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF IC's," in *Symposium on VLSI Circuits Digest of Technical Papers*, June 1997, pp. 85-86.

[98] D. Shaeffer, A. Shahani, S. Mohan, H. Samavati, H. Rategh, M. Hershenson, M. Xu, C. Yue, D. Eddleman, and T. Lee, "A 115mW CMOS GPS receiver," in *ISSCC Digest of Technical Papers*, Feb. 1998, pp. 122-123.

[99] S. R. Taub and S. A. Alterovitz, "Silicon technologies adjust to RF applications," *Microwave & RF*, pp. 60-74, Oct. 1994.

[100] J. R. Long and M. A. Copeland, "A 1.9GHz low-voltage silicon bipolar receiver front-end for wireless personal communications systems," *IEEE J. Solid-State Circuits*, Vol. 30, pp. 1438-1448, December 1995.

[101] D. Ferguson, P. Bauhahn, J. Keuper, R. Lokken, J. Culp, C. Chao, and A. Podell, "Transformer coupled high-density circuit technique for MMIC", in *1984 IEEE MTT-S Digest*, 1984, pp. 34-36.

[102] S. A. Jamison, A. Podell, M. Helix, P. Ng, and C. Chao, "Inductively coupled push-pull amplifiers for low cost monolithic microwave ICs," in *1982 IEEE GaAs IC Symposium*, 1982, pp. 91-93.

[103] L. Selmi and B. Ricco, "Design of an X-band transformer-coupled amplifier with improved stability and layout," *IEEE J. Solid-State Circuits,* vol. 28, pp. 701-703, June 1993.

[104] A. Boulouard and M. L. Rouzic, "Analysis of rectangular spiral transformers for MMIC applications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 37, no. 8, pp. 1257-1260, August 1989.

[105] E. Frlan, S. Meszaros, M. Cuhaci, and J. S. Wight, "Computer aided design of square spiral transformers and inductors", in *1989 IEEE MTT-S Digest*, 1989, pp. 661-664.

[106] J. R. Long and M. A. Copeland, "Modeling of monolithic inductors and transformers for silicon RF IC design," in *Proc. IEEE MTT-S Int. Symp. Tech. Wireless Appl.,* Vancouver, Canada, Feb. 1995, pp. 129-134.

[107] J. Hogerheiden, M. Ciminera, and G. Jue, "Improved planar spiral transformer theory applied to a miniature lumped element quadrature hybrid," *IEEE Trans. on Microwave Theory and Techniques*, vol. 45, no. 4, pp. 543-545, April 1997.

[108] W. B. Kuhn, "Design of integrated, low power, radio receivers in BiCMOS technologies," *Ph.D. Dissertation*, Virginia Polytechnic Institute and State University, 1995.

[109] W. B. Kuhn, A. Elshabini-Rian, and F. W. Stephenson, "Centre-tapped spiral inductors for monolithic bandpass filters," *Electron. Lett.*, vol. 31, no.8, pp. 625-626, Apr. 13, 1995.

[110] L. M. Devlin, B. J. Buck, J. C. Clifton, A. W. Dearn, and A. P. Long, "A 2.4 GHz single chip transceiver," in *IEEE 1993 Microwave and Millimeter-Wave Monolithic Circuits Symp.*, June 1993, pp. 23-26.

[111] N. Camilleri, J. Kirschgessner, J. Costa, D. Ngo, and D. Lovelace, "Bonding pad models for silicon VLSI technologies and their effects on the noise figure of RF NPN's," in *IEEE 1994 Microwave and Millimeter-Wave Monolithic Circuits Symp.*, 1994, pp. 225-228.

[112] R. H. Yan, K. F. Lee, D. Y. Jeon, Y. O. Kim, B. G. Park, M. R. Pinto, C. S. Rafferty, D. M. Tennant, E. H. Westerwick, G. M. Chin, M. D. Morris, K. Early, P. Mulgrew, W. M. Mansfield, R. K. Watts, A. M. Voshchenkov, J. Bokor, R. G. Swartz, and A. Ourmazd, "High performance 0.1-micron room temperature Si MOSFETs," in *1992 Symposium on VLSI Technology Digest of Technical Papers*, June 1992, pp. 86-87.

APPENDICES

Appendix A Matlab Programs for Inductor and Transformer Simulations

The use of computer analysis and optimization programs is of great importance in the design of monolithic inductors and transformers. In this appendix, the Matlab programs which include closed-form expressions for the computation of shunt parasitics C_{ox}, C_c, R_{si}, and C_{si}, derived in Chapter 4, to conduct the simulation and optimization for monolithic spiral inductors and transformers are illustrated. The programs also generate the input files to *FastHenry* which is used for the computation of the self-inductance, mutual coupling coefficient, and metal resistance including the skin effect.

A.1 Programs for Inductor Simulations

The main program is `ind_simulation.m` which takes the number of turns, the center hole spacing, the trace width, the trace spacing, and a file name as the input parameters. Its output is a graphical presentation of the inductor's quality factor, self-resonant frequency, and one-port impedance, as illustrated in Chapter 5. Its subroutine, `z0input.m` generates an input file to *FastHenry* for the computation of the inductance and metal resistance including the skin effect of the inductor. It also invokes `parasitics.m` to calculate the inductor parasitics C_{ox}, C_c, R_{si}, and C_{si}. and draws a 3-D plot of the inductor. All programs are listed below.

