CIRCUIT AND SYSTEM DESIGN FOR MM-WAVE RADIO AND RADAR APPLICATIONS

BY

IOANNIS SARKAS

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

© Ioannis Sarkas, 2013

Circuit and System Design for mm-wave Radio and Radar Applications

Ioannis Sarkas
Doctor of Philosophy, 2013

Graduate Department of Electrical and Computer Engineering
University of Toronto

Abstract

Recent advancements in silicon technology have paved the way for the development of integrated transceivers operating well inside the mm-wave frequency range (30 - 300 GHz). This band offers opportunities for new applications such as remote sensing, short range radar, active imaging and multi-Gb/s radios. This thesis presents new ideas at the circuit and system level for a variety of such applications, up to 145 GHz and in both state-of-the-art nanoscale CMOS and SiGe BiCMOS technologies.

After reviewing the theory of operation behind linear and power amplifiers, a purely digital, scalable solution for power amplification that takes advantage of the significant f_T/f_{max} improvement in pFETs as a result of strain engineering in nanoscale CMOS is presented. The proposed Class-D power amplifier, features a stacked, cascode CMOS inverter output stage, which facilitates high voltage operation while employing only thin-oxide devices in a 45 nm SOI CMOS process.

Next, a single-chip, 70-80 GHz wireless transceiver for last-mile point-to-point links is described. The transceiver was fabricated in a 130 nm SiGe BiCMOS technology and can operate at data rates in excess of 18 Gb/s. The high bitrate is accomplished by taking advantage of the ample bandwidth available at the W-band frequency range, as well as by employing a direct QPSK modulator, which eliminates the need for separate upconversion

and power amplification.

Lastly, the system and circuit level implementation of a mm-wave precision distance and velocity sensor at 122 and 145 GHz is presented. Both systems feature a heterodyne architecture to mitigate the receiver 1/f noise, as well as self-test and calibration capabilities along with simple packaging techniques to reduce the overall system cost.

Acknowledgements And Dedication

I would like to thank Professor Sorin P. Voinigescu for his invaluable advice and ample guidance over all the years of my graduate studies.

I am also grateful to all my colleagues from BA4182, especially Andreea Balteanu and Kenneth Yau. Without your help, the many long nights would have been impossible. Many thanks go to Juergen Hasch and Mekdes Girma of Robert Bosch for their great assistance with the mm-wave radar sensors. The aid of Pascal Chevalier and Gregory Avenier of STMicroelectronics with the SiGe BiCMOS process was also valuable.

Finally, I would like to thank my parents, my brother and all my friends for their endless care and support.

This work is dedicated to the memory of my grandmother, Andriana.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BEOL	Back-end of the line
BER	Bit Error Rate
BERT	Bit Error Rate Tester
BIST	Built-in self-test
BOX	Buried oxide
CB	Common-base
CE	Common-emitter
CW	Continuous Wave
DAC	Digital-to-Analog Converter
DDS	Direct Digital Synthesizer
DSB	Double-Sideband
DSP	Digital Signal Processor
EER	Envelope elimination and restoration
FCC	Federal Communications Commission
FET	Field Effect Transistor
FFT	Fast Fourier Transform
FMCW	Frequency Modulated Continuous Wave

FOM	Figure-of-Merit
HBT	Heterojunction Bipolar Transistor
IC	Integrated Circuit
IDFT	Inverse Discrete Fourier Transform
IF	Intermediate Frequency
IP1dB	1-dB Input Compression Point
LINC	Linear Amplification with Non-Linear Components
LNA	Low-Noise Amplifier
LO	Local Oscillator
LTI	Linear Time Invariant
MiM	Metal-Insulator-Metal
MIMO	Multiple-In Multiple-Out
MoM	Metal-Oxide-Metal
MUX	Multiplexer
PA	Power amplifier
PAE	Power Added Efficiency
PLL	Phase Locked Loop
PSD	Power Spectral Density
QPSK	Quadrature phase shift keying
RMS	Root mean square
RX	Receiver
SFR	Step Frequency Radar
SiGe	Silicon Germanium
SNR	Signal-to-Noise ratio

SOI	Silicon-on-Insulator
SSB	Single-Sideband
STI	Shallow trench isolation
TA	Transmit Amplifier
TDR	Time Domain Reflectometer
TIA	Trans-Impedance Amplifier
TOF	Time-of-flight
TX	Transmitter
TXRX	Transceiver
VCO	Voltage Controlled Oscillator
VNA	Vector Network Analyzer

1 Introduction

It is an undisputable fact that wireless technology has, and keeps, changing our everyday lives. From the cell-phone revolution of the 1990s to the advent of the smart-phone, a large percentage of the population uses mobile devices that integrate at least three radios.

This unprecedented success has led engineers not only to continuously improve on the existing wireless standards, but also to explore new potential applications, such as gigabit radio, radar and imaging. A natural choice for many of the new systems was to exploit the large, mostly unlicensed bandwidth available at the less crowded mm-wave frequency range, extending from 30 GHz to 300 GHz. This was also made possible with the aid of silicon technology which, through constant scaling, has reached cut-off frequencies in excess of 200 GHz. At the time of writing of this thesis, products have already appeared in the market at 57-64 GHz and 71-76 GHz for data communications and at 77 GHz for automotive radar.

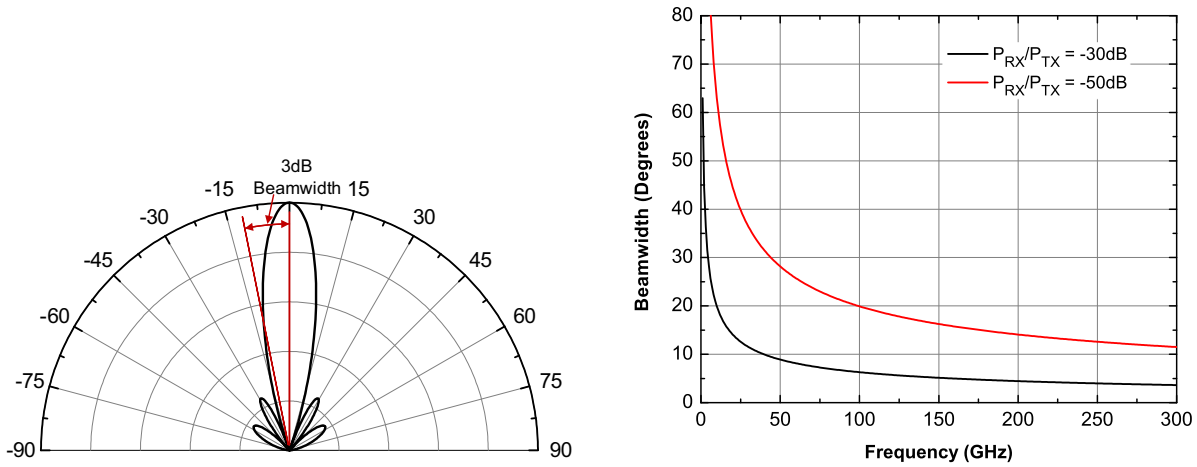
The purpose of this thesis is to present new ideas at the circuit and system level for a variety of high frequency applications.

1.1. Motivation

The mm-wave frequency range offers several GHz of unlicensed bandwidth that can be employed for bandwidth-intensive applications such as data communication and radar. However, there are limitations stemming from the fundamental laws of propagation and from antenna design constraints, as well as from shortcomings of the semiconductor technology that can limit the application space.

If a transmitter generates a signal of power P_{TX} at frequency f, that travels a distance R before reaching the receiver, the received power is predicted by the Friis' transmission equation:

$$P_{\mathrm{RX}} = P_{\mathrm{TX}} G_{\mathrm{RX}} G_{\mathrm{TX}} \left(\frac{c}{4\pi R f} \right)^2 \tag{1.1}$$



(a) Typical directional antenna radiation pattern and 3 dB beamwidth.

(b) Antenna 3 dB beamwidth for two values of $P_{\mathrm{RX}}/P_{\mathrm{TX}}$ for a distance of 10 m.

Figure 1.1: Directional antenna characteristics.

where G_{RX} and G_{TX} are the transmitter and receiver antenna gains, respectively, and c is the speed of light. The term $(c/4\pi Rf)^2$, known as the free space loss, indicates that the received power reduces by the square of the frequency.

To recover the lost power, either the transmitter power or the antenna gain need to increase. The first option is the most challenging and will be the subject of chapter 3. The second option is more attractive since the antenna gain is proportional to the antenna size and to the frequency of operation:

$$G = k\frac{d^2 f^2}{c^2} \tag{1.2}$$

where d is the antenna diameter and k is an antenna type dependent constant.

Increasing the gain comes at a cost. The antenna 3 dB-beamwidth, i.e. the angle at which the received power drops by 3 dB with respect to the maximum (figure 1.1a), is inversely proportional to its gain[1]:

$$\theta = \frac{4}{\sqrt{G}} \tag{1.3}$$

Therefore, although it is beneficial to design high gain antennas at higher frequencies, due to their smaller size, the resulting beamwidth is significantly smaller. This highlights the fact that the antenna "gain" is not a typical power gain, as for example defined in amplifiers. It is merely a focusing factor that compares the antenna to the isotropic radiator.

Low antenna beamwidth implies that only one, or a few closely spaced receivers, that have their antennas well aligned with the transmitter, will be able to receive the necessary

[1]The antenna beamwidth is inversely proportional to its directivity. For simplicity, the directivity is assumed equal to the gain.

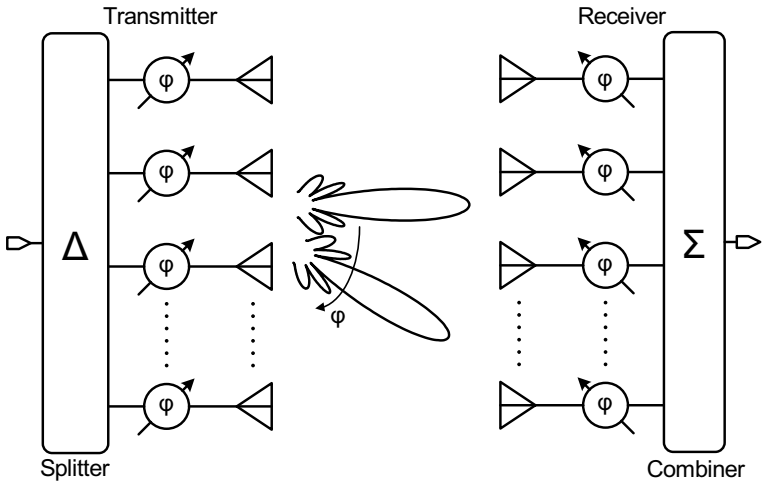


Figure 1.2: Phased array transceiver.

power, i.e, the link will be *directional*. Assuming that $G_{\mathrm{RX}} = G_{\mathrm{TX}}$ and R=10 m, figure 1.1b illustrates the beamwidth of the antennas that are required to maintain a certain $P_{\mathrm{RX}}/P_{\mathrm{TX}}$ ratio over frequency. The beamwidth reduces significantly over frequency, even for a small distance of 10 m.

The above treatment did not consider the system bandwidth. To examine the implications of wideband operation, the radio receiver input referred noise level needs to be considered:

$$N_{\mathrm{RX}} = k_B T B F \tag{1.4}$$

where k_B is the Boltzmann's constant, T is the temperature, F is the receiver noise figure and B is the bandwidth. As the bandwidth increases, the receiver noise will also increase, degrading the overall system signal-to-noise ratio (SNR) $P_{\mathrm{RX}}/N_{\mathrm{RX}}$. Consequently, to maintain a constant SNR over frequency, an even higher P_{RX} is required which translates to even more directional antennas.

A directive link can be made more versatile by invoking phased arrays, as illustrated in figure 1.2. In this case, instead of a single antenna, several lower gain elements are fed with certain phase shift. When the outputs combine in the free space, an equivalent directive pattern is formed, but the direction of its peak can be manipulated based on the applied phase shifts. This approach was adopted in 60 GHz radio and 77 GHz automotive radar, but the increased complexity of the system comes at the cost of considerably increased power consumption.

Consequently, the mm-wave frequency range offers significant bandwidth and antenna size benefits, but is suitable only for applications were either a high transmitter output power can be generated, or a very directive link can be tolerated. This thesis will investigate both cases.

Chapter 3 will examine the problem of efficient power generation at high frequencies. On the directive link side, two systems that take advantage of the small form factor, high gain antennas and wide bandwidth available at mm-wave frequencies are proposed. Chapter 4 presents a wideband transceiver for point-to-point, last-mile communication links, where reception at only one point is required, and thus highly directive antennas are employed to compensate most of the free-space loss. Chapters 5 and 6 propose a mm-wave precision distance sensor system, where the transmitter and receiver are co-located and employ the same partially integrated, high gain antenna.

1.2. State of the Art

The design of high-frequency circuits, even deep inside the mm-wave band, can be considered today a well developed art. Current Gallium-Arsenide (GaAs) and Indium-Phosphide (InP) integrated circuits (ICs) can operate at terahertz frequencies. For example, in [3], a 650 GHz low-noise amplifier with 30 dB gain is reported in a 30 nm InP High Electron Mobility Transistor (HEMT) process along with a 670 GHz receiver with 15 dB noise figure. Similarly, [4] presents 15 dB gain, 460 GHz amplifiers in a GaAs HEMT process and [5] reports a 630 GHz transmitter with an integrated phased lock loop in a 130 nm InP Heterojunction Bipolar Transistor (HBT) process.

Silicon integrated circuits, which were deemed as the last frontier for silicon, have also broken into the mm-wave band after approximatively 2004. Today, there are products at 60 GHz for gigabit-radio applications in 65 nm CMOS [6] and at 77 GHz for automotive adaptive cruise control radar [7, 8] in Silicon-Germanium (SiGe). More recently, there has been considerable interest in the 70-76 GHz and 80-86 GHz bands (E-band) for point-to-point communications. Several books on mm-wave design in silicon have also been published [9–12].

Transceivers in SiGe have been demonstrated at 160 GHz [13] and transmitters and receivers up to 650 GHz [14] and 820 GHz [15], albeit with inferior performance compared to corresponding GaAs and InP ICs. In CMOS, amplifiers operating at 145 GHz with 12 dB gain [16] and at 150 GHz with 8 dB gain [17] in 65 nm have been reported. More recently, oscillators at 300 GHz [18, 19] in 65 nm and arrays of frequency multipliers at 180 and 370 GHz [20, 21] in 45 nm have been demonstrated. Furthermore, considerable research interest has been focused on the operation of Field-Effect Transistors (FETs) as a power detectors at terahertz frequencies [22]. Terahertz imagers [23, 24] and even a camera have recently been demonstrated [25]. Nevertheless, in almost all cases, CMOS circuits cannot match the performance of their SiGe and III-V counterparts.

State-of-the-art SiGe HBTs have reached a maximum frequency of oscillation, f_{max}, in the order of 500 GHz [26,27]. Such technologies provide adequate margin for the implementation

of production-quality circuits up to 200 GHz. The performance of CMOS on the other hand, after significant performance improvement in the 90 and 65 nm nodes, seems to have flattened at f_{max} frequencies of approximately 250 GHz for fully metalized transistors. This is confirmed by the measurements presented in chapter 2 as well as in [20] for 45 nm SOI and in [28, 29] for 32 nm.

1.3. Thesis Outline

The remainder of this thesis consists of six chapters. Chapter 2 reviews the theoretical framework behind high frequency design. The necessary background of linear amplifier design for maximum power gain and low noise is provided. The design and performance of passive elements is reviewed, and the most important high-frequency layout techniques are presented. Chapter 3 proceeds to the analysis of power amplifiers, which need to be considered as non-linear elements. The most important challenges in their design are summarized, along with the existing solutions. A new circuit is proposed and the equations that govern its operation and its design methodology are presented.