`ind_simulation.m`

```
function ind_simulation(n,center,w,d,fname);

%% function ind_simulation(n,center,w,d,fname);
%% w=the trace width; n=the number of turns (any positive real number);
%% d=the spacing between adjacent traces;
%% center=the center hole spacing (between center lines of traces);
%% fname=output filename

%% simulations for spiral inductors;
```

```
henry=[fname, '.henry'];
par=[fname, '.par'];

%% generating input file to FastHenry; dumping the output to fname.henry;
%% computing the parasitics; dumping the outputs to fname.par
z0input(n,center,w,d,fname);

%% read in data from output files: fname.henry and fname.par
fid=fopen(henry,'r');dat1=fscanf(fid,'%f');
fid=fopen(par,'r');dat2=fscanf(fid,'%f');
n1=size(dat1,1);n2=size(dat2,1);

Cox=dat2(1:5:n2); %% in pF
Cc=dat2(2:5:n2);%% in pF
Csi=dat2(3:5:n2);%% in pF
Rsi=dat2(4:5:n2);%% in ohm
length=dat2(5:5:n2);%% in um

freq=dat1(1:3:n1)*1e-9;%% in GHz
resistance=dat1(2:3:n1);%% in ohm
inductance=dat1(3:3:n1)./freq/2/pi;%% in nH

%eddy current loss; changed with different technologies
Rloss=freq.^2*2.8e-4;%% ohm/um; Ref: Hasegawa et al

%% total series resistance
Rt=resistance+Rloss*length;%% in ohm;

%% computing the quality factor, the self-resonant frequency,
%% and the one-port impedance
m=4; %% determine m by gradual refinement, fitting to scalable models
for i=1:n1/3,
  [Q(i),Fr(i),imag_imp(i),real_imp(i)]=q(freq(i),inductance(i),Rt(i),Cox,Cc/m,Rsi);
end

%% plotting the quality factor, the self-resonant frequency,
%% and the one-port impedance
figure;
subplot(311);
semilogx(freq*1e9,real_imp,'-',freq*1e9,Rt,'--');
grid on;
ylabel('Real Impedance Ohms');
axis([1e8 1e10 0 265]);

subplot(312);
semilogx(freq*1e9,imag_imp, '-',freq*1e9,inductance,'--');
grid on;
ylabel('Imaginary Impedance nH');
axis([1e8 1e10 -7 9]);

subplot(313);
semilogx(freq*1e9,Q);
grid on;
ylabel('Quality Factor');
axis([1e8 1e10 -1 6]);
xlabel('Frequency Hz');

fclose('all');
```

z0input.m

```
function z0input(n,center,w,d,fname);

%% function z0input(n,center,w,d,fname);
%% FastHenry input file generator;
%% inductor computation using fasthenry and parasitics.m
%% inductors structure--2D
%% 03/28/97
%% w=the trace width; n= the number of turns(any positive real number);
%% d=the spacing between adjacent traces;
%% center=the center hole spacing (between center lines of traces);
%% fname=output filename

%%--------------------------------------
%% process information

rho=0.042;          %% metal resistivity; ohm-um
h=1.2;              %% trace height; um

%%--------------------------------------

node=ceil(4*n+1);  %% number of nodes
spacing=w+d;       %% spacing between two center lines of metal trace
radius=(ceil(n)-1)*spacing+center/2;

length1=0; %% total length of in-between microstrip lines, um
length2=0; %% total length of outer-most microstrip lines, um
length3=0; %% total length of inner-most microstrip lines, um


%% generate (x,y)

m=n-floor(n);
x(node)=0;
y(1)=-radius;%% if m==0, this defines y(1);

if (0<m & m<=0.25)
  y(1)=-radius;
  x(1)=(2*radius-spacing)*m*4+spacing-radius;
  y(2)=y(1);
elseif (0.25<m & m<=0.5)
  y(1)=2*radius*4*(m-0.25)-radius;
  x(1)=radius;
  y(2)=-radius;
  x(2)=radius;
  y(3)=y(2);
elseif (0.5<m & m<=0.75)
  y(1)=radius;
  x(1)=(-2)*radius*4*(m-0.5)+radius;
  y(2)=radius;
  x(2)=radius;
  y(3)=-radius;
  x(3)=radius;
  y(4)=y(3);
elseif m~=0
  x(1)=-radius;
  y(1)=(-2)*radius*4*(m-0.75)+radius;
  y(2)=radius;
  x(2)=-radius;
  y(3)=radius;
  x(3)=radius;
```

```
  y(4)=-radius;
  x(4)=radius;
  y(5)=y(4);
end

j=ceil(4*m)+1;
for i=j:4:node-1,
  x(i)=ceil((i-1)/4)*spacing-radius;
  x(i+1)=x(i);
  x(i+2)=-x(i);
  x(i+3)=x(i+2);
  y(i+1)=x(i+2);
  y(i+2)=x(i+2);
  y(i+3)=-x(i+3);
  y(i+4)=y(i+3);
end

%%-------------------------------------
%% generate fasthenry input file: fname.inp

filename=[fname '.inp'];
fid = fopen(filename,'w');
dd=date; tt=fix(clock);

fprintf(fid, '*** planar spiral inductors--2D ***\n\n');
fprintf(fid, '*** %d:%d:%d PST, %s ***\n\n', tt(4),tt(5),tt(6),dd);
fprintf(fid, '.units um\n');

if w>25
  fprintf(fid, '.default z=0 w=%d h=%2.4f rho=%2.4f nhinc=1 nwinc=7\n\n\',w,h,rho);
else
  fprintf(fid, '.default z=0 w=%d h=%2.4f rho=%2.4f nhinc=1 nwinc=5\n\n\',w,h,rho);
end

fprintf(fid, '*** z0input(n=%d, center=%d, w=%d, d=%d, %s)\n\n', n,center,w,d,fname);
fprintf(fid, '*** spacing between two lines: d=%dum\n\n\n',d);

substrate(1.5*radius,x,y,0,0,w,filename);

if m==0
  j=5;
end

for i=1:node,
  fprintf(fid,'n%d  x=%4.2f  y=%4.2f\n',i,x(i),y(i));
  if (i<=5 &i>1)
    length2=length2+(abs(x(i)-x(i-1))+abs(y(i)-y(i-1)));
  elseif (i<=node & i>node-4)
    length3=length3+(abs(x(i)-x(i-1))+abs(y(i)-y(i-1)));
  elseif (i<=node-4 & i~=1)
    length1=length1+(abs(x(i)-x(i-1))+abs(y(i)-y(i-1)));
  end
end


fprintf(fid,'\n\n');

for i=1:node-1,
  fprintf(fid,'e%d  n%d  n%d\n',i,i,i+1);
end;

fprintf(fid,'\n\n');

fprintf(fid,'.external  n1  n%d\n', node);
```

```
fprintf(fid,'\n\n.freq fmin=1e+08 fmax=1e+10 ndec=9\n\n.end\n');

%%-------------------------------------
%% fasthenry computation:output fname.henry

eval(['!fasthenry -S _' fname ' ' filename]);
output1=[fname '.henry'];

eval(['!awk -f ind_henry.awk Zc_' fname '.mat >>' output1]);


%%-------------------------------------
%% parasitics computation:output fname.par

c1=(length1+length2+length3)*w*0.0092;
length=(length1+length3)+(w+d)*ceil(4*(n-1));
if length<0
  length=0;
end

[Cox, Cc, Csi, Rsi]=parasitic(w,d,length1,length2,length3,length);
total_length=length1+length2+length3;

output2=[fname '.dat'];
fid = fopen(output2,'a');

fprintf(fid, '========================== planar spiral inductors--2D \n');
fprintf(fid, '*** %d:%d:%d PST, %s ***\n\n', tt(4),tt(5),tt(6),dd);

fprintf(fid, '*** z0input(n=%d, center=%d, w=%d, d=%d, %s)\n\n', n,center,w,d,fname);
fprintf(fid, '*** estimated capacitance: c1=%2.4ffF\n\n\n',c1);
fprintf(fid, 'length1=%2.2fum length2=%2.2fum length3=%2.2fum length=%2.2fum\n\n', length1,
length2, length3, length);
fprintf(fid, 'total_length= %2.2f um\n\n', total_length);
fprintf(fid, 'Cox= %2.4f pF Cc= %2.4f pF Csi= %2.4f pF Rsi= %2.4f ohm\n\n\n',Cox, Cc, Csi, Rsi);

output3=[fname '.par'];
fid = fopen(output3,'a');
fprintf(fid, '%2.4f %2.4f %2.4f %2.4f %2.2f\n',Cox, Cc, Csi,Rsi,total_length);


fclose('all');


%%-------------------------------------
%% 3D plot;

xmax=max([max(x),max(y)]);
xmin=min([min(x),min(y)]);

figure;
polyfill(0,node,h,w,x,y,'y');

axis([xmin,xmax,xmin,xmax,xmin/4,xmax/4]);
axis('equal');
axis off;
hold off;
%%-------------------------------------
```