The following chapters deal progressively more with system level analysis and design. Chapter 4 presents a transceiver for E-band communications that features an alternative transmitter architecture to facilitate QPSK modulation. Chapter 5 reviews some of the system level considerations behind radars, and selects the parameters for the proposed distance measurement system. Chapter 6 presents two implementations of the radar sensor proposed in chapter 5, at 122 GHz and 145 GHz. Finally, chapter 7 concludes this thesis and discusses potential future research.

1.4. Contributions

The key contributions of this thesis are summarized below.

- A new circuit for high frequency power amplification in nanoscale CMOS, that is applicable in the lower mm-wave frequency range and in the more traditional 2 and 5 GHz bands is proposed. It allows for class-D switch mode operation, while minimizing the use of passive components and impedance transformation networks.

- A novel direct QPSK modulator, which eliminates the need for separate upconversion and power amplification, is proposed and is integrated inside a QPSK transceiver for E-band communications. The resulting high modulation quality allows for very high bitrates, up to 18 Gb/s.

- Two mm-wave precision distance sensors, at 122 GHz and 145 GHz, the first of their

kind, are presented. Both sensors feature a new heterodyne architecture that minimizes the impact of 1/f noise, along with self-calibration and self-test capabilities.

- The formulas (2.19), (2.25) and (2.26) are derived in chapter 2 and allow for the study of the amplifier gain sensitivity to impedance mismatches. The sensitivity terms are linked to the transistor equivalent circuit, allowing to examine the impact of semiconductor technology on amplifier design. Furthermore, a methodology for designing amplifiers for minimum noise contribution is proposed and electromagnetic simulations for various layout techniques are presented for the first time.

2 Elements of High Frequency Integrated Circuit Design

THE aim of this chapter is two-fold. First, to revisit some aspects of the high frequency design theory that the author has found to be very crucial in the design of the circuits presented throughout this work. Second, to attempt to shed some new light by presenting new formulations, methodologies and simulations when necessary. The ultimate goal is to provide useful, practical design insight to a future designer.

Section 2.1 deals with the small-signal amplifier design theory, which is the cornerstone of high frequency integrated circuit design. Almost all other active circuit elements, such as mixers, oscillators and even attenuators can be considered as extensions of amplifiers and can thus be designed and analyzed as such, along with some necessary additions [30,31]. For example, an oscillator can be modeled as an amplifier with frequency selective feedback, and an active mixer as an amplifier along with a switching element.

Section 2.2 reviews the subject of amplifier noise and in conjunction with section 2.1 provides a new design methodology for low-noise, high frequency amplifier design. Section 2.3 presents examples of integrated passive components and compares their performance in two different IC processes.

Finally, sections 2.4 and 2.5 bring forward some physical circuit layout design problems that arise at high frequencies, especially at mm-waves and beyond. Various techniques and methodologies to alleviate them are presented.

2.1. Transistor Gain and Small-Signal Amplifier Design

Many aspects of the theory of high frequency amplifier design, as well as almost all the associated circuit topologies used today, were developed in the 1930s and 1940s using vacuum tubes. At the time, amplifiers were not required to operate at very high frequency and the size of the passive elements used for impedance matching was not a concern, since the vacuum tubes were already bulky. In the widely adopted radio-frequency engineering book of the

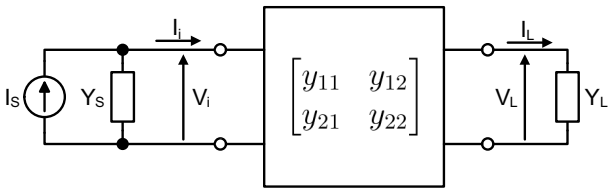


Figure 2.1: Two-port amplifier.

time by Frederick Terman [32] (its first edition was published in 1932), one can see that the tubes were considered almost ideal elements compared to today's complicated transistor models. Using transformers with magnetic cores for impedance transformation was the norm. Furthermore, because vacuum tubes suffered from significant grid-plate capacitance (the equivalent of the gate-drain and base-collector capacitance), neutralization, i.e. the cancelation of the internal feedback, was very common and many techniques to do so were developed [32].

As technology progressed and the equivalent circuits of active components became increasingly complicated, engineers understood the necessity to treat the amplifying element as a "black-box" and develop a general theory of amplifier design. The research on this topic started at least as early as 1948 [33] at Bell Labs, and was definitely accelerated by the invention of the bipolar transistor [34,35]. The theory was brought to its modern form by the end of the 1950s, with significant contributions by researchers such as John Linville [36,37], Arthur Stern [38], R. Pritchard [39] and S. Mason [40] (the list is definitely not complete). Later, in the 1960s , with the advent of the network analyzer and the S-parameter measurements, all the original formulas were re-derived in terms of S-parameters instead of y-parameters that were originally used. The core of the theory, however, remained unchanged.

This section reviews the classical two-port amplifier design theory and provides a slightly different point of view than most textbooks (e.g. [41]). The available and operating power gains (defined below) will be considered to be deviations from the maximum power gain, an approach that was originally introduced by H. Fukui [42]. To do so, a new formula (2.19) will be derived for the operating power gain that is the equivalent of Fukui's equation for the available gain. This formulation leads to equations (2.25) and (2.26) that allow to draw conclusions on the sensitivity of the power gain of the amplifier to its termination impedances and therefore, to impedance mismatches introduced by non-idealities in the transistor models or deviations of the matching networks, both commonly encountered in Integrated Circuits (ICs).

2.1.1. Amplifier Design Theory

A small-signal amplifier of any configuration can be adequately described as a two-port network. As illustrated in figure 2.1, the amplifier design theory deals with the analysis and calculation of the optimum source and load admittances Y_S and Y_L that maximize the power gain. Techniques to maximize the gain could also rely on the use of feedback e.g. with the neutralization of the internal feedback of the device [43], but they will not be considered here.

The input current source is generally a sinusoidal tone of RMS amplitude I_S at frequency ω. The two-port network can then be described by its Y-parameters which are generally complex numbers:

$$\begin{aligned} y_{11} &= g_{11} + jb_{11} \\ y_{12} &= g_{12} + jb_{12} \\ y_{21} &= g_{21} + jb_{21} \\ y_{22} &= g_{22} + jb_{22} \end{aligned} \tag{2.1}$$

The source and load admittances are also represented by complex numbers:

$$\begin{aligned} Y_S &= G_S + jB_S \\ Y_L &= G_L + jB_L \end{aligned} \tag{2.2}$$

For arbitrary Y_S and Y_L, the power gain of the amplifier is best described by its *transducer* power gain [44]:

$$G_{\mathrm{T}} = \frac{P_{\mathrm{L}}}{P_{\mathrm{avs}}} = \frac{4G_S G_L |y_{21}|^2}{|(y_{11} + Y_S)(y_{22} + Y_L) - y_{12}y_{21}|^2} \tag{2.3}$$

where $P_{\mathrm{L}} = \mathrm{Re}(V_{\mathrm{L}} I_{\mathrm{L}}^*)$ is the power delivered by the two-port to the load and

$$P_{\mathrm{avs}} = \frac{|I_S|^2}{4\,\mathrm{Re}(Y_S)} \tag{2.4}$$

is the *available* power from the source, which corresponds to the maximum power that the source can provide to a conjugately matched load [45]. Essentially the transducer gain compares the power delivered to the load by the two-port, to the power that the source would deliver if it was directly connected to a load conjugately matched to its admittance, Y_S. The transducer gain depends on the amplifier Y-parameters as well as on the source and load admittances. In a $50\,\Omega$ environment, where $Y_S = Y_L = 50\,\Omega$ then $G_{\mathrm{T}} = |S_{21}|^2$.

The question that immediately arises is whether the source and load admittances can be

optimized in order to maximize the transducer power gain. The answer is intertwined with the concept of stability, which is determined by the (Linvill) stability factor associated with the two port network [44]:

$$K = \frac{2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})}{|y_{12}y_{21}|} \tag{2.5}$$

When $K \geq 1$, the two port network is *unconditionally stable*, while if $K < 1$, the two port is *conditionally stable*. A conditionally stable two port can provide infinite gain when Y_S and Y_L are selected appropriately, i.e. it can turn into an oscillator.

For an unconditionally stable two-port without external feedback, the maximum transducer gain, which is referred to as the *maximum power gain* or *maximum available gain*, $G_{\max}$, can be calculated by [44]:

$$G_{\max} = \frac{|y_{21}|^2}{2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21}) + \left[\left(2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})\right)^2 - |y_{12}y_{21}|^2\right]^{1/2}} \tag{2.6}$$

and is obtained when the source and load admittances are equal to $Y_{S,og}$ and $Y_{L,og}$:

$$Y_{S,og} = G_{S,og} + jB_{S,og} \tag{2.7}$$

$$Y_{L,og} = G_{L,og} + jB_{L,og} \tag{2.8}$$

where

$$G_{S,og} = \frac{1}{2g_{22}}\left[\left(2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})\right)^2 - |y_{12}y_{21}|^2\right]^{1/2} \tag{2.9}$$

$$B_{S,og} = -b_{11} + \frac{\mathrm{Im}(y_{12}y_{21})}{2g_{22}} \tag{2.10}$$

$$G_{L,og} = \frac{1}{2g_{11}}\left[\left(2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})\right)^2 - |y_{12}y_{21}|^2\right]^{1/2} \tag{2.11}$$

$$B_{L,og} = -b_{22} + \frac{\mathrm{Im}(y_{12}y_{21})}{2g_{11}} \tag{2.12}$$

Selecting $Y_S = Y_{S,og}$ and $Y_L = Y_{L,og}$ is equivalent to *simultaneously* conjugately matching both the input an output of the two-port, i.e. setting $Y_S = Y_{\mathrm{in}}^*$ and $Y_L = Y_{\mathrm{out}}^*$. As expected from the maximum power transfer theorem, these conditions maximize the power flow in and out of the two port and thus maximize the power gain.

In most practical cases, the circuit designer is not able to arbitrarily select Y_S and Y_L and set them equal to $Y_{S,og}$ and $Y_{L,og}$. Instead, the amplifier is presented with fixed source and load admittances which most commonly depend on the preceding or following stages, or are $50\,\Omega$ terminations. Gain optimization is still possible by employing impedance trans-

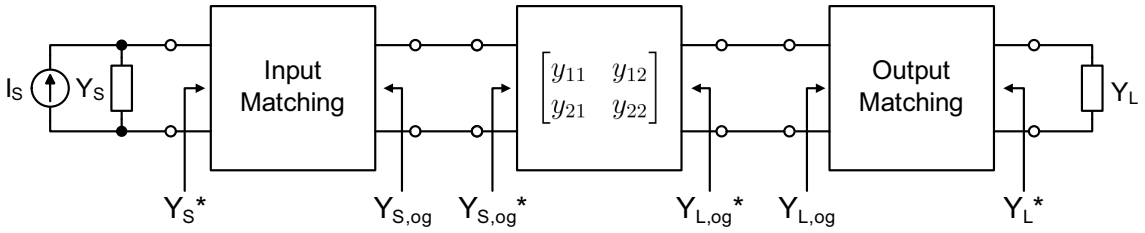


Figure 2.2: Two-port amplifier with matching networks.

formation networks, or matching networks, as illustrated in figure 2.2, which transform Y_S and Y_L into $Y_{S,og}$ and $Y_{L,og}$. There are many variants of such networks that one may select from [41, 45]. The choice usually depends on the ratio of the transformed impedances as well as on loss and bandwidth considerations. Very often, the selection simply depends on minimizing the number of components in the matching network.

Equation (2.6) for $G_{\max}$ does not provide adequate insight into its sensitivity on $Y_{S,og}$ and $Y_{L,og}$. This is especially important in integrated circuits where modeling inaccuracies lead to deviations from the optimum terminations. For this purpose, two different power gains need to be defined along with their dependence on $G_{\max}$.

The impact of the deviation from the optimum source matching can best be analyzed using the *available gain*[1], which is defined as the available power at the load, i.e. the maximum power that the two-port of figure 2.1 can provide to a conjugately matched load, over the power available from the source [44]:

$$G_{\text{av}} = \frac{P_{\text{avl}}}{P_{\text{avs}}} = \frac{|y_{21}|^2 G_S}{g_{22}|y_{11} + Y_S|^2 - \text{Re}[y_{12}y_{21}(y_{11} + Y_S)^*]} \tag{2.13}$$

The available power gain can also be interpreted as follows. The source admittance is first arbitrarily set to Y_S, resulting in the output admittance of the two-port to become:

$$Y_{\text{out}} = y_{22} - \frac{y_{12}y_{21}}{y_{11} + Y_S} \tag{2.14}$$

In order to achieve the maximum power delivery to the load, the load admittance is set to $Y_L = Y_{\text{out}}^*$, leading to equation (2.13). Obviously, G_{av} depends only on Y_S and when $Y_S = Y_{S,og}$, $G_{\text{av}} = G_{\max}$, i.e. G_{av} is maximized, but generally $G_{\text{av}} \leq G_{\max}$.

As proven in [42], the available gain can be expressed in terms of $G_{\max}$:

$$\frac{1}{G_{\text{av}}} = \frac{1}{G_{\max}} + \frac{R_{sg}}{G_S}|Y_S - Y_{S,og}|^2 = \frac{1}{G_{\max}} + \frac{R_{sg}}{G_S}\left[(G_S - G_{S,og})^2 + (B_S - B_{S,og})^2\right] \tag{2.15}$$

[1]The available gain is also referred to as associated gain in the literature.

where

$$R_{sg} = \frac{g_{22}}{|y_{21}|^2} \tag{2.16}$$

This expression is similar to the more familiar one of noise figure (F) in terms of $F_{\min}$ (section 2.2). It essentially describes G_{av} as a deviation from $G_{\max}$ due to the departure of the source admittance from its optimum value $Y_{S,og}$. Since the term $\left[(G_S - G_{S,og})^2 + (B_S - B_{S,og})^2\right]/G_S$ depends only on the source mismatch, the resistance R_{sg} essentially expresses the sensitivity of G_{av} to the departure from the optimum, with amplifiers with small R_{sg} being more robust.

As with the available gain, a similar expression can be defined that accounts only for the load mismatch. This role is fulfilled by the *operating power gain*[2] which is defined as the ratio of the power delivered to the the load, $P_{\text{L}} = \text{Re}(V_{\text{L}} I_{\text{L}}^*)$, over the power that is absorbed at input of the two-port, $P_{\text{i}} = \text{Re}(V_{\text{i}} I_{\text{i}}^*)$ (figure 2.1) [44]:

$$G_{\text{P}} = \frac{P_{\text{L}}}{P_{\text{i}}} = \frac{|y_{21}|^2 G_L}{|y_{22} + Y_L|^2 - \text{Re}[y_{11} - \frac{y_{12}y_{21}}{y_{22}+Y_L}]} \tag{2.17}$$

In accordance with G_{av}, G_{P} can be interpreted as the gain that would result if the two-port is terminated with an arbitrary load admittance Y_L, leading to the input admittance becoming:

$$Y_{\text{in}} = y_{11} - \frac{y_{12}y_{21}}{y_{22} + Y_L} \tag{2.18}$$

The source admittance is then set to $Y_S = Y_{\text{in}}^*$, resulting to maximum power transfer from the source. G_{P} depends only on Y_L and $G_{\text{P}} \leq G_{\max}$ with $G_{\text{P}} = G_{\max}$ when $Y_L = Y_{L,og}$.

Following a derivation similar to the one in Fukui's seminal work [42] for the available gain, the operating power gain can be shown to depend on $G_{\max}$ as:

$$\frac{1}{G_{\text{P}}} = \frac{1}{G_{\max}} + \frac{R_{lg}}{G_L}|Y_L - Y_{L,og}|^2 = \frac{1}{G_{\max}} + \frac{R_{lg}}{G_L}\left[(G_L - G_{L,og})^2 + (B_L - B_{L,og})^2\right] \tag{2.19}$$

where

$$R_{lg} = \frac{g_{11}}{|y_{21}|^2} \tag{2.20}$$

The term $\left[(G_L - G_{L,og})^2 + (B_L - B_{L,og})^2\right]/G_L$ expresses the deviation of the load admittance Y_L from its optimum value $Y_{L,og}$, while the resistance R_{lg} expresses the sensitivity of G_{P} to this mismatch.

Equations (2.15) and (2.19) represent constant G_{av} and G_{P} circles when plotted versus Y_S

[2]The operating power gain is sometimes referred to simply as power gain in the relevant literature. This can be confusing since all other gains are also power gains.

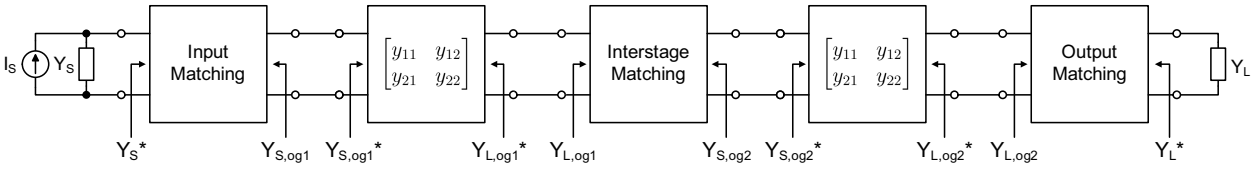


Figure 2.3: Multistage amplifier design.

and Y_L respectively. The center of these circles are the optimum source and load admittances $Y_{S,og}$ and $Y_{L,og}$ and the gain at the center of the circle is $G_{\max}$ in both cases.

Most textbooks that deal with amplifier design (e.g. [41,44]) provide lengthy equations for the constant G_{av} and G_{P} circles in terms of S-parameters. These equations are considerably more complicated that (2.15) and (2.19) and mask the simple interpretation of G_{av} and G_{P} as deviations from $G_{\max}$ due to source and load mismatch. In this work, the role of R_{sg} and R_{lg} as the sensitivities of the power gain to the termination admittances will be brought forward and will be further analyzed in terms of circuit parameters.

The above treatment has focused only on cases where $K \geq 1$. This is usually a valid assumption for mm-wave circuit design, since the transistor gain decreases with increasing frequency and naturally, the corresponding stability factor increases. Nevertheless, deeply-scaled SiGe HBTs [46] and FET devices can exhibit $K < 1$ well above 100 GHz. Even in these cases, when the unavoidable loss of the passive components employed in the matching networks is accounted for as part of the amplifier two-port, the resulting amplifier frequently becomes unconditionally stable. There are certain cases, most notably the HBT-only cascode, where its stability factor will persistently remain below unity, especially when biased at high currents. There are various work-arounds for these cases. The current density through the amplifier can be reduced, neutralization of the internal feedback [43, 47], resistive feedback stabilization [48] or inductive degeneration can be employed, or even selecting optimally mismatched source and load admittances [38, 49].

Finally, as the frequency of operation increases, it becomes increasingly difficult to squeeze the necessary power gain out of the single stage amplifier of figure 2.2. Therefore, it is common for mm-wave amplifiers to employ several amplification stages. The design methodology for multistage amplifiers is almost identical to the single stage case. Specifically, as illustrated in figure 2.3 for the two-stage case, an interstage matching network is introduced between the two amplifiers. The role of this matching network is to transform the optimum source admittance of second stage $Y^*_{S,og2}$ to the optimum load admittance of the first stage $Y^*_{L,og1}$ and vice-versa. Essentially the load of the first stage is the input admittance of the second, and therefore, it has to be appropriately transformed to properly load the first stage. The same approach can then be applied to an amplifier with any number of stages, or between

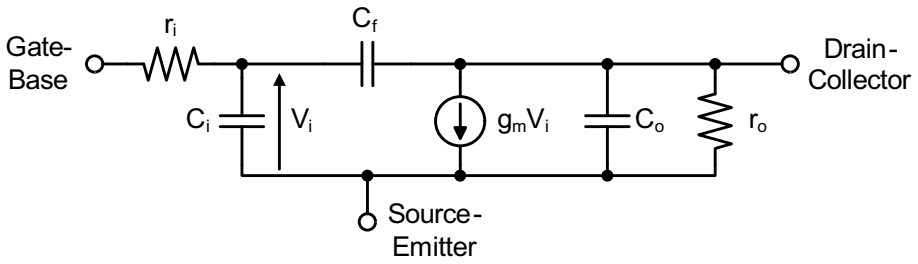


Figure 2.4: Generic transistor model.

any two circuits that need to be interfaced, e.g. between the amplifier and the mixer in a radio receiver.

2.1.2. Dependence of Gain on Transistor Parameters

The theory developed in the previous subsection is generic and can be employed for any amplifier, provided that its two-port parameters are known. Although they can be easily acquired from simulation, or even measurement, deriving the important formulas from the previous subsection in terms of circuit parameters can provide significant circuit design insight.

Both bipolar and field effect transistors at high frequencies can be represented by the same simplified hybrid-Π equivalent circuit [9] of figure 2.4. The correspondence between a typical FET hybrid-Π (e.g. in [50, 51]) and the circuit of figure 2.4 is: $r_i \simeq R_g$, $C_i = C_{gs}$, $C_f = C_{gd}$, $r_o = 1/g_{ds}$ and $C_o = C_{db}$ while for the bipolar transistor: $r_i \simeq R_b$, $C_i = C_\pi$, $C_f = C_\mu$ and $C_o = C_{cs}$. The equivalent circuit of the bipolar transistor also includes a resistor r_π in parallel to $C_i = C_\pi$. However, at high frequencies, the impedance of C_π is much smaller than r_π and effectively shunts it.

The y-parameters of the equivalent circuit of figure 2.4 in common-source/emitter configuration become [51, 52]:

$$\begin{aligned} y_{11} &= \frac{j\omega C_{in}}{1 + j\omega r_i C_{in}} \\ y_{12} &= \frac{-j\omega C_f}{1 + j\omega r_i C_{in}} \\ y_{21} &= \frac{g_m - j\omega C_f}{1 + j\omega r_i C_{in}} \\ y_{22} &= g_o + j\omega C_{out} + \frac{\omega^2 r_i C_f^2 + j\omega g_m r_i C_f}{1 + j\omega r_i C_{in}} \end{aligned} \tag{2.21}$$

where $C_{in} = C_i + C_f$, $C_{out} = C_o + C_f$ and $g_o = 1/r_o$.

Plugging the y-parameters into equation (2.6), one can calculate the maximum gain that

the common-source/emitter amplifier provides:

$$G_{\max} = \frac{g_m^2}{4(g_o r_i C_{in}^2 + g_m r_i C_{in} C_f)} \cdot \frac{1}{\omega^2} \simeq \frac{g_m}{4 C_{in} C_f r_i} \cdot \frac{1}{\omega^2} \tag{2.22}$$

where the second expression is derived by dropping the g_o terms, since $g_m \gg g_o$ in all modern semiconductor devices. Inspection of the above expression reveals that the maximum power gain drops by the *square* of the frequency, highlighting the most notorious difficulty of high frequency design.

The frequency at which $G_{\max} = 1$ is a widely adopted Figure-of-Merit (FoM) for semiconductor devices, known as the maximum frequency of oscillation $\omega_{\max}$ or $\mathrm{f_{max}}$:

$$\omega_{\max} = \sqrt{\frac{g_m}{4 C_{in} C_f r_i}} \tag{2.23}$$

Essentially $\omega_{\max}$ represents the theoretical frequency limit beyond which a semiconductor device is unusable for the design of amplifiers and oscillators. In practice, this frequency is considerably lower, close to a factor of two, than $\omega_{\max}$ due to the loss of the matching networks that are required to realize $G_{\max}$. Another important fact that stems from the above equations is that $G_{\max}$ and $\omega_{\max}$ are strongly coupled with the resistive parasitics of the device, most notable r_i, and they can be significantly improved by minimizing them. This is especially true for FET devices where r_i can be improved by optimizing the physical layout of the transistor, leading to $\mathrm{f_{max}}$ in excess of 200 GHz.

Another widely reported FoM of semiconductor devices is their cut-off frequency ω_{T} or $\mathrm{f_T}$. ω_{T} corresponds to the frequency up to which the device has current gain larger than unity. The current gain can be calculated by transforming the two-port y-parameters to h-parameters and solving for $h_{21} = 0$ [51], yielding:

$$\omega_{\mathrm{T}} = \frac{g_m}{C_{in}} \tag{2.24}$$

There is no straightforward link between high frequency amplifier performance and $\mathrm{f_T}$. A device that exhibits high $\mathrm{f_T}$ but low $\mathrm{f_{max}}$ due to increased r_i will not be suitable for high frequency applications. However, as will be shown below, a device with high $\mathrm{f_{max}}$ but low $\mathrm{f_T}$ will also suffer.

It is also instructive to calculate the sensitivity resistances R_{sg} and R_{lg} in terms of the transistor equivalent circuit elements. Substituting the expressions of the y-parameters from

(2.21) in (2.16) and (2.20) yields:

$$R_{sg} \simeq \frac{g_o(1+\omega^2 r_i^2 C_{in}^2)}{g_m^2} + \frac{\omega^2 r_i^2 C_f C_{in}}{g_m} \simeq \frac{\omega^2 r_i^2 C_f C_{in}}{g_m} = \frac{r_i}{4}\frac{\omega^2}{\omega_{\max}^2} \tag{2.25}$$

$$R_{lg} \simeq \frac{\omega^2 r_i C_{in}^2}{g_m^2} = r_i \frac{\omega^2}{\omega_{\mathrm{T}}^2} \tag{2.26}$$

where the original expressions have been simplified by dropping the C_f^2 terms from the denominators, without degrading the accuracy.

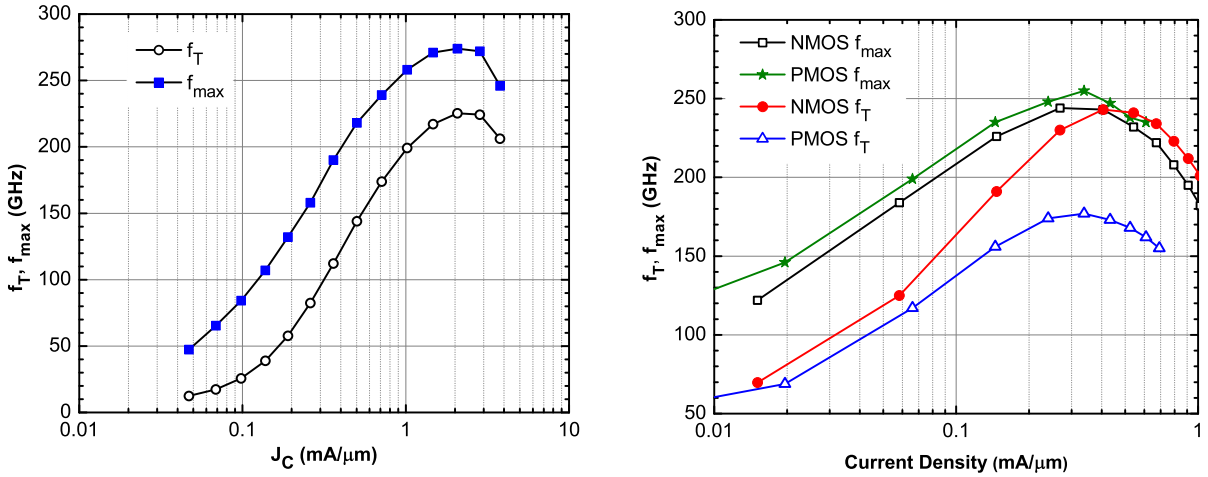
The above expressions highlight several previously unknown facts. First, the sensitivity of the power gain to source and load admittance mismatches is increasing with the *square* of the frequency. Therefore, not only the maximum power gain drops by the square of frequency but also the difficulty of realizing it in practice increases by the same rate. Second, as indicated by R_{lg}, the f_T of the device is also important in amplifier design. A device with low f_T but high f_{max} will suffer from increased sensitivity to its output matching network. This can be intuitively explained by considering the power gain as a product of voltage and current gains. A device with high power gain and low current gain needs to have very high voltage gain, which is attained by increasing the output resistance r_o. The maximum power transfer theorem requires that the optimum load for such a device is also r_o and any deviation from it will have an immediate impact on the voltage gain.

2.1.3. Performance of 130 nm SiGe HBTs and 45 nm FETs

In order to better visualize the above theory, as well as to assess the transistor performance, simulations and measurements were performed on transistors from the two integrated circuit processes that were used in this thesis. The first process is a 130 nm SiGe BiCMOS process from STMicroelectronics (BiCMOS9MW) [53] which was used for the transceiver circuits at 75, 120 and 145 GHz in chapters 4 and 6 respectively. The second process is a 45 nm Silicon-on-Insulator (SOI) CMOS process from IBM [54] which was employed for the power amplifier presented in chapter 3.

As expected from equations (2.23) and (2.24), f_{max} and f_T will depend on the transistor bias conditions and are thus usually plotted as a function of the transistor current density. Figure 2.5a reproduces the measured[3] f_T and f_{max} versus current density of a 4.5 μm SiGe HBT when its V_{CE} is 1.2 V. Its f_{max} is 280 GHz while its f_T is slightly lower, at 230 GHz. Similarly, figure 2.5b illustrates the measured[4] f_T and f_{max} of minimum gate length 45 nm CMOS FETs with $W = 40 \times 0.77\,\mu$m when V_{DS}=1.1 V. The nFET achieves almost symmetrical f_T

[3]The measurements of the SiGe HBTs were kindly provided by Kenneth Yau.
[4]The measurements of the CMOS transistors were kindly provided by Eric Dacquay.



(a) 130 nm SiGe BiCMOS CBEBC HBT with $L_e = 4.5\,\mu$m and $V_{CE} = 1.2$ V.

(b) 45 nm CMOS FETs with $W = 40 \times 0.77\,\mu$m, minimum L and $V_{DS} = 1.1$ V.

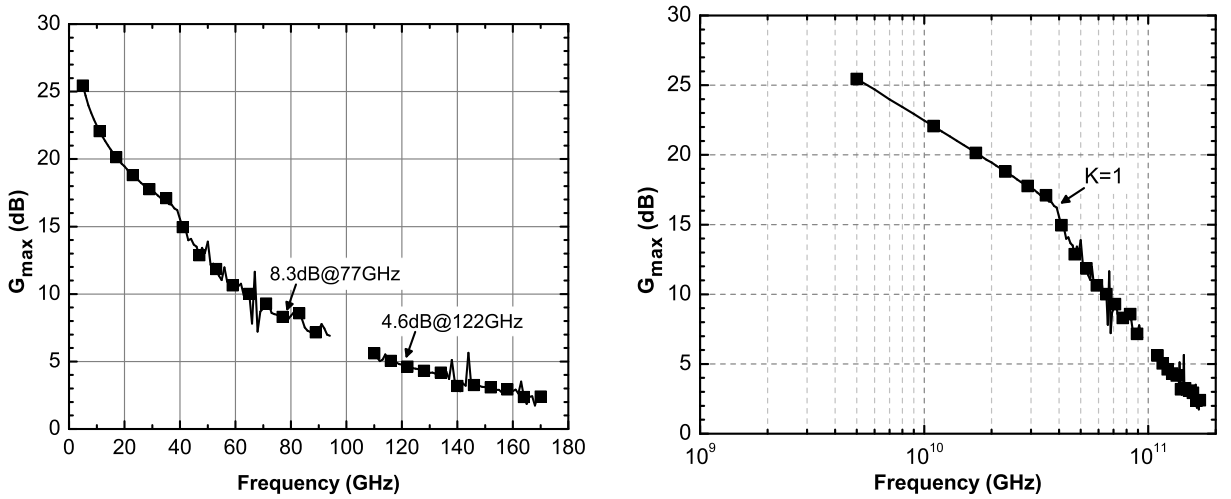
Figure 2.5: Measured f_T and f_{max} versus current density of the 130 nm SiGe BiCMOS and 45 nm CMOS processes.

and f_{max} of 250 GHz which occur at slightly different current densities. Surprisingly, the pFET also achieves a very high f_{max} of 250 GHz albeit with a lower f_T of 175 GHz, indicating that the pFET has lower g_m than the nFET but improved parasitics.