q.m

```
function [Q,fr,imag_imp,real_imp]=q(f,L,Rt,Coxx,Cc,Rsi);

%% function [Q,fr,imag_imp,real_imp]=q(f,L,Rt,Coxx,Cc,Rsi); 06/22/98
%% f: frequency, in GHz
%% L: inductance, in nH
%% Rt:resistive losses including skin effect and eddy current loss;
%% Coxx: oxide cap to substrate; pF
%% Cc: side coupling capacitance; pF
%% Rsi: substrate parasitic shunt resistance

%% quality factor and one-port impedance computation for
%% transformers and inductors;
%% Q: quality factor; fr: self-resonance, in Hz;
%% imag_imp,real_imp: imaginary and real part of impedance, in nH and ohm;


w=2*pi*f*1e9;
Cox=Coxx+Cc;

wr=1/sqrt(L*1e-9*Cox*1e-12)*sqrt((1-Rt^2*Cox/L*1e-3)/(1-Rsi^2*Cox/L*1e-3));
%% self-resonance
fr=wr/2/pi;

x1=w*L*1e-9*(1-Rt^2*Cox/L*1e-3);
x2=Rt+Rsi*w^4*L^2*Cox^2*1e-42+(w*Cox*1e-12)^2*(Rt*Rsi^2+Rt^2*Rsi);
Q=x1/x2*(1-(w/wr)^2);  %% quality factor

%% make use of complex computation
z1=j*w*L*1e-9+Rt;
y1=z1^-1;
z2=(j*w*Cox*1e-12)^-1+Rsi;
y2=z2^-1;
y=y1+y2;
z=y^-1;
imag_imp=imag(z)/2/pi/f;
real_imp=real(z);
```

substrate.m

```
function substrate(length,x1,y1,x2,y2,w,fname);

%% function substrate(length,height,x,y,w,fname);
%% Substrate specification for FastHenry input file generator
%% 03/30/97
%% length: outer length of inductor (square)
%% x1,y1,x2,y2: node's (x,y) of the bottom inductor's trace
%% w: the width of trace
%% fname=output filename

height=4;    %% the height of the top metal plane over substrate; um
thickness=250;  %% thickness of substrate; um
rho=0.14e4;      %% substrate resistivity, Ohm-um
z=-(height+thickness/2);
kk=size(x1,2);

%%-------------------------------------
%% output fname.inp
```

```
fid = fopen(fname,'a');

fprintf(fid,'\n\n*** Define substrate, resistivity=%4.2f Ohm-cm ***\n',rho/1e4);
fprintf(fid, 'g_substrate\n');
fprintf(fid, '+ x1=%4.2f y1=%4.2f z1=%4.2f \n', -length,-length,z);
fprintf(fid, '+ x2=%4.2f y2=%4.2f z2=%4.2f \n', length,-length,z);
fprintf(fid, '+ x3=%4.2f y3=%4.2f z3=%4.2f \n', length,length,z);
fprintf(fid, '+ thick=%4.2f rho=%4.2f file=NONE \n', thickness,rho);
fprintf(fid, '***** under the trace \n');
for i=1:kk-1,
 fprintf(fid,'+ contact trace (%4.2f,%4.2f,%4.2f,',x1(i),y1(i),z);
 fprintf(fid,'%4.2f,%4.2f,%4.2f,%4.2f,1)\n',x1(i+1),y1(i+1),z,w);
end

if size(x2,2)>1
 for i=1:kk-1,
  fprintf(fid,'+ contact trace (%4.2f,%4.2f,%4.2f,',x2(i),y2(i),z);
  fprintf(fid,'%4.2f,%4.2f,%4.2f,%4.2f,1)\n',x2(i+1),y2(i+1),z,w);
 end
end

fprintf(fid, '+ nhinc=3 rh=2 \n\n');
fprintf(fid,'***************\n\n');
```

parasitic.m

```
function [Cox,Cc,Csi,Rsi]=parasitic(w,s,length1,length2,length3,length);

%% function [Cox, Cc, Csi, Rsi]=parasitic(w,h,s);
%% Cox: oxide cap to substrate; Cc: sidewall cap;  %%pF
%% Csi, Rsi: substrate parasitics  %% pF
%% w=width of trace, um;
%% s=spacing between adjacent traces, um;
%% length1: in-between; length2: outtermost; length3: innermost; um
%% length: effective length for side wall, um

%% characteristics of coupled microstriplines
%% R. Garg and I. J. Bahl, IEEE MTT, July 1979


%%-------------------------constants
C0=2.998e2;    %% light speed, um/ps
E0=8.854e-6;    %% permitivity of free space, pF/um
Er=3.9;        %% relative dielectric constant of SiO2
Err=11.9;       %% relative dielectric constant of Si
t=1.2;           %% trace height, um
Tep=7;         %%thickness of epitaxial layer, um
Pep=1e5;        %% resistivity of epi layer, ohm-um
h=4;           %% oxide thickness, um;


%%%----------oxide capacitance and side coupling capacitance
Ere=(Er+1)/2+(Er-1)/2/sqrt(1+12*h/w);  %%equivalent dielectric constant
if (w/h<=1)
 Z0=60/sqrt(Ere)*log(8*h/w+w/4/h);
else
 Z0=120*pi/sqrt(Ere)/(w/h+1.393+0.667*log(w/h+1.444));
end;
```

```
Cp=E0*Er*w/h;
C1=0.5*(sqrt(Ere)/C0/Z0-E0*Er*w/h);
A=exp(-0.1*exp(2.33-2.53*w/h));
C2=C1/(1+A*h/s*tanh(8*s/h));
k=s/h/(s/h+2*w/h);
kp=sqrt(1-k^2);
if(0<=k^2 & k^2<=0.5)
  kk=1/pi*log(2*(1+sqrt(kp))/(1-sqrt(kp)));
else
  kk=pi/log(2*(1+sqrt(k))/(1-sqrt(k)));
end;
Cga=E0/2*kk;
Cgd=E0*Er/pi*log(coth(pi/4*s/h))+0.65*C1*(0.02/s*h*sqrt(Er)+1-Er^(-2));

Cox=(Cp/2+C2)*length1+(Cp+C1+C2)/2*(length2+length3);

Cc=(2*(Cga+Cgd)-C2+E0*t/s)*length/2;



%%%----------substrate parasitics
Peff=Pep/(0.5+0.5/sqrt(1+12*Tep/w));
if (w/Tep<=1)
  Teff=w/2/pi*log(8*Tep/w+w/4/Tep);
else
  Teff=w/(w/Tep+1.393+0.667*log(w/Tep+1.444));
end;
Rsi=2*Peff*Teff/w/(length1+length2+length3);
Csi=E0*Err*Pep/Rsi;
```

polyfill.m

```
function polyfill(zz,node,h,w,x,y,color);
%% 3D plot for inductors
%% 03/27/97
%% w=width of trace;
%% h=trace height
%% zz=z-axis position
%% node=number of nodes
%% x,y= position of each node


for i=1:node-1,
 j=(i-1)*6;
 zq(1,j+1)=zz-h/2;
 zq(2,j+1)=zz-h/2;
 zq(3,j+1)=zz+h/2;
 zq(4,j+1)=zz+h/2;
 for k=1:4,
   zq(k,j+5)=zz+h/2;
   zq(k,j+6)=zz-h/2;
   zq(k,j+2)=zq(k,j+1);
   zq(k,j+3)=zq(k,j+1);
   zq(k,j+4)=zq(k,j+3);
 end
 xx1=min(x(i),x(i+1));
 xx2=max(x(i),x(i+1));
 yy1=min(y(i),y(i+1));
 yy2=max(y(i),y(i+1));
```

```
  if xx1==xx2
    xq(1,j+1)=xx1+w/2;
    xq(2,j+1)=xx1-w/2;
    xq(3,j+1)=xq(2,j+1);
    xq(4,j+1)=xq(1,j+1);

    yq(1,j+3)=yy2-w/2;
    yq(2,j+3)=yy1+w/2;
    yq(3,j+3)=yy1+w/2;
    yq(4,j+3)=yy2-w/2;
    yq(1,j+5)=yy1+w/2;
    yq(2,j+5)=yy1+w/2;
    yq(3,j+5)=yy2-w/2;
    yq(4,j+5)=yy2-w/2;

    %% increase the length of first trace
    if i==1 | i==node-1
      yq(2,j+3)=yy1-w/2;
      yq(3,j+3)=yy1-w/2;
      yq(1,j+5)=yy1-w/2;
      yq(2,j+5)=yy1-w/2;
    end

    for k=1:4,
      yq(k,j+1)=yy1+w/2;
      %% increase the length of first trace
      if i==1 | i==node-1
        yq(k,j+1)=yy1-w/2;
      end
      yq(k,j+2)=yy2-w/2;
      xq(k,j+3)=xx1+w/2;
      xq(k,j+4)=xx1-w/2;

      xq(k,j+2)=xq(k,j+1);
      xq(k,j+5)=xq(k,j+1);
      xq(k,j+6)=xq(k,j+5);
      yq(k,j+4)=yq(k,j+3);
      yq(k,j+6)=yq(k,j+5);
    end

    %%cancel panel 1
    if i~=1 & i~=node-1
      for k=1:4,
xq(k,j+1)=0;
yq(k,j+1)=0;
zq(k,j+1)=0;
      end
    end

  else
    yq(1,j+1)=yy1+w/2;
    yq(2,j+1)=yy1-w/2;
    yq(3,j+1)=yq(2,j+1);
    yq(4,j+1)=yq(1,j+1);

    xq(1,j+3)=xx2+w/2;
    xq(2,j+3)=xx1-w/2;
    xq(3,j+3)=xx1-w/2;
    xq(4,j+3)=xx2+w/2;
    xq(1,j+5)=xx1-w/2;
    xq(2,j+5)=xx1-w/2;
    xq(3,j+5)=xx2+w/2;
    xq(4,j+5)=xx2+w/2;
```

```
    for k=1:4,
     xq(k,j+1)=xx1-w/2;
     xq(k,j+2)=xx2+w/2;
     yq(k,j+3)=yy1-w/2;
     yq(k,j+4)=yy1+w/2;

     yq(k,j+2)=yq(k,j+1);
     yq(k,j+5)=yq(k,j+1);
     yq(k,j+6)=yq(k,j+5);
     xq(k,j+4)=xq(k,j+3);
     xq(k,j+6)=xq(k,j+5);
    end
   end
  end

  hq =fill3(xq, yq, zq, color);
  hold on;
```