In both transistor cases the trends are the same: f_{max} and f_T first increase rapidly with current, until they reach a maximum at which the transistor operates at its maximum speed. The range of currents that result in this maximum is relatively wide, and is referred to as the peak-f_T current density [55]. Biasing at higher current than this will result in a rapid reduction of f_{max} and f_T. This occurs primarily due to the deterioration of g_m because of mobility degradation in FETs and the Kirk effect in bipolar transistors [51, 55]. Therefore, the peak-f_T current density is determined by the interaction between g_m and the parasitic capacitances and their associated dependence on the bias current.

Figure 2.6a shows the measured G_{max} versus frequency for the same SiGe HBT when biased at its peak-f_T current density. It is also plotted in figure 2.6b in a logarithmic frequency axis graph. As marked in the figure, after approximately 40 GHz, the slope changes to 20 dB per decade, which is expected from the $1/\omega^2$ decay of equation (2.22). At frequencies lower than 40 GHz, the transistor exhibits $K < 1$ and is potentially unstable. When $K < 1$, the G_{max} formula (2.6) reduces to $|y_{21}/y_{12}|$ which is known as the Maximum Stable Gain (MSG) and decays at a rate of $1/\omega$ (10 dB per decade). Contrary to what the name might imply, the MSG is generally *not* the maximum power gain that could be squeezed out of the device when it is stabilized to $K = 1$ [49].

Figure 2.7 illustrates the G_{max} of the 45 nm FETs with both linear and logarithmic



(a) Linear frequency axis.

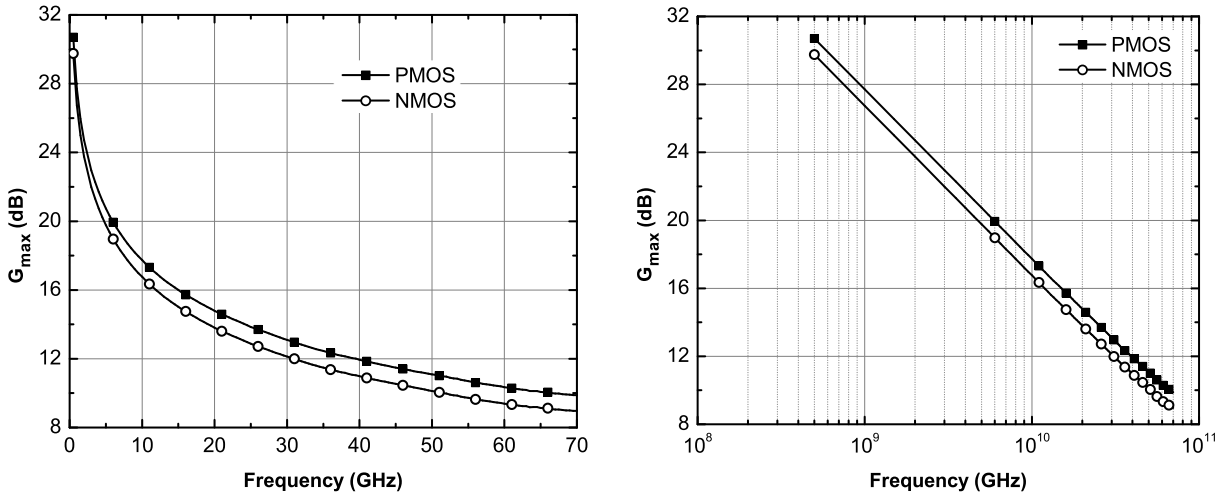
(b) Logarithmic frequency axis highlighting the point where $K = 1$.

Figure 2.6: Measured $G_{\max}$ versus frequency of the 130 nm SiGe BiCMOS 4.5 μm HBT at its peak-f_T current density.

frequency axes. Interestingly, the point where $K = 1$ is located at a frequency higher than 70 GHz, revealing that FETs are potentially unstable devices for most of their usable frequency range. This phenomenon stems from the fact that the high f_{max} of these devices is primarily obtained through layout optimization that leads to very low gate resistance R_g and in turn, from equation (2.21), to very small g_{11}. As implied by the formula for the stability factor (2.5), low g_{11} will lead to a degraded stability factor. Consequently, care needs to be exercised when designing FET amplifiers, because they are prone to stability problems.

To examine the sensitivity of the gain to admittance mismatches, the R_{sg} and R_{lg} of a common-emitter and of an HBT-only cascode amplifier have been simulated and are reproduced in figure 2.8. As expected from equations (2.25) and (2.26), they both increase with frequency. Since the SiGe HBT exhibits higher f_{max} than f_T, both amplifiers have higher R_{lg} than R_{sg} and thus higher sensitivity to output mismatches, with the cascode topology demonstrating almost twice as high R_{lg} than the common-emitter and slightly improved R_{sg}. As a result, although the cascode amplifier can provide higher power gain than the common-emitter, it is more difficult to realize it in practice due to its higher sensitivity to load mismatches.

The same simulations were repeated for the $40 \times 0.77\,\mu$m nFET common-source and cascode configurations in figure 2.9. Interestingly, although the nFET transistor has lower f_{max} than the HBT, it exhibits considerably lower sensitivity to mismatches, especially at the load side, due to its higher f_T and lower r_i. The nFET cascode shows similar behavior with its HBT-only counterpart, with slightly lower sensitivity to output matching, but still exceed-



(a) Linear frequency axis.

(b) Logarithmic frequency axis.

Figure 2.7: Measured $G_{\max}$ versus frequency of the 45 nm FETs at their peak-$f_{\max}$ current density.

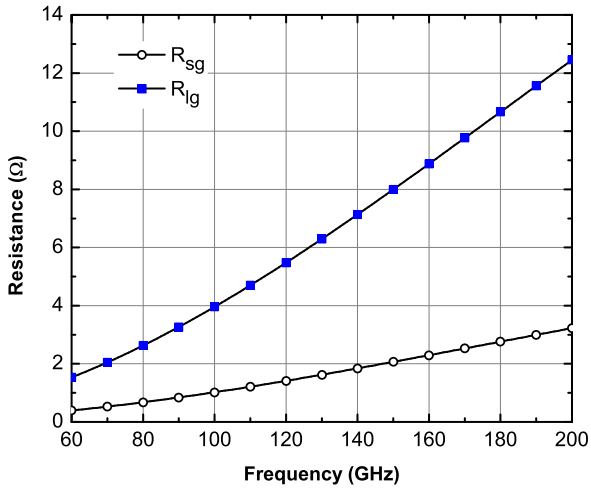
ingly higher than the common-source case. These results explain how several researchers achieved substantial success in designing CMOS mm-wave amplifiers at 94 GHz [20, 56–58] and even above 100 GHz [16, 17], with technologies that had inferior $f_{\max}$ than SiGe HBTs.

Lastly, figure 2.10 reproduces the simulated R_{sg} and R_{lg} versus current density for the common-emitter HBT configuration at 122 GHz. Both resistances exhibit a minimum value which is located at the peak-f_T current density of the transistor. This was anticipated since R_{sg} and R_{lg} are inversely proportional to $f_{\max}$ and f_T respectively (equations (2.25) and (2.26)). Therefore, biasing a transistor at its peak-f_T current density will not only yield its maximum power gain, but also the resulting amplifier will be less sensitive to deviations from the optimum termination impedances.

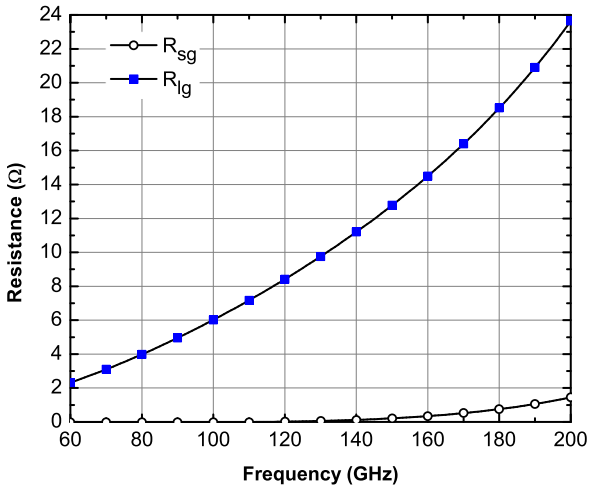
2.2. Noise Figure and Noise Measure

Every amplifier contaminates the signal it processes with noise. Therefore, along with the power gain theory presented in section 2.1, one needs to also consider the theory behind the high frequency noise. Historically, its development started along with the first vacuum tube radio receivers, e.g. the paper by Harald Friis that formally defined the concept of Noise Figure (F) appeared in 1944 [59] and refers to papers from back to 1930 within. However, the noise theory was brought to its modern, abstract form that employs two-port networks also in the 1950s by H. Rothe and W. Dahlke in [60] and with significant contributions by Herman Haus and Richard Adler [61].

In general, the internal noise sources of every two port network, and hence of every

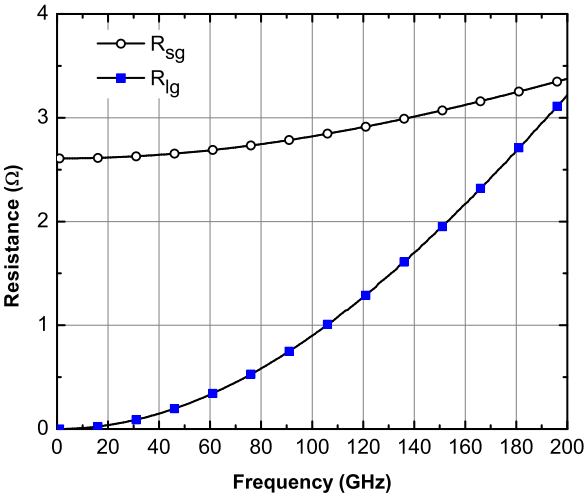


(a) Common-emitter.

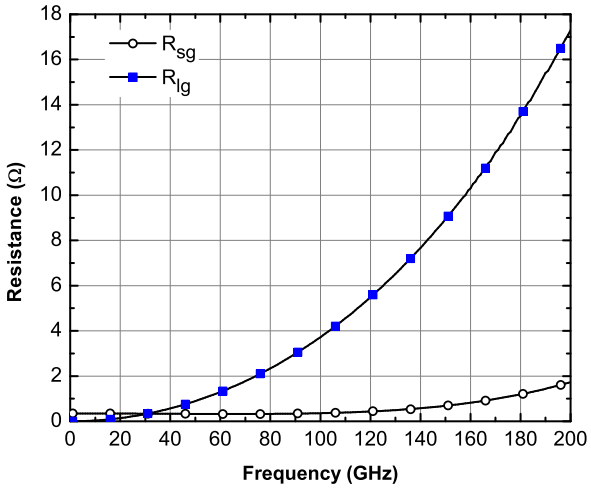


(b) HBT-only cascode.

Figure 2.8: Simulated R_{sg} and R_{lg} of 4.5 μm HBT amplifier configurations.



(a) Common-source.



(b) Cascode.

Figure 2.9: Simulated R_{sg} and R_{lg} of $40 \times 0.77\,\mu$m nFET amplifier configurations.

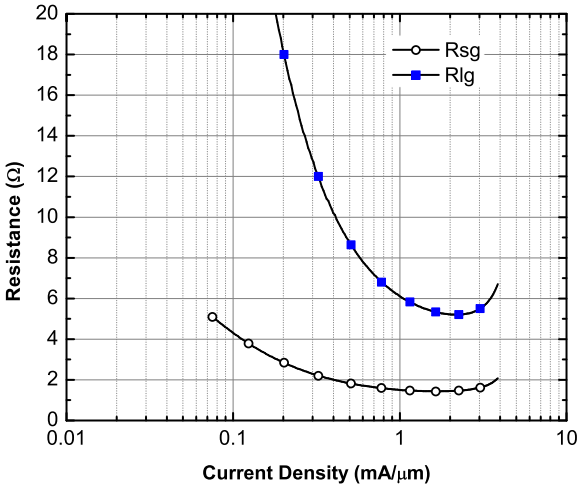


Figure 2.10: Simulated R_{sg} and R_{lg} versus current density of the 4.5 μm HBT at 122 GHz.

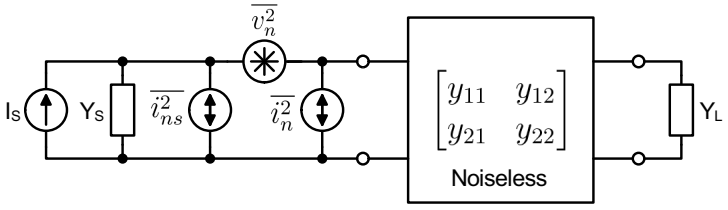


Figure 2.11: Noisy two-port representation.

amplifier that can be described as such, can be lumped into two input noise sources, $\overline{v_n^2}$ and $\overline{i_n^2}$ [60], as illustrated in figure 2.11. The $\overline{v_n^2}$ and $\overline{i_n^2}$ sources, which are generally correlated can be calculated from the internal noise sources of the two port thru simple circuit analysis or more systematically, by the noise correlation matrix method [62]. Expressions of varying degree of complexity for bipolar transistors can be found in [1, 63–65] while for FETs in [52, 66–69].

The noise figure, F, of the two-port is defined as the Signal-to-Noise ratio (SNR) at the input of the amplifier over the SNR at its output[5]:

$$F = 1 + \frac{SNR_i}{SNR_o} = 1 + \frac{\overline{(i_n + Y_s v_n)^2}}{\overline{i_{ns}^2}} \tag{2.27}$$

where $\overline{i_{ns}^2}$ is the noise current due to the real part of the source admittance Y_S. This expression indicates that F depends on Y_S and its value can be optimized even if the designer has no access to the internal components of the two-port and hence cannot alter the value of $\overline{v_n^2}$ and $\overline{i_n^2}$. This dependence is emphasized by the following equivalent expression for F [60, 70]:

$$F = F_{\min} + \frac{R_{sn}}{G_S}|Y_S - Y_{S,on}|^2 = F_{\min} + \frac{R_{sn}}{G_S}\left[(G_S - G_{S,on})^2 + (B_L - B_{S,on})^2\right] \tag{2.28}$$

where [62]:

$$Y_{S,on} = G_{S,on} + jB_{S,on} = \sqrt{\frac{\overline{i_n^2}}{\overline{v_n^2}} - \left[\operatorname{Im}\left(\frac{\overline{v_n i_n^*}}{\overline{v_n^2}}\right)\right]^2} + j\operatorname{Im}\left(\frac{\overline{v_n i_n^*}}{\overline{v_n^2}}\right) \tag{2.29}$$

is the optimum source admittance that yields the minimum noise figure [62]:

$$F_{\min} = 1 + \frac{\overline{v_n i_n^*} + \overline{v_n^2} Y_{S,on}^*}{2k_B T} \tag{2.30}$$

[5]The paper by H. Friis [59] defines the noise figure as the ratio of the *available* SNR at the input over the *available* SNR at the output, i.e. the SNR under conjugate matching conditions. More recent textbooks use the SNR under arbitrary source and load impedances. Due to the division in the fraction, the two definitions yield identical results.

The resistance R_{sn}, than determines the sensitivity of F to deviations from the optimum source admittance for noise $Y_{S,on}$, is calculated by [62]:

$$R_{sn} = \frac{\overline{v_n^2}}{4k_B T} \tag{2.31}$$

where k_B is the Boltzmann's constant, T is the temperature and $\overline{v_n i_n^*}$ represents the correlation between the two two-port noise sources. The noise sources have been assumed to be represented by double sided spectra (hence the division by $4k_B T$).