A.2 Programs for Transformer Simulations

The main program is `trans_simulation.m` which takes the number of turns, the center hole spacing, the trace width, the trace spacing, and a file name as the input parameters. Its output is a graphical presentation of the primary's quality factor, self-resonant frequency, and one-port impedance, as illustrated in Chapter 5. Its subroutine, `z2input.m` generates an input file to *FastHenry* for the computation of the self-inductance, mutual coupling coefficient, and metal resistance including the skin effect of the transformer. It also invokes `parasitics.m` to calculate the primary's parasitics C_{ox}, C_c, R_{si}, and C_{si}. and draws a 3-D plot of the transformer. All programs, except for `q.m`, `substrate.m`, and `parasitic.m` which are listed in A.1, are listed below.

`trans_simulation.m`

```
function trans_simulation(n,center,w,d,fname);

%% function trans_simulation(n,center,w,d,fname);
%% w=the trace width; n=the number of turns for each inductor;
%% d=the spacing between adjacent traces;
%% center=the center hole spacing between center lines of traces(each ind);
%% fname=output filename

%% simulations for transformers;
```

```
henry=[fname, '.henry'];
par=[fname, '.par'];

%% generating input file to FastHenry; dumping the output to fname.henry;
%% computing the parasitics; dumping the outputs to fname.par
z2input(n,center,w,d,fname);

%% read in data from output files: fname.henry and fname.par
fid=fopen(henry,'r');dat1=fscanf(fid,'%f');
fid=fopen(par,'r');dat2=fscanf(fid,'%f');
n1=size(dat1,1);n2=size(dat2,1);

Cox=dat2(1:5:n2); %% in pF
Cc=dat2(2:5:n2); %% in pF
Csi=dat2(3:5:n2); %% in pF
Rsi=dat2(4:5:n2); %% in ohm
length=dat2(5:5:n2); %% in um

freq=dat1(1:4:n1)*1e-9; %% in GHz
resistance=dat1(2:4:n1); %% in ohm
inductance=(dat1(3:4:n1)+dat1(4:4:n1))./freq/2/pi; %% in nH
kc=dat1(4:4:n1)./dat1(3:4:n1); %% coupling coefficent


%eddy current loss; changed with different technologies
Rloss=freq.^2*2.8e-4;  %% ohm/um; Ref: Hasegawa et al

%% total series resistance
Rt=resistance+Rloss*length; %% in ohm;

%% computing the quality factor, the self-resonant frequency,
%% and the one-port impedance of the primary or the secondary
for i=1:n1/4,
  [Q(i),Fr(i),imag_imp(i),real_imp(i)]=q(freq(i),inductance(i),Rt(i),2*Cox,Cc,0);
end

%% plotting the quality factor, the self-resonant frequency,
%% and the one-port impedance of the primary or the secondary
figure;
subplot(311);
semilogx(freq*1e9,real_imp,'-',freq*1e9,Rt,'--');
grid on;
ylabel('Real Impedance Ohms');
axis([1e8 1e10 0 265]);

subplot(312);
semilogx(freq*1e9,imag_imp, '-',freq*1e9,inductance,'--');
grid on;
ylabel('Imaginary Impedance nH');
axis([1e8 1e10 -7 9]);

subplot(313);
semilogx(freq*1e9,Q);
grid on;
ylabel('Quality Factor');
axis([1e8 1e10 -1 6]);
xlabel('Frequency Hz');

fclose('all');
```