By comparing formula (2.29) for the minimum noise source admittance $Y_{S,on}$ with (2.9) for the optimum source admittance $Y_{S,og}$ for maximum power gain, it can be concluded that the two admittances generally assume different values and depend on entirely different parameters. As a result, it is unlikely for an amplifier to simultaneously exhibit the minimum possible noise figure and the maximum possible gain without some ad hoc design intervention. Furthermore, similarly to $G_{\max}$, the transistor $F_{\min}$ will depend on its current density and the one that yields the minimum $F_{\min}$, is lower [64] than the one that results in the largest $G_{\max}$.

When a transistor has a large $G_{\max}$ at the required frequency of operation and the minimum possible noise is required, a simple "brute-force" amplifier design approach would proceed as follows. The transistors are biased at their minimum-$F_{\min}$ current density and the input admittance is set equal to $Y_{S,on}$ [71]. The amplifier output is then conjugately matched to maximize its power gain, which can be calculated by the available gain formula (2.13). If the resulting available gain is not sufficient for the target application, a compromise can be made by selecting Y_S between the optimum admittance for gain $Y_{S,og}$ and the optimum for noise $Y_{S,on}$ or by adjusting the transistor current density.

Another well-known method of compromising between maximum gain and minimum noise figure is by the use of external feedback in the amplifier. Its proper use can manipulate $Y_{S,og}$ while leaving $Y_{S,on}$ almost intact [64, 72]. Consequently, one can use feedback in order to make $Y_{S,og}$ and $Y_{S,on}$ equal and then match directly for maximum power gain at both the input and output of the amplifier. This method has the important advantage of leading to an amplifier that is conjugately matched at its input (i.e. its S_{11} is theoretically $-\infty$) compared to the "brute-force" method. However, the feedback element used to alter $Y_{S,og}$, usually a degeneration inductor, will also result in decreased $G_{\max}$, leading to a similar noise figure - power gain trade-off.

The two above methods suffer from the same drawback. Specifically, they tend to sacrifice power gain in order to improve the noise figure of the amplifier. This is a perfectly valid approach at lower frequencies where the transistors typically exhibit very high power gain.

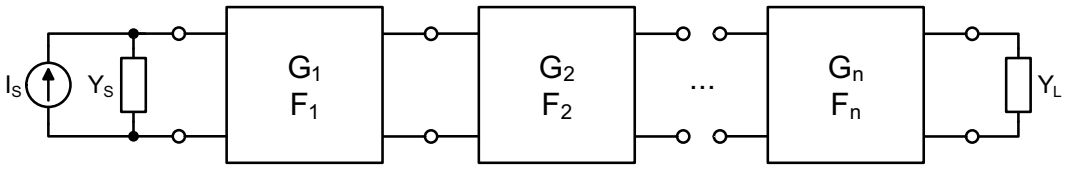


Figure 2.12: Noise figure of a two-port cascade.

However, at upper mm-wave frequencies, especially above 100 GHz, where the transistor $G_{\max}$ is very small, a different design methodology is required.

The reason why the reduced power gain matters is illustrated is figure 2.12, where n two-ports with available power gains $G_{\text{av,1}}, G_{\text{av,2}}, \ldots, G_{\text{av,n}}$ and noise figures $F_1, F_2, \ldots, F_n$ have been cascaded. These two-ports can represent the stages of a multi-stage amplifier as well as the mixer that will succeed them in a radio receiver. The noise figure of the cascade is calculated by Friis' noise formula [59]:

$$F = F_1 + \frac{F_2 - 1}{G_{\text{av,1}}} + \frac{F_3 - 1}{G_{\text{av,1}} G_{\text{av,2}}} + \cdots + \frac{F_n - 1}{G_{\text{av,1}} G_{\text{av,2}} \cdots G_{\text{av,n}-1}} \tag{2.32}$$

where the available gains $G_{\text{av,1}}, G_{\text{av,2}}, \ldots, G_{\text{av,n}}$ and noise figures $F_1, F_2, \ldots, F_n$ are calculated with respect to the output admittances of the preceding stages, i.e. $Y_S, Y_{\text{out,1}}, Y_{\text{out,2}}, \ldots, Y_{\text{out,n}-1}$ respectively [73].

Therefore, if the first stage of the amplifier has inadequate power gain, the noise figure of the following stage will add to the overall noise figure by $\frac{F_2-1}{G_{\text{av,1}}}$, i.e. by its noise figure divided by the power gain of the first stage. As a result, if the designer reduces the power gain of the first stage excessively in favor of its noise figure, the overall F might actually *degrade* due to the contribution of the subsequent stages.

To overcome this shortcoming of the traditional low-noise design methodologies, a new Figure-of-Merit needs to be considered that will account the aforementioned cascading effect. This role is fulfilled by the noise measure M [70, 74–76], which is defined as the noise figure of an infinite cascade of the same two-port [70]:

$$F_\infty = 1 + M = 1 + \left[\frac{F - 1}{1 - 1/G_{\text{av}}} \right] \tag{2.33}$$

where the assumption that all the two-ports are presented the same source admittance Y_S is made. The noise measure accounts for both the noise figure and power gain. Designing for minimum M will properly balance between the two, making sure that the gain of the stage in not excessively degraded, preventing succeeding stages from dominating the noise [76].

Since both G_{av} and F depend on Y_S, the noise measure can be cast into an equivalent

expression by employing equations (2.15) for G_{av} and (2.28) for F, as was first described by H. Fukui [42]:

$$M = \frac{F-1}{1-1/G_{\text{av}}} = \frac{F_{\min} - 1 + \dfrac{R_{sn}}{G_S}\left[(G_S - G_{S,on})^2 + (B_L - B_{S,on})^2\right]}{1 - \dfrac{1}{G_{\max}} + \dfrac{R_{sg}}{G_S}\left[(G_S - G_{S,og})^2 + (B_S - B_{S,og})^2\right]} \tag{2.34}$$

The right side of above equation has a unique minimum [42, 74]:

$$M_{\min} = \frac{M_2 + \sqrt{M_2^2 - M_1 M_3}}{M_1} \tag{2.35}$$

where

$$\begin{aligned}
M_1 &= \Big(1 - \frac{1}{G_{\max}}\Big)^2 + 4\Big(1 - \frac{1}{G_{\max}}\Big)R_{sg}G_{S,og} \\
M_2 &= \Big(1 - \frac{1}{G_{\max}} + 2R_{sg}G_{S,og}\Big)(F_{\min} - 1 - 2R_{sn}G_{S,on}) + 2R_{sg}R_{sn}(|Y_{S,og}|^2 + |Y_{S,on}|^2 - 2B_{S,og}B_{S,on}) \\
M_3 &= (F_{\min} - 1)^2 - 4(F_{\min} - 1)R_{sn}G_{S,on}
\end{aligned} \tag{2.36}$$

The minimum noise measure occurs at the corresponding optimum source admittance $Y_{S,om} = G_{S,om} + jB_{S,om}$, where:

$$\begin{aligned}
G_{S,om} &= \frac{M_{\min}\Big(2R_{sg}G_{S,og} - \dfrac{1}{G_{\max}} + 1\Big) + 2R_{sn}G_{S,on} - F_{\min} + 1}{2(M_{\min}R_{sg} + R_{sn})} \\
B_{S,om} &= \frac{M_{\min}R_{sg}B_{S,og} + R_{sn}B_{S,on}}{M_{\min}R_{sg} + R_{sn}}
\end{aligned} \tag{2.37}$$

The noise measure can then be rewritten in the minimum plus mismatch form, as was the case with G_{av} and F [74]:

$$M = M_{\min} + \frac{R_{sm}}{G_S}|Y_S - Y_{S,om}|^2 = M_{\min} + \frac{R_{sm}}{G_S}\left[(G_S - G_{S,om})^2 + (B_L - B_{S,om})^2\right] \tag{2.38}$$

where

$$R_{sm} = R_{sn} + M_{\min}R_{sg} \tag{2.39}$$

A similar expression is valid for the noise figure of the infinite cascade:

$$F_\infty = F_{\infty,\min} + \frac{R_{sm}}{G_S}|Y_S - Y_{S,om}|^2 = F_{\infty,\min} + \frac{R_{sm}}{G_S}\left[(G_S - G_{S,om})^2 + (B_L - B_{S,om})^2\right] \tag{2.40}$$

where $F_{\infty,\min} = 1 + M_{\min}$.

The methodology to design an amplifier for optimum noise measure in straightforward. First, the current density that minimizes $M_{\min}$ at the required frequency of operation is found and the corresponding bias current is set through the transistor. Subsequently, there are two options regarding the source admittance. If conjugate matching at the input is not necessary for the particular application, the input admittance can be set directly to $Y_{S,om}$. If conjugate matching is required, the same feedback techniques describes in [64, 72] can be employed to increase $Y_{S,og}$ and attempt to make it equal to $Y_{S,om}$. When the transistor power gain is high, i.e. $G_{\text{av}} \to \infty$, then $M_{\min} + 1 = F_{\min}$ and designing for minimum noise measure reduces to designing for minimum noise figure.

Fig. 2.13 illustrates the simulated and measured $F_{\min}$ and $G_{\max}$ at 122 GHz versus current density for the 4.5 μm SiGe HBT. Also shown are the minimum total noise figure $F_{\infty,\min} = 1 + M_{\min}$ and the total noise figure $F_{\infty,G_{\max}} = 1 + M_{G_{\max}}$ corresponding to the source admittance for maximum gain $Y_{S,og}$. The noise factors $F_{\infty,\min}$ and $F_{\infty,G_{\max}}$ of the infinite chain are based on the measured data, obtained using the methodology described in [1]. Due to the small power gain of the device, less than 5 dB, the minimum total noise figure, $F_{\infty,\min}$ is 2 dB higher than the minimum noise figure $F_{\min}$ of a single transistor. Notably, the current density for minimum $F_{\infty,\min}$, approximately 0.7-1 mA/μm, is considerably higher than the one that yields the minimum $F_{\min}$. Furthermore, if the transistor is matched directly for maximum gain, the resulting overall noise factor $F_{\infty,G_{\max}}$ would be only 0.7 dB higher than $F_{\infty,\min}$.

Figure 2.14 illustrates $F_{\infty,\min}$ of an infinite cascade of 4.5 μm HBTs along with the $F_{\min}$ and $G_{\max}$ of a single device versus frequency. As the frequency increases and the gain drops, $F_{\infty,\min}$ rapidly increases, as it becomes dominated by the gain of the device. On the contrary, $F_{\min}$ increases only moderately by approximately 1 dB, highlighting the fact that at mm-wave frequencies $F_{\min}$ can be misleading as far as the resulting amplifier performance is going to be.

2.3. Passive Components

Sections 2.1 and 2.2 have focused on the calculation of the optimum source and load admittances than maximize the power gain and minimize the noise contribution of amplifiers. However, the circuit designer, having to work with fixed source and load admittances, needs to resort to matching networks, as was depicted in figures 2.2 and 2.3.

Matching networks almost exclusively involve the use of passive components [41], most often inductors and capacitors, but also more complicated structures such as transmission lines, transformers and coupled lines, or even varactors. Examples of some commonly en-

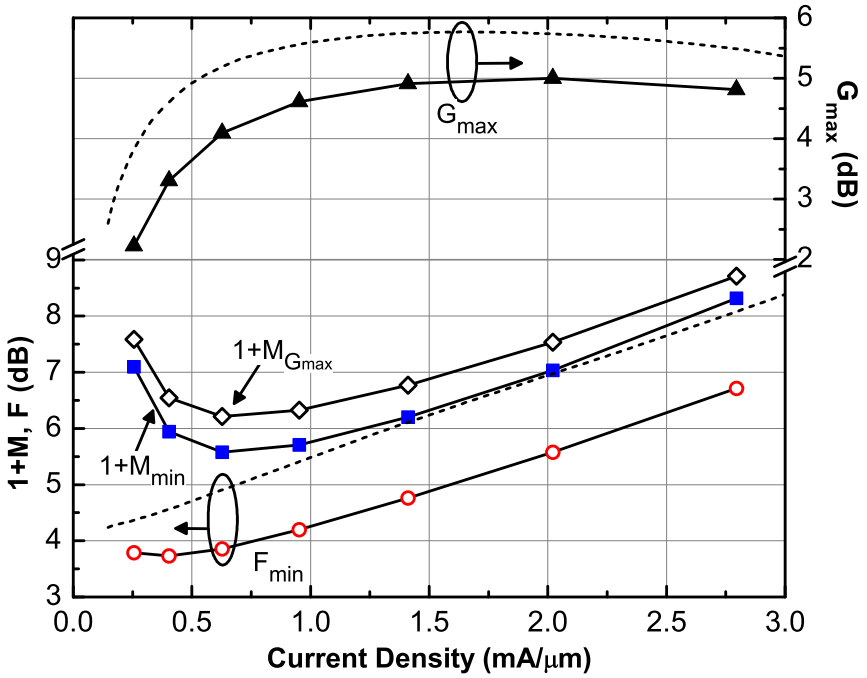


Figure 2.13: Measured (lines and symbols) and simulated (dashed lines) $G_{\max}$ and $F_{\min}$ for a 4.5 μm HBT at 122 GHz. $M_{\min}$ and $M_{G_{\max}}$ also shown, are extracted from the measured data.

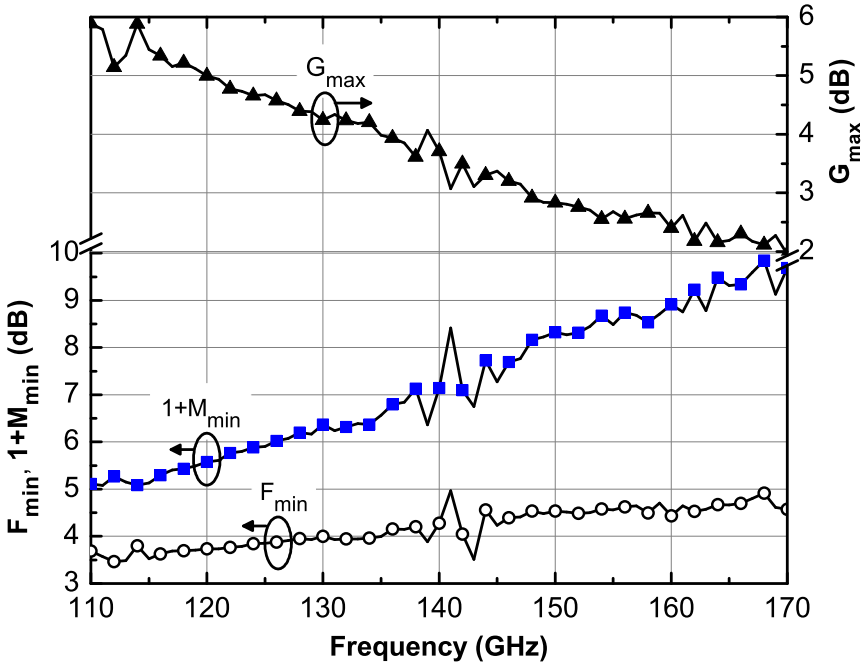
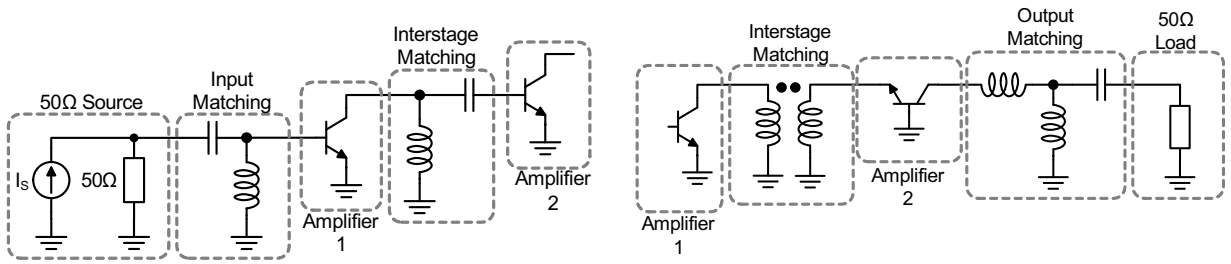


Figure 2.14: $F_{\infty,\min}$, $F_{\min}$ and $G_{\max}$ versus frequency of a 4.5 μm HBT [1].



(a) Input and interstage matching.

(b) Interstage and output matching.

Figure 2.15: Two examples of matching networks.

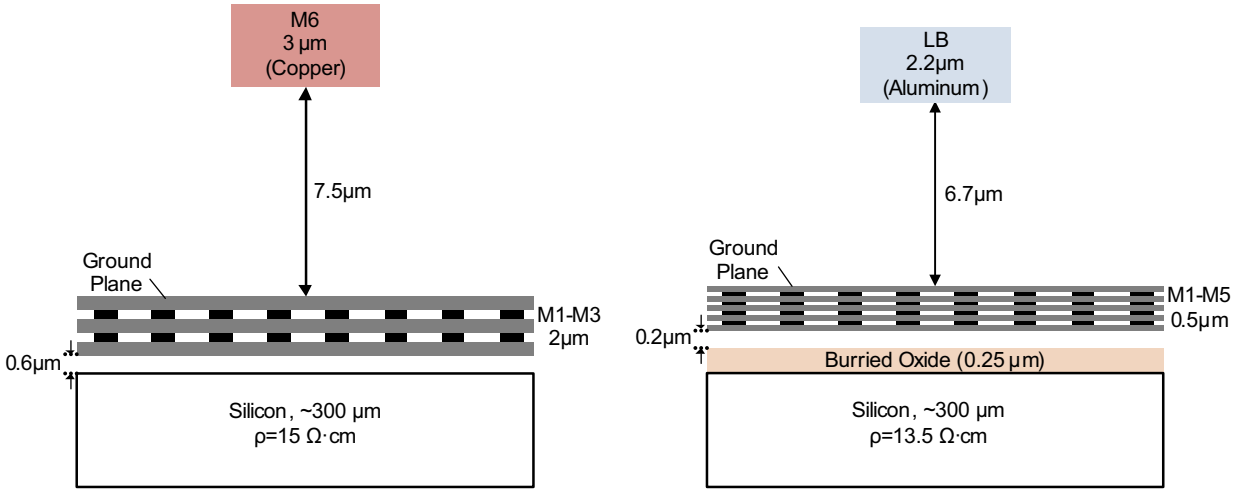
countered transformation networks are illustrated in figure 2.15. Clearly, a large number of passive components is required.

There are advantages and disadvantages in passive component design at high frequencies. Specifically, as the frequency increases, the same impedance Z can be realized with smaller inductors and capacitors. Smaller values of passives implies that their size will also shrink, rendering them much friendlier to integrated circuit implementation and reducing the overall die size. However, at increasing frequencies, the resistive loss of the metal required to create passives also increases, due to the skin effect, by a factor of $\sqrt{f}$. In addition, the loss due to the conductive silicon substrate underneath the components, especially under the spiral inductors (also known as the eddy current loss), also increases with frequency by a factor of f [77]. This unwanted loss in passives is captured in their quality factor, Q, defined as the ratio of reactive energy stored over the energy dissipated in resistive losses:

$$Q = \frac{E_S}{E_D} = \frac{X}{R} \tag{2.41}$$

The loss limitations might initially seem as a show-stopper for high-frequency and particularly mm-wave passives. However, a more careful consideration proves the contrary. First, the substrate losses might increase with frequency but they also decrease with footprint [78], which actually decreases with frequency. Second, if the frequency scales by a factor s, the same reactance $X = \omega L$ can be realized by an inductor that is s times smaller L/s, whereas the metal loss due to the skin-effect increases only by $\sqrt{s}$, thereby improving the resulting Q [79]. Third, modern semiconductor processes, even the ones aimed purely for digital circuits, have to improve their back-end of the line (BEOL) metallization, in terms of metal thickness and distance from the substrate[6]. This is the only way to reduce the $I \cdot R$ losses in the supply distributions. Examples of such improvements have been the migration from aluminum to copper as the main BEOL material, as well as the introduction of an ultra-

[6]The distance to the substrate of the top metals increases due to the fact that the total number of metals and the thickness increases.



(a) 130 nm SiGe BiCMOS process [53]. (b) 45 nm SiGe BiCMOS process [20].

Figure 2.16: Simplified cross-sections of the IC processes under consideration.

thick metal in microprocessor processes [29]. Overall, as also demonstrated in [80], at higher frequencies, the passive components will demonstrate similar, or even improved, Q provided that their size scales.

To further examine the above point, three characteristic passive components, inductors, capacitors and microstrip transmission lines, were designed and simulated using a 3D electromagnetic simulator in both the 130 nm SiGe BiCMOS and 45 nm SOI CMOS technologies. Simplified cross-sections of the BEOL of the two technologies, which will ultimately determine the performance of the passives, are shown in figure 2.16. Interestingly, the two cross-sections of two entirely different processes are very similar, the main difference being the material of the top metal. It should be noted that aluminum is almost 1.6 times more resistive than copper.

The inductors, which are the elements that dominate the loss of the matching networks due to their lower quality factors, are formed using the top metal layer. A hole is also opened on the ground plane underneath them, exposing the silicon substrate. The resulting structure resembles a co-planar waveguide inductor [81], on top of the lossy silicon substrate which can complicate the mode of field propagation [78, 80, 81]. Although the opening of the ground plane introduces the substrate loss, it is customary done in inductors because it reduces their parasitic capacitance to the ground, which would cause the inductor to self-resonate and become a capacitor [78].

Figure 2.17a shows the simulated effective L and Q[7] of a 270 pH spiral inductor designed in

[7] L was calculated from $L = \frac{\mathrm{Im}(y_{11}^{-1})}{\omega}$ while $Q = \frac{\mathrm{Im}(y_{11})}{\mathrm{Re}(y_{11})}$. These definitions are also known as the effective L and effective Q [82].

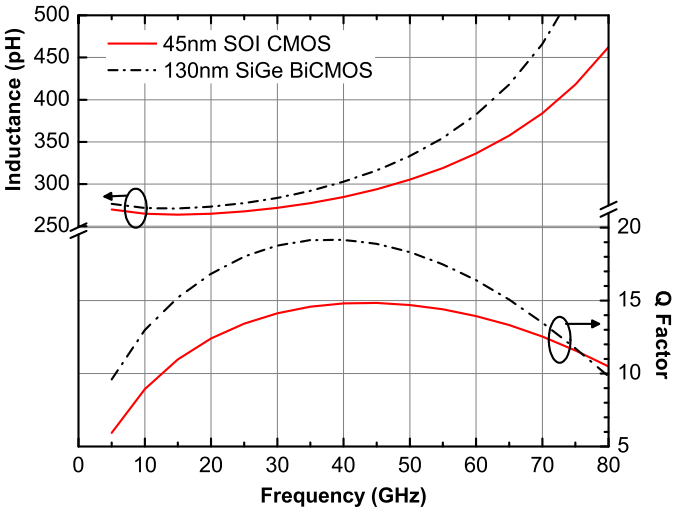
both SiGe and CMOS process and whose top-view is sketched in figure 2.17b. An inductance of this magnitude would be useful in matching networks around 40 GHz. In both cases, the effective inductance is initially relatively constant with frequency and then starts increasing. This happens due to the interaction between the inductance of the coil and the parasitic capacitance between its turns [78]. Since the top metal of the SiGe process is thicker than the CMOS (3 μm versus 2.2 μm), this effect manifest more intensely, with a sharper increase of the inductance. The peak Q of the SiGe inductor is considerably higher than the CMOS (19 versus 15), which is attributed primarily to the lower metallic losses in the thicker, copper, top metal layer, as opposed to the aluminum in CMOS. Keeping in mind that the distance to the substrate and the footprint of the spiral is almost the same in both processes, this demonstrates that the conductor and not the substrate losses dominate, which is in agreement with the analysis in [80].

Figure 2.17c shows the L and Q of the 85 pH inductor sketched in 2.17d. The peak Q of the 85 pH inductors is improved compared to the 270 pH ones, which again indicates that the conductor skin-effect losses dominate. Moreover, it becomes clear that high-Q inductors are possible at mm-wave frequencies and that their loss will limit, but will not dominate the circuit performance.

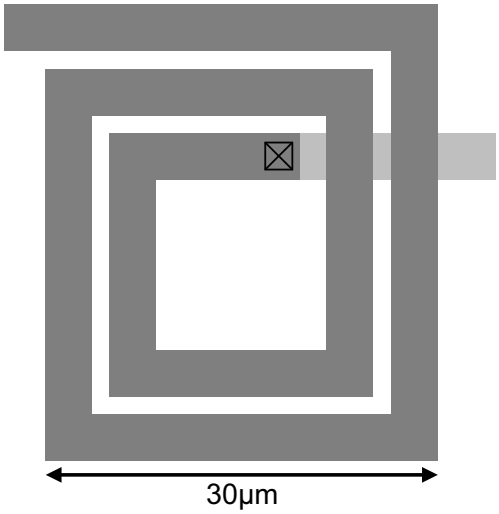
Integrated capacitors usually have higher Q than inductors, unless their value is very large. There are two types of capacitors typically found in IC processes. The Metal-Insulator-Metal (MiM) capacitors which are standard, parallel plate capacitors with a high-k dielectric between them. Most often, the MiM capacitor is not part of the standard process metallization and comes at an additional cost. The Metal-Oxide-Metal (MoM) capacitor is formed by interdigitating metal "fingers", and bringing them as close to each other as the design rules permit, taking advantage of their natural vertical and lateral capacitive coupling. The metal fingers are realized using the standard process metals, which especially in nanoscale CMOS, can be extremely dense, maximizing the capacitance per unit area.

Figure 2.18 illustrates the capacitance and Q of an 80 fF MiM capacitor in the SiGe process, as compared to a MoM capacitor of the same value in the 45 nm CMOS process. This capacitance value would be useful in matching network design in the 100-140 GHz region. In both cases, the Q of the capacitor is very high and the loss burden it will introduce in the matching network will be negligible. However, proper optimization of the structure and of the aspect ratio of the MoM capacitor led to markedly higher Q than the MiM. The reason for the reduced MiM Q is the fact that in order to bring the two plates as close as possible, their corresponding thickness has to be very small, increasing the associated resistance.

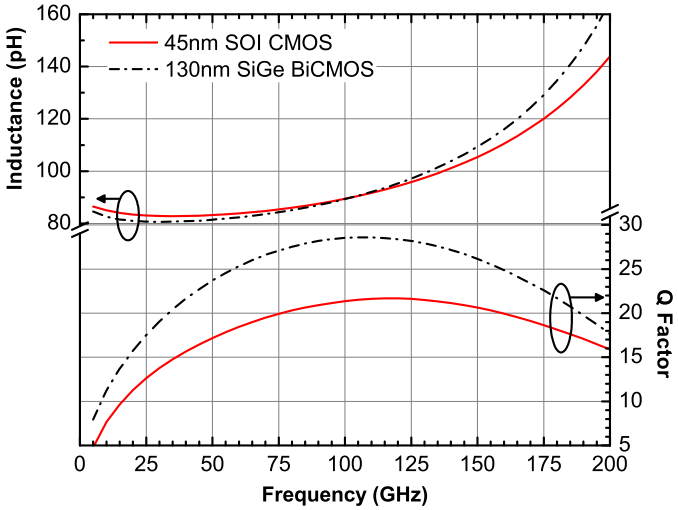
Finally, figure 2.19 illustrates the simulated parameters of a 50 Ω microstrip transmission line in both processes. A comparison between these simulations and measured data can



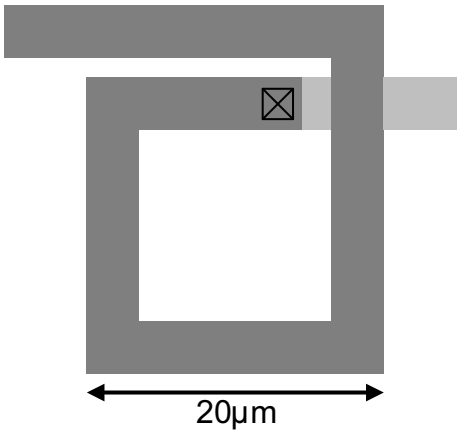
(a) 270 pH inductors L and Q.



(b) 270 pH inductor top-view.



(c) 85 pH inductor L and Q.



(d) 85 pH inductor top-view.

Figure 2.17: Simulated L and Q versus frequency of two inductors in the SiGe and CMOS processes.

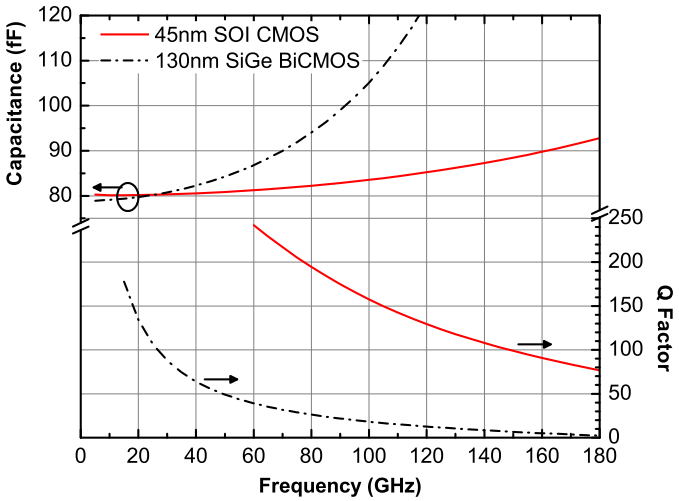
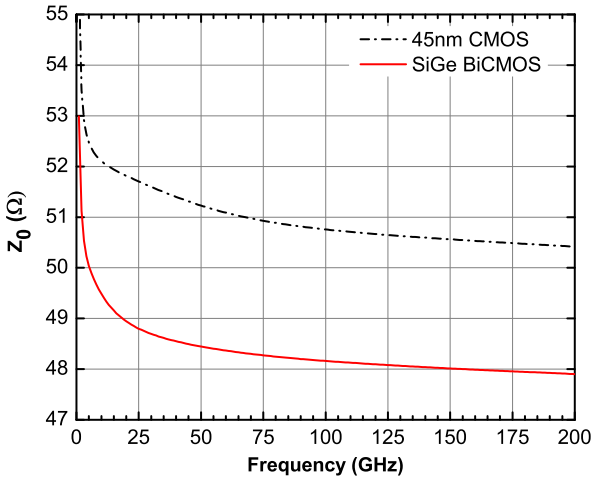


Figure 2.18: L and Q of an 80 fF capacitor in the SiGe and CMOS processes.

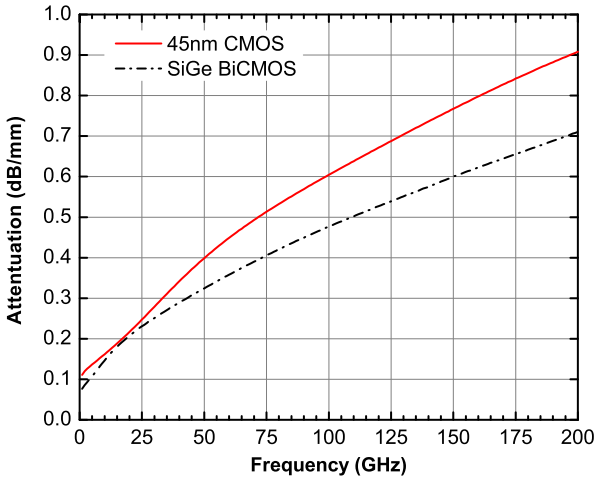
be found in [2, 83]. In both cases the top metal layer was used for signal conductor of the line whereas the ground planes were formed as indicated in figure 2.16. The characteristic impedance shows the expected quasi-TEM [41] propagation behavior with little variation at high frequencies. The almost dispersion-less quasi-TEM propagation is also verified by the linear behavior of the imaginary part of the propagation constant, β. The attenuation, which is dominated by the conductor loses, is very low in both cases, lower than 1 dB/mm even at 200 GHz. The CMOS transmission line exhibits slightly higher loss than the SiGe case for the same reason as in the inductor Q. Figure 2.19d shows the quality factor, $Q = \beta/2\alpha$, that would be attained if the lines were used as stub inductors or capacitors. The resulting quality factor is exceptionally high due to the low attenuation of the lines.