z2input.m

```
function z2input(n,center,w,d,fname);

%% function z2input(n,center,w,d,fname);
%% FastHenry input file generator; no plot
%% transformer computation using fasthenry and parasitics.m
%% Transformer--2D coupling inductors
%% 03/26/97
%% w=the trace width; n=the number of turns for each inductor;
%% d=the spacing between adjacent traces;
%% center=the center hole spacing between center lines of traces(each ind);
%% fname=output filename

%% fasthenry output stored in fname.henry
%% parasitics output stored in fname.par and fname.dat


%% process information
rho=0.042;              %% metal resistivity; ohm-um
h=1.2;                  %% trace height, um


%%-------------------------------------

node=4*n+1;             %% number of nodes (each ind)
spacing=2*(w+d);        %% spacing between two center lines (each ind)
radius=(n-1)*spacing+center/2;

%% for the first inductor
length1=0;  %% total length of in-between microstrip lines, um
length2=0;  %% total length of outer-most microstrip lines, um
length3=0;  %% total length of inner-most microstrip lines, um


%% generate (x1,y1) for first inductor

x1(node)=0;
y1(1)=-radius;
for i=1:4:node-1,
  x1(i)=(i-1)/4*spacing-radius;
  x1(i+1)=x1(i);
  x1(i+2)=-x1(i);
  x1(i+3)=x1(i+2);
  y1(i+1)=x1(i+2);
  y1(i+2)=x1(i+2);
  y1(i+3)=-x1(i+3);
  y1(i+4)=y1(i+3);
end


%% generate (x2 ,y2) for the second inductor

for i=1:node,
  x2(i)=-x1(i)+spacing/2;
  y2(i)=-y1(i)-spacing/2;
end;


%%-------------------------------------
%% generate fasthenry input file: fname.inp
```

```
filename=[fname '.inp'];
fid = fopen(filename,'w');
dd=date; tt=fix(clock);

fprintf(fid, '*** planar spiral transformers--2D ***\n\n');
fprintf(fid, '*** %d:%d:%d PST, %s ***\n\n', tt(4),tt(5),tt(6),dd);
fprintf(fid, '.units um\n');

if w>25
 fprintf(fid, '.default z=0 w=%d h=%2.4f rho=%2.4f nhinc=1 nwinc=7\n\n\',w,h,rho);
else
 fprintf(fid, '.default z=0 w=%d h=%2.4f rho=%2.4f nhinc=1 nwinc=5\n\n\',w,h,rho);
end

fprintf(fid, '*** z2input(n=%d, center=%d, w=%d, d=%d, %s)\n\n', n,center,w,d,fname);
fprintf(fid, '*** spacing between two lines: d=%dum\n\n\n',d);

substrate(1.5*radius,x1,y1,x2,y2,w,filename);

fprintf(fid,'*** first inductor\n');

for i=1:node,
 fprintf(fid,'n%d  x=%4.2f  y=%4.2f\n',i,x1(i),y1(i));
 if i<3 %% in transformer, only node 1-2-3 is outmost edge
  length2=length2+(abs(x1(i+1)-x1(i))+abs(y1(i+1)-y1(i)));
 elseif (i<=node & i>node-2)
  length3=length3+(abs(x1(i)-x1(i-1))+abs(y1(i)-y1(i-1)));
 elseif i<node-2
  length1=length1+(abs(x1(i+1)-x1(i))+abs(y1(i+1)-y1(i)));
 end
end


fprintf(fid,'\n\n*** second inductor\n');

for i=1:node,
 fprintf(fid,'n%d  x=%4.2f  y=%4.2f\n',i+node,x2(i),y2(i));
end


fprintf(fid,'\n\n');

for i=1:node-1,
 fprintf(fid,'e%d  n%d  n%d\n',i,i,i+1);
end;
for i=node+1:2*node-1,
 fprintf(fid,'e%d  n%d  n%d\n',i,i,i+1);
end;

fprintf(fid,'\n\n');

fprintf(fid,'.external  n1  n%d\n', node);
fprintf(fid,'.external  n%d  n%d\n\n', node+1, 2*node);

fprintf(fid,'.freq fmin=0.9e+09 fmax=0.9e+09 ndec=9\n\n.end\n');


%%-------------------------------------
%% fasthenry computation:output fname.henry

eval(['!fasthenry -S _' fname ' ' filename]);
output1=[fname '.henry'];

eval(['!awk -f trans_henry.awk Zc_' fname '.mat >>' output1]);
```

```
%%-------------------------------------
%% parasitics computation:output fname.par

c1=(length1+length2+length3)*w*0.0092;
length=(length1+length3)+(w+d)*(4*(n-1)+2); %% pay attention
[Cox, Cc, Csi, Rsi]=parasitic(w,d,length1,length2,length3,length);
total_length=length1+length2+length3;

output2=[fname '.dat'];
fid = fopen(output2,'a');


fprintf(fid, '============================== planar spiral transformers--2D\n\n');
fprintf(fid, '*** %d:%d:%d PST, %s ***\n\n', tt(4),tt(5),tt(6),dd);

fprintf(fid, '*** z2input(n=%d, center=%d, w=%d, d=%d, %s)\n\n', n,center,w,d,fname);
fprintf(fid, '*** estimated capacitance: c1=%2.4ffF\n\n\n',c1);
fprintf(fid, 'length1=%2.2fum length2=%2.2fum length3=%2.2fum length=%2.2fum\n\n', length1,
length2, length3, length);
fprintf(fid, 'total_length= %2.2f um\n\n', total_length);
fprintf(fid, 'Cox= %2.4f pF Cc= %2.4f pF Csi= %2.4f pF Rsi= %2.4f ohm\n\n\n',Cox, Cc, Csi, Rsi);

output3=[fname '.par'];
fid = fopen(output3,'a');
fprintf(fid, '%2.4f %2.4f %2.4f %2.4f %2.2f\n',Cox, Cc, Csi,Rsi,total_length);


fclose('all');

%%-------------------------------------
%% 3D plot;

xmax=max([max(x1),max(y1),max(x2),max(y2)])-w-d;
xmin=min([min(x1),min(y1),min(x2),min(y2)])+w+d;

figure;
polyfill1(0,node,h,w,x1,y1,'y');
polyfill2(0,node,h,w,x2,y2, 'r');

axis([xmin,xmax+w+d,xmin-w-d,xmax,xmin/4,xmax/4]);
axis('equal');
axis off;
hold off;
%%-------------------------------------
```

polyfill1.m and polyfill2.m

```
function polyfill1(zz,node,h,w,x,y,color);
%% 3D plot for transformers
%% 03/27/97
%% w=width of trace;
%% h=trace height
%% zz=z-axis position
%% node=number of nodes
%% x,y= position of each node

d=10; %% trace spacing for plotting

for i=1:node-1,
```

```
j=(i-1)*6;
zq(1,j+1)=zz-h/2;
zq(2,j+1)=zz-h/2;
zq(3,j+1)=zz+h/2;
zq(4,j+1)=zz+h/2;
for k=1:4,
 zq(k,j+5)=zz+h/2;
 zq(k,j+6)=zz-h/2;
 zq(k,j+2)=zq(k,j+1);
 zq(k,j+3)=zq(k,j+1);
 zq(k,j+4)=zq(k,j+3);
end
xx1=min(x(i),x(i+1));
xx2=max(x(i),x(i+1));
yy1=min(y(i),y(i+1));
yy2=max(y(i),y(i+1));

if xx1==xx2
 xq(1,j+1)=xx1+w/2;
 xq(2,j+1)=xx1-w/2;
 xq(3,j+1)=xq(2,j+1);
 xq(4,j+1)=xq(1,j+1);

 yq(1,j+3)=yy2-w/2;
 yq(2,j+3)=yy1+w/2;
 yq(3,j+3)=yy1+w/2;
 yq(4,j+3)=yy2-w/2;
 yq(1,j+5)=yy1+w/2;
 yq(2,j+5)=yy1+w/2;
 yq(3,j+5)=yy2-w/2;
 yq(4,j+5)=yy2-w/2;

 %% increase the length of first trace
 if i==1
  yq(2,j+3)=yy1-3*w/2-d;
  yq(3,j+3)=yy1-3*w/2-d;
  yq(1,j+5)=yy1-3*w/2-d;
  yq(2,j+5)=yy1-3*w/2-d;
 end

 for k=1:4,
  yq(k,j+1)=yy1+w/2;
  %% increase the length of first trace
  if i==1
   yq(k,j+1)=yy1-3*w/2-d;
  end
  yq(k,j+2)=yy2-w/2;
  xq(k,j+3)=xx1+w/2;
  xq(k,j+4)=xx1-w/2;

  xq(k,j+2)=xq(k,j+1);
  xq(k,j+5)=xq(k,j+1);
  xq(k,j+6)=xq(k,j+5);
  yq(k,j+4)=yq(k,j+3);
  yq(k,j+6)=yq(k,j+5);
 end

 %%cancel panel 1
 if i~=1 & i~=node-1
  for k=1:4,
xq(k,j+1)=0;
yq(k,j+1)=0;
zq(k,j+1)=0;
  end
```

```
   end

  else
   yq(1,j+1)=yy1+w/2;
   yq(2,j+1)=yy1-w/2;
   yq(3,j+1)=yq(2,j+1);
   yq(4,j+1)=yq(1,j+1);

   xq(1,j+3)=xx2+w/2;
   xq(2,j+3)=xx1-w/2;
   xq(3,j+3)=xx1-w/2;
   xq(4,j+3)=xx2+w/2;
   xq(1,j+5)=xx1-w/2;
   xq(2,j+5)=xx1-w/2;
   xq(3,j+5)=xx2+w/2;
   xq(4,j+5)=xx2+w/2;

   for k=1:4,
    xq(k,j+1)=xx1-w/2;
    xq(k,j+2)=xx2+w/2;
    yq(k,j+3)=yy1-w/2;
    yq(k,j+4)=yy1+w/2;

    yq(k,j+2)=yq(k,j+1);
    yq(k,j+5)=yq(k,j+1);
    yq(k,j+6)=yq(k,j+5);
    xq(k,j+4)=xq(k,j+3);
    xq(k,j+6)=xq(k,j+5);
   end
  end
end


hq =fill3(xq, yq, zq, color);
hold on;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


function polyfill2(zz,node,h,w,x,y,color);
%% 3D plot for transformers
%% 03/27/97
%% w=width of trace;
%% h=trace height
%% zz=z-axis position
%% node=number of nodes
%% x,y= position of each node

d=10; %% trace spacing for plotting

for i=1:node-1,
 j=(i-1)*6;
 zq(1,j+1)=zz-h/2;
 zq(2,j+1)=zz-h/2;
 zq(3,j+1)=zz+h/2;
 zq(4,j+1)=zz+h/2;
 for k=1:4,
  zq(k,j+5)=zz+h/2;
  zq(k,j+6)=zz-h/2;
  zq(k,j+2)=zq(k,j+1);
  zq(k,j+3)=zq(k,j+1);
  zq(k,j+4)=zq(k,j+3);
 end
```

```
  xx1=min(x(i),x(i+1));
  xx2=max(x(i),x(i+1));
  yy1=min(y(i),y(i+1));
  yy2=max(y(i),y(i+1));

  if xx1==xx2
   xq(1,j+1)=xx1+w/2;
   xq(2,j+1)=xx1-w/2;
   xq(3,j+1)=xq(2,j+1);
   xq(4,j+1)=xq(1,j+1);

   yq(1,j+3)=yy2-w/2;
   yq(2,j+3)=yy1+w/2;
   yq(3,j+3)=yy1+w/2;
   yq(4,j+3)=yy2-w/2;
   yq(1,j+5)=yy1+w/2;
   yq(2,j+5)=yy1+w/2;
   yq(3,j+5)=yy2-w/2;
   yq(4,j+5)=yy2-w/2;

   %% increase the length of first trace
   if i==1
   yq(1,j+3)=yy2+3*w/2+d;
   yq(4,j+3)=yy2+3*w/2+d;
   yq(3,j+5)=yy2+3*w/2+d;
   yq(4,j+5)=yy2+3*w/2+d;
   end

   for k=1:4,
    yq(k,j+1)=yy1+w/2;
    %% increase the length of first trace
    if i==1
     yq(k,j+2)=yy2+3*w/2+d;
    end
    yq(k,j+2)=yy2-w/2;
    xq(k,j+3)=xx1+w/2;
    xq(k,j+4)=xx1-w/2;

    xq(k,j+2)=xq(k,j+1);
    xq(k,j+5)=xq(k,j+1);
    xq(k,j+6)=xq(k,j+5);
    yq(k,j+4)=yq(k,j+3);
    yq(k,j+6)=yq(k,j+5);
   end

   %%cancel panel 1
   if i~=1 & i~=node-1
    for k=1:4,
xq(k,j+1)=0;
yq(k,j+1)=0;
zq(k,j+1)=0;
    end
   end

  else
   yq(1,j+1)=yy1+w/2;
   yq(2,j+1)=yy1-w/2;
   yq(3,j+1)=yq(2,j+1);
   yq(4,j+1)=yq(1,j+1);

   xq(1,j+3)=xx2+w/2;
   xq(2,j+3)=xx1-w/2;
   xq(3,j+3)=xx1-w/2;
   xq(4,j+3)=xx2+w/2;
```

```
    xq(1,j+5)=xx1-w/2;
    xq(2,j+5)=xx1-w/2;
    xq(3,j+5)=xx2+w/2;
    xq(4,j+5)=xx2+w/2;

    for k=1:4,
     xq(k,j+1)=xx1-w/2;
     xq(k,j+2)=xx2+w/2;
     yq(k,j+3)=yy1-w/2;
     yq(k,j+4)=yy1+w/2;

     yq(k,j+2)=yq(k,j+1);
     yq(k,j+5)=yq(k,j+1);
     yq(k,j+6)=yq(k,j+5);
     xq(k,j+4)=xq(k,j+3);
     xq(k,j+6)=xq(k,j+5);
    end
   end
  end


  hq =fill3(xq, yq, zq, color);
  set(hq,'EdgeColor', 'k');
  hold on;
```