2.4. Routing of High Speed and Control Signals

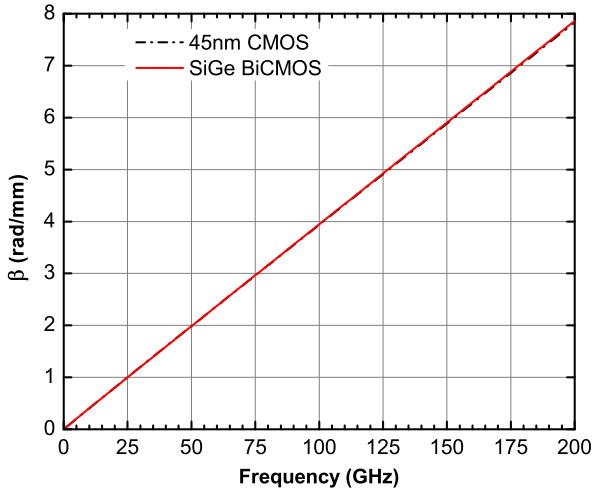
Physical layout of circuits at mm-wave frequencies presents new, significant challenges that are not encountered in such intensity at lower frequency (<5 GHz) design. Especially at frequencies above 100 GHz, a seemingly "innocent" interconnect between two circuit elements can significantly affect their performance. Even worse, as will be shown in the next section, an inconsistency in the on-chip power and ground distribution networks can even cause the entire design to fail, a problem that cannot be easily identified by the available electromagnetic simulation tools due to the complexity of these networks. The purpose of this and of the following section is to highlight the potential problems and provide design rules that can help alleviate them.



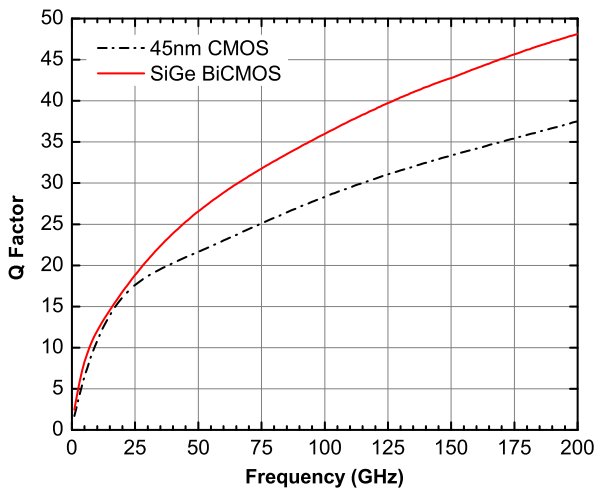
(a) Characteristic impedance Z_0.



(b) Attenuation α.



(c) Phase delay β.



(d) Quality factor $\beta/2\alpha$.

Figure 2.19: Parameters a 50 Ω microstrip line in the SiGe and CMOS processes.

2.4.1. Routing of High Speed Signals

A wire at high frequencies behaves essentially like a transmission line. It can thus transform the impedance Z_L, seen at its one end, to Z_i at the other end:

$$Z_i = Z_0 \frac{Z_L + Z_0 \tanh(\gamma l)}{Z_0 + Z_L \tanh(\gamma l)} = Z_0 \frac{Z_L + Z_0 \tanh(\alpha l + j\beta l)}{Z_0 + Z_L \tanh(\alpha l + j\beta l)} \tag{2.42}$$

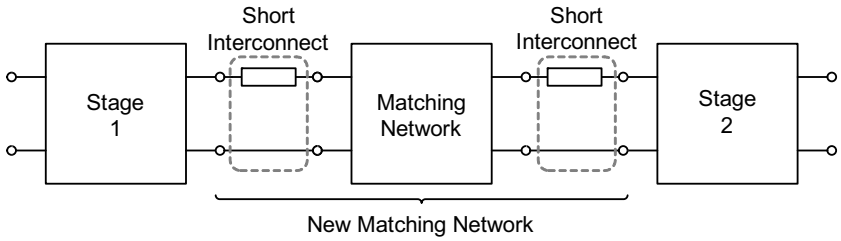
where l is the wire length, Z_0 its characteristic impedance and $\gamma = \alpha + j\beta$ is the propagation constant. Even if the line is lossless ($\alpha = 0$), it will still perform transformation of impedances. Z_0 and γ depend on the characteristics of the line, i.e. the width and the metal layer that the wire is formed at, as well as the distance of the wire from its ground return path [10]. Therefore, not only the wire itself matters, but also the ground plane that surrounds the circuits, which needs to be wide enough in order to minimize its impact.

Whenever two circuit elements need to be connected, a wire will have to be inserted in between. For example, in the matching networks of figure 2.15, wires will have to be inserted between the transistors and the inductors and capacitors of the matching networks. Since the very role of the matching networks is to transform impedances, any unwanted further impedance transformation that will be introduced by the interconnects will have an adverse impact. In extreme cases where the designer is working with conditionally stable devices, the impedance shifts can push the transistor into its unstable region, causing the amplifier to oscillate.

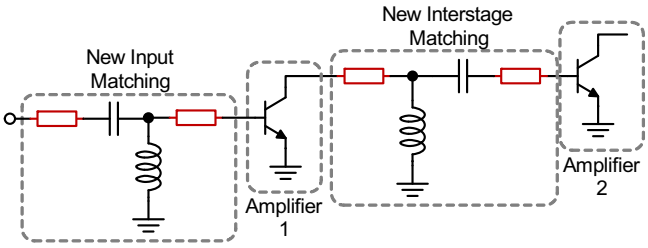
It thus becomes clear that systematic methods to handle the interconnects are required. In this work, two categories of interconnects are identified and corresponding methods are proposed to address them. The first category consists of cases where the circuit blocks can be as close to each-other as possible, e.g. between components of the same amplifier. The second category has to do with connections between blocks that have to be spaced further apart due to layout and isolation requirements. During the design of large systems, the decision to which of the two categories each block falls into can be made early on in the design phase and appropriate action can be taken to minimize design iterations.

Consider the first case where short interconnects are possible. As illustrated in figure 2.20a, the wires that connect the amplifier to the matching network can be considered part of the matching network itself. Since these wires are electrically short ($l << \lambda/2$), they will behave like small inductors. As a result, it will be possible to absorb them into the matching network only by adjusting its component values, i.e. without having to radically change its topology, as shown in the example of figure 2.20b.

In this process, only minimal design iteration will be necessary after the transition from the schematic design to the physical layout. This iteration will mainly involve measuring the



(a) Absorption of short interconnects in the matching network.



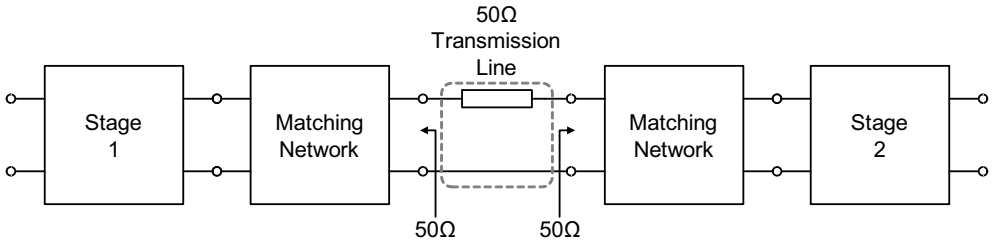
(b) Example of adjusting the matching networks to accommodate short interconnects.

Figure 2.20: Handling of short interconnects.

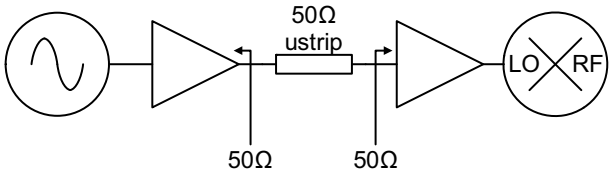
length of the wires and simulating their performance with an electromagnetic simulator (or even using scalable models). However, it should be stressed out that in order to accurately predict their behavior, some knowledge of the return path is necessary. This is easy in differential circuits because in the odd mode of operation each of the two wires at their input and output will be the return path for the other. However, in single-ended circuits a consistent, wide ground plane that is not very far from the main wire is necessary in order to minimize its impact and safely ignore it in simulation. This becomes possible with the grounding technique discussed in the next section.

The second category, that involves long interconnects between blocks placed physically apart, can be dealt with the method illustrated in figure 2.21a. Since it is very difficult to know the exact distance of the two blocks until late in the design phase, the easiest method to preemptively handle the problem is to introduce introduce fixed $50\,\Omega$ interfaces[8] at the input and output of the blocks that need to be connected. Subsequently, these two blocks can be connected with a $50\,\Omega$ transmission line of any length, without affecting their operation, other than the loss of the line which, as shown in the previous section, is quite low. On the downside, this procedure requires using two separate matching networks at the output and input of the circuits to be interfaced. Since these matching networks will introduce loss, occupy area and increase the design effort, the use of this method of interfacing should be minimized.

[8]Any other real impedance other than $50\,\Omega$ can be used for this interface, provided that a low loss transmission line of the same characteristic impedance is realizable.



(a) Interconnection of two blocks spaced physically apart using 50 Ω transmission lines.



(b) Example of using 50 Ω transmission lines to connect the oscillator to the mixer.

Figure 2.21: Handling of long interconnects.

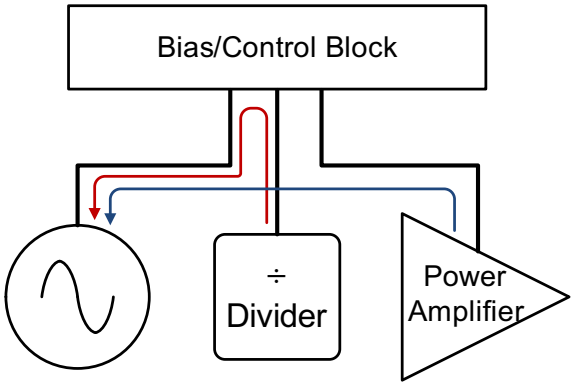


Figure 2.22: Coupling from noisy blocks through bias and control lines.

Figure 2.21b depicts a typical example where interfacing with long transmission lines is necessary. In radio transceivers, the synthesizer that generates the oscillator signal is usually placed further apart from the other circuits to increase the isolation and due to layout limitations. The routing of the oscillator signal can be easily handled if two 50-Ω buffers are included at the output of the synthesizer and the input of the mixer.

2.4.2. Routing of Bias and Control Signals

In every system, it is necessary to route bias and control signals from the control blocks that generate them over very long distances in the chip. This does not present significant challenges in low frequency design, where those lines are routed with the only consideration being to minimize the space they occupy. However, at higher frequencies where coupling between adjacent structures increases, even the DC control lines can pose new problems.

As illustrated in figure 2.22, if control lines whose destinations are different blocks are routed relatively close to each other, noise and other interfering signals can couple from "noisy" blocks, such as a frequency divider or a power amplifier to a sensitive block, like the oscillator. This deleterious effect can degrade the noise-sensitive blocks and impact the overall system performance.

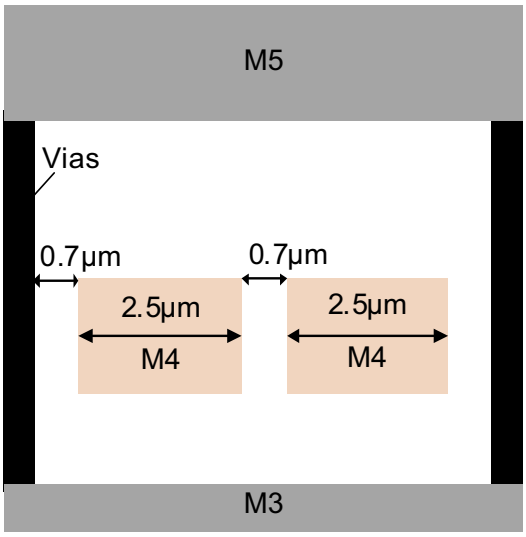
A solution to this problem would be to route the bias/control lines as far from each other as possible. However, this is not always possible due to space limitations, especially in big systems where hundreds of control lines need to be routed. Therefore, it is necessary to figure out a method to allow control lines to be routed close to each other and yet ensure that they are still adequately isolated. Figure 2.23a shows a typical example of two control lines routed with the minimum allowed distance between them. In order to isolate the lines from surrounding structures, they are enclosed in a ground Faraday "cage" formed with the metal layers above and below them along with vias. Figure 2.23b illustrates an improved method of routing the lines side-by-side but still increasing their isolation. In this case the lines are still enclosed in a cage but their distance is slightly increased in order to introduce vias between them.

Figure 2.23 illustrates the simulated isolation between the two control lines for the two cases, when their length is 100 μm. Surprisingly, when no via is present, the isolation is only 20 dB after approximately 40 GHz, even though the two lines are still surrounded by grounding structures. This affirms the problems that can be caused at high frequencies by routing control lines next to each other. On the contrary, significant isolation improvement, by almost 30 dB, was attained by introducing the vias between the lines, and without significantly increasing the occupied space.

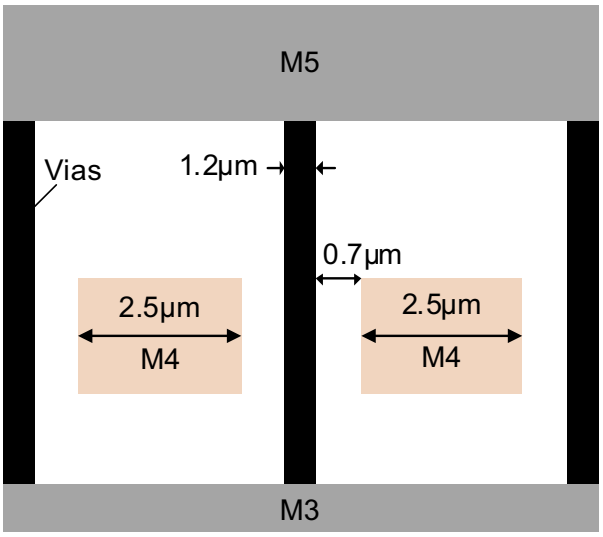
2.5. Supply Distribution and Grounding

The previous section dealt with the impact that the non-ideal behavior of wires has on high speed interconnects between blocks. Most often, these blocks contain transistor circuits that need to be connected to the power supply and the ground. During the design phase, it is common to consider both the supply and the ground to be ideal zero potential nodes at any frequency other than DC. As was the case with interconnects, this is not always a valid assumption. As the frequency increases, the supply and ground nodes start behaving quite differently due to the fact that the metals used to route them also act as transmission lines, generating finite impedances and thus finite potential at these nodes.