Appendix B Unity-Gain Frequencies of RF MOSFETs

Two figures of merit are commonly used to describe RF transistor performance, especially for BJTs and GaAs FETs: The unity-current-gain frequency f_T and the unity-power-gain frequency f_{max}. In this appendix, we briefly review the derivation of these two figures of merit for MOSFETs.

f_T is the frequency at which the short-circuit current gain approximates unity. Based on the small-signal equivalent circuit of a MOSFET as shown in Fig. A-1, f_T is easily obtained as

$$f_T \approx \frac{g_m}{2\pi(C_{gs}+C_{gd})} \tag{A.1}$$

For a MOSFET in saturation, C_{gd} is usually much smaller than C_{gs}. Thus the above expression can be further simplified to

$$f_T \approx \frac{g_m}{2\pi C_{gs}} \tag{A.2}$$

f_{max} is the frequency at which the maximum available power gain G_A of the MOSFET is equal to 1. The maximum available power gain G_A is independent of the load and so is the f_{max}. They can be obtained by conjugately matching the source impedance to the transistor input impedance and the load to the transistor output impedance. Given an input current i_{in}, it is seen from the small-signal circuit shown in Fig. A-1 that the input power is $i_{in}^2 R_g$. The output current i_{out} is

$$i_{out} = \frac{2\pi f_T}{\omega} \cdot i_{in} \tag{A.3}$$

and the output conductance of the MOSFET is given by

$$g_{out} = \frac{g_m C_{gd}}{C_{gd} + C_{gs}} + g_{ds} \approx 2\pi f_T C_{gd} + g_{ds} \tag{A.4}$$

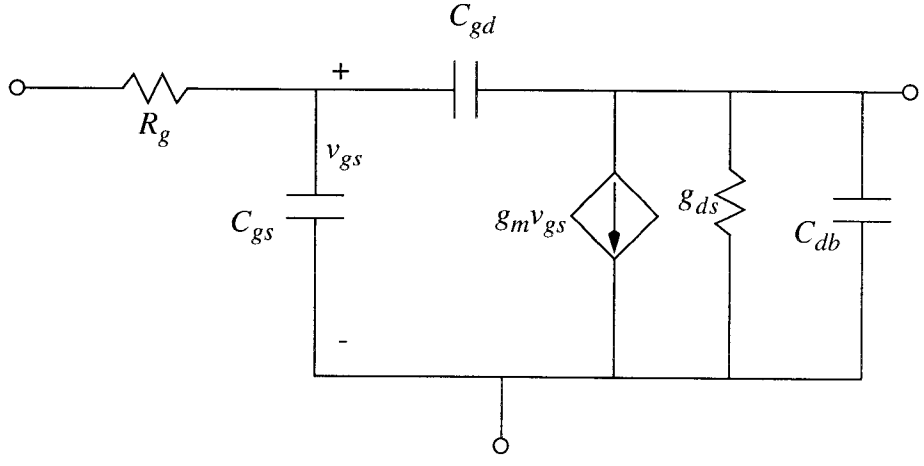


Figure A.1: Small-signal circuit of a MOSFET.

Thus the maximum output power is

$$P_{out} = \frac{i_{out}^2}{4g_{out}} \approx \frac{(2\pi f_T)^2 i_{in}^2}{4\omega^2 g_{out}} \tag{A.5}$$

The maximum available power gain G_A is then given by

$$G_A \approx \frac{(2\pi f_T)^2}{4\omega^2 g_{out} R_g} \tag{A.6}$$

and

$$f_{max} = \frac{f_T}{2\sqrt{g_{out} R_g}} \tag{A.7}$$

If $2\pi f_T C_{gd} \gg g_{ds}$, then

$$f_{max} \approx \sqrt{\frac{f_T}{8\pi C_{gd} R_g}} \tag{A.8}$$

If $2\pi f_T C_{gd} \ll g_{ds}$, then

$$f_{max} \approx \frac{f_T}{2\sqrt{g_{ds} R_g}} \tag{A.9}$$

In either case, f_{max} is considerably larger than f_T since the gate resistance R_g of a MOSFET is usually very small.

f_T and f_{max} are a first-order indications of the high-frequency performance of RF transistors. Transistors are useful as amplifiers and oscillators as determined roughly by these figures of merit. It is relative easier to deal with power than voltages or currents at radio frequencies. Therefore, f_{max} is more often used as an indication of the maximum frequency that a transistor is active to amplify signal power or to oscillate.