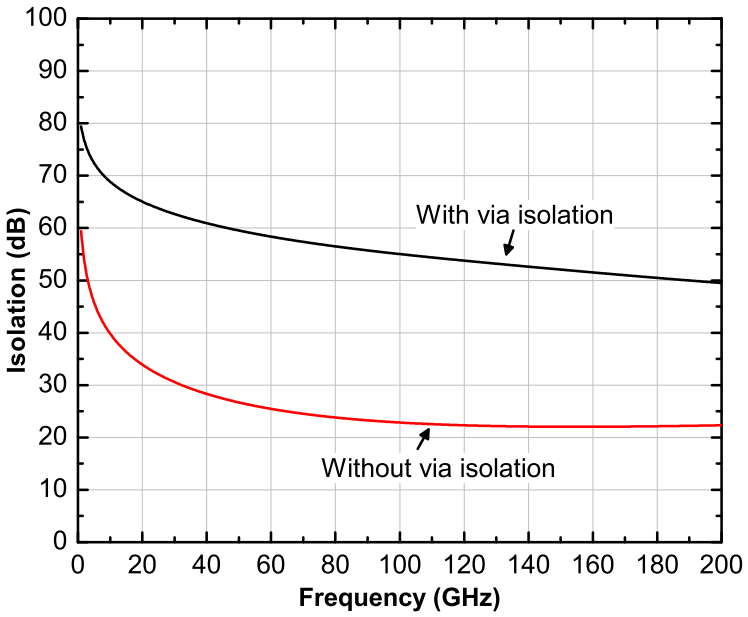
The study of this detrimental effect needs to consider single-ended and differential circuits separately, as the impact on each is quite different. Figure 2.24 illustrates a differential circuit along with a realistic model of what the supply and ground nodes consist of. The



(a) Control line routing without isolation between them.



(b) Isolated control lines with vias.



(c) Electromagnetic simulation of the isolation in the two cases when the line length is 100 μm.

Figure 2.23: Impact of bias/control signal routing.

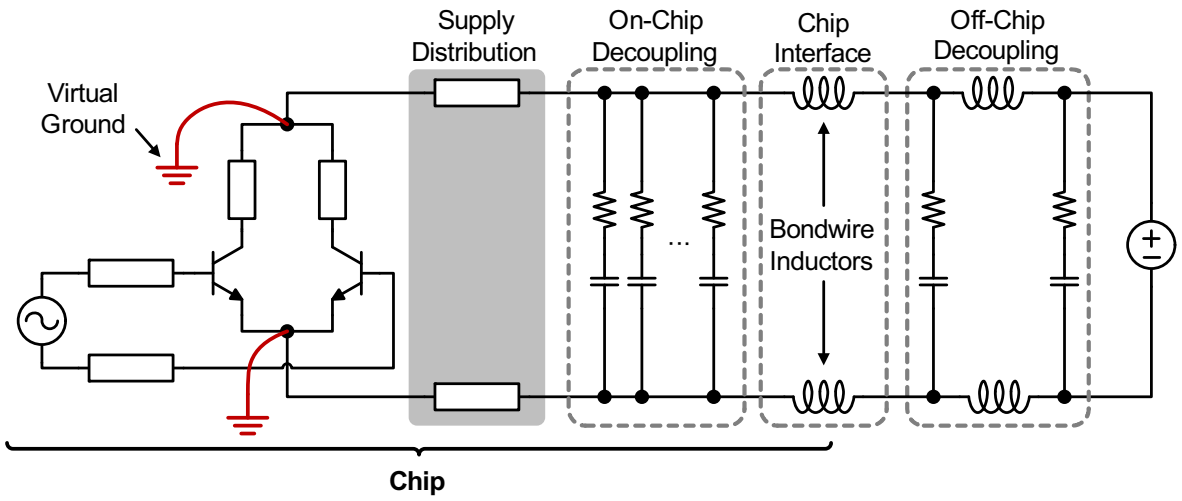


Figure 2.24: Grounding in differential circuits.

DC supply voltage is usually provided from off-chip through on-chip to off-chip interfaces, such as bondwires in simple packages or bumps in more advanced flip-chip packages. In any case, these interfaces are never optimized for high frequency performance and act as inductors whose magnitude can be up to several nH. On chip, there is a decoupling network of low Equivalent Series Resistance (ESR) capacitors, whose magnitude is significantly smaller compared to the off-chip capacitors due to size limitations. Finally, there is distribution network that routes the supply and ground nodes from the edges of the chip to the location of the circuit.

All the aforementioned components can produce a complicated frequency response in the impedance looking from the circuit towards the supply/ground network. As highlighted in figure 2.24 there is great benefit stemming from the differential nature of the circuit. Specifically, because it is operating in the odd mode, virtual grounds are formed at both the supply and ground nodes where the two sides of the differential circuit as summed together. Therefore, with only a few exceptions [84], such as common-mode stability, that are relatively easy to deal with, what happens at the supply and ground nodes is largely irrelevant to the operation of the circuit.

The situation is quite different for single-ended circuits, as shown in figure 2.25. The absence of the virtual ground due to the single-ended operation necessitates the use of a local decoupling capacitor. Ideally, this capacitor behaves like a short-circuit and provides a low-impedance return path for the AC current. However, on-chip capacitors can be far from ideal, suffering from series inductance, especially due to the vias that are required to connect to them. Moreover, the difficulty of physically connecting the capacitor from the supply back to the circuit ground can introduce a ground inconsistency in the form of series

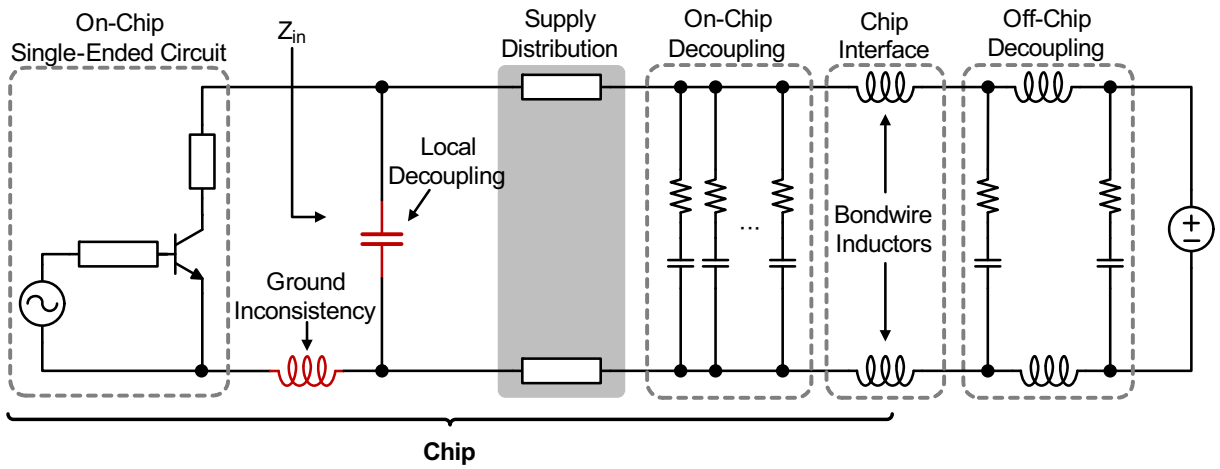


Figure 2.25: Grounding in single-ended circuits.

inductance (fig. 2.25).

To examine these issues in more detail, several MiM capacitors from the SiGe BiCMOS process have been simulated with a 3D electromagnetic simulator. Figure 2.26 illustrates the magnitude of the capacitor input impedance when their other end is short-circuited, i.e. when the capacitor is decoupling. Interestingly, each capacitors resonates at frequency that depends on its size, and is therefore suitable for decoupling only in a certain frequency region. If the decoupling capacitor is made too big, it will resonate early on and might appear as an inductor at the desired frequency, large enough to interfere with the operation of the circuit. Furthermore, as explained in [85], using multiple capacitors of various sizes will only help minimally.

Although proper selection of the decoupling capacitor size for the desired frequency range in narrow-band circuits can cause it to have a negligible impact on the circuit operation, the problem of ground consistency still remains. The physical distance between the other end of the capacitor and the circuit ground needs to be covered with the widest possible wires in order to minimize their inductance. This inductance is almost impossible to simulate using electromagnetic simulators due to the complexity of the ground structure.

Using differential circuits would solve these problems. There are cases however, where it is necessary to use single ended-circuits. Examples are the circuits that are right before and after the antennas, which are usually single ended. Therefore, the question arises whether a supply/ground distribution structure can be developed that will provide the necessary, broadband decoupling, without having to rely on the proper selection of the decoupling capacitor. In addition, the same network should provide a risk-free, low-inductance connection to the ground that will relieve the designer from having to examine every single ground connection for its consistency.

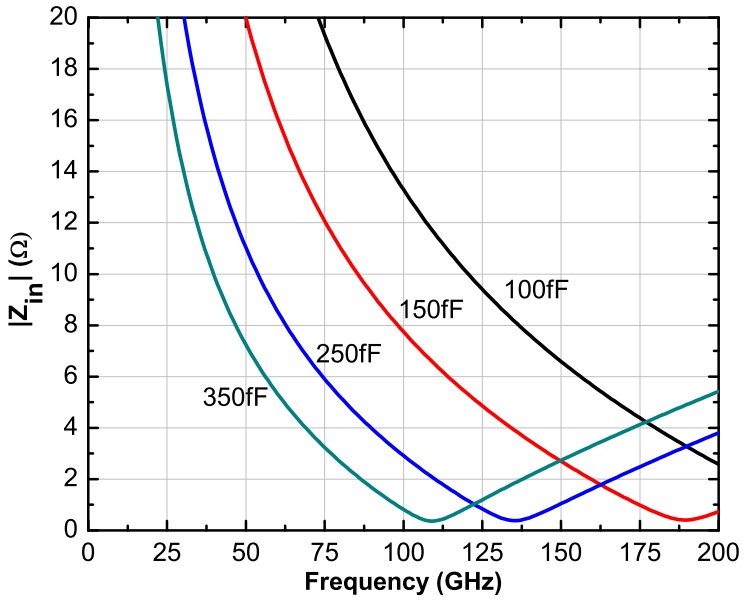


Figure 2.26: Electromagnetic simulation of $|Z_{in}| = |y_{11}^{-1}|$ of decoupling capacitors of various sizes.

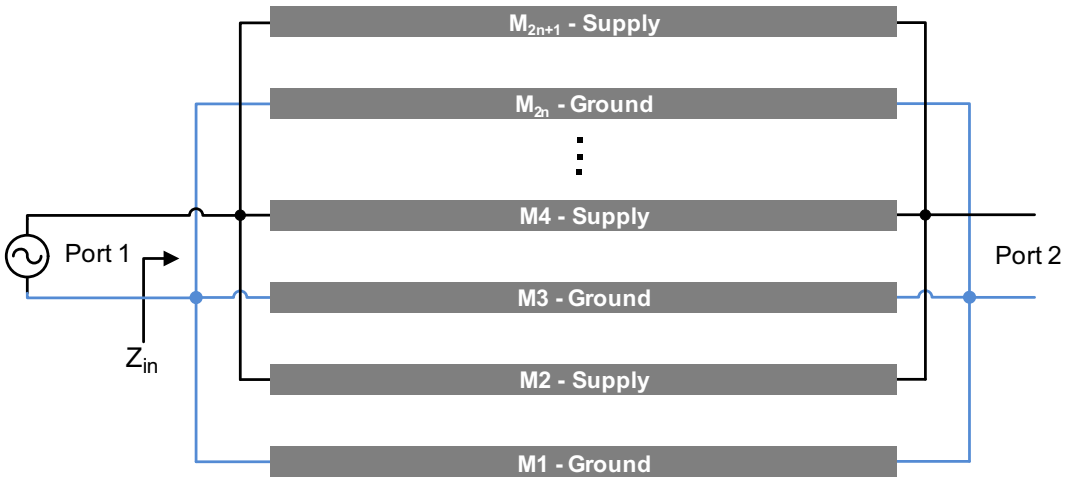


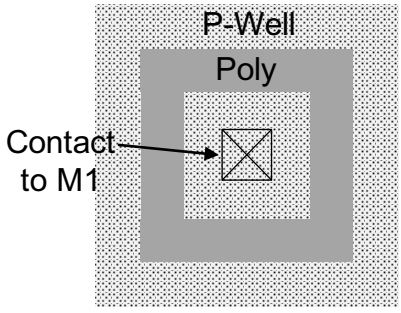
Figure 2.27: Cross-section of the supply-ground stack.

Modern semiconductor processes offer several metal layers in their BEOL, in many cases, more than ten. Moreover, because of the interconnect density required by the digital circuits, these metals can be quite dense and tightly packed. The metal layers can be used alternatively for supply and ground distribution, i.e. metal layer 1 for ground, metal 2 for supply, metal 3 for ground etc, as sketched in the cross-section of figure 2.27. This results in a "sandwich" between supply and ground planes that introduce significant decoupling capacitance between them. The "sandwich" can cover large areas of the chip that would otherwise be empty, generating a large, distributed decoupling capacitor. The large width will also result in very low inductance, which in conjunction with the large distributed decoupling capacitor will generate very low impedance, Z_{in} in figure 2.27, between the supply and ground.

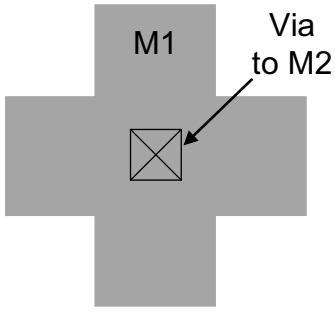
An equivalent view of this structure is as a transmission line where the signal conductor is the supply stack. The ports of such a transmission line are denoted in figure 2.27. The low inductance - high capacitance combination will generate a very low characteristic impedance $Z_0 = \sqrt{L/C}$, which according to equation (2.42), implies that the input impedance will be very low regardless of the termination at the other port.

The only obstacle for generating this structure is the metal density rules enforced by the integrated circuit foundries. These rules require from the circuit designer not only to have a metal density within a certain range, but also prohibits the creation of solid metals of very large dimensions. This problem can be surpassed by replacing the solid metals with a mesh-like structure [86,87], whose unit cell is sketched in figure 2.28. Slots of different shapes (cross vs. square) are introduced in the alternating metal layers, not only to fulfill the metal density rules, but also to allow for vias to shunt the alternating ground and supply layers with each other. Furthermore, contacts are introduced in the silicon substrate (figure 2.28a), connecting it to ground, in order to reduce the substrate coupling at high frequencies. The dimensions of the unit cell are conveniently chosen, e.g. $5\,\mu\text{m} \times 5\,\mu\text{m}$, and by repeating several cells next to each other, both horizontally and vertically, a supply distribution network of arbitrary width can be generated. The top view of a portion of this network is depicted in figure 2.29.

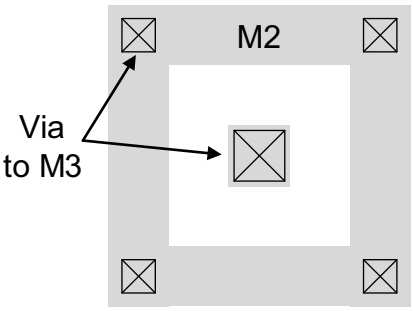
To evaluate the performance of the mesh structure, an electromagnetic simulation was performed on a $W \times L = 100\,\mu\text{m} \times 200\,\mu\text{m}$ section that utilizes five of the six metal layers of the SiGe BiCMOS process. Figure 2.30a reproduces the simulated characteristic impedance of the structure. Z_0 as low as $0.4\,\Omega$ was obtained at high frequencies, demonstrating the capability of the mesh to decouple without the use of an external capacitor in that range. However, the lack of adequate capacitance at low frequencies, below 0.5 GHz, causes a sharp increase of the characteristic impedance and shows that decoupling capacitors are still needed



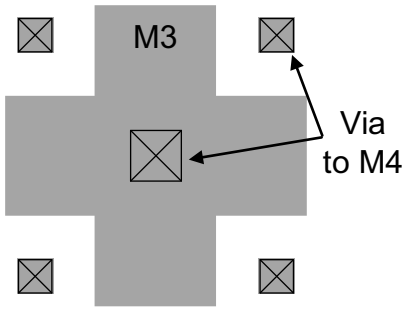
(a) Substrate to metal 1



(b) Metal 1 to metal 2.



(c) Metal 2 to metal 3



(d) Metal3 to metal 4

Figure 2.28: Details of the power - ground mesh unit cell.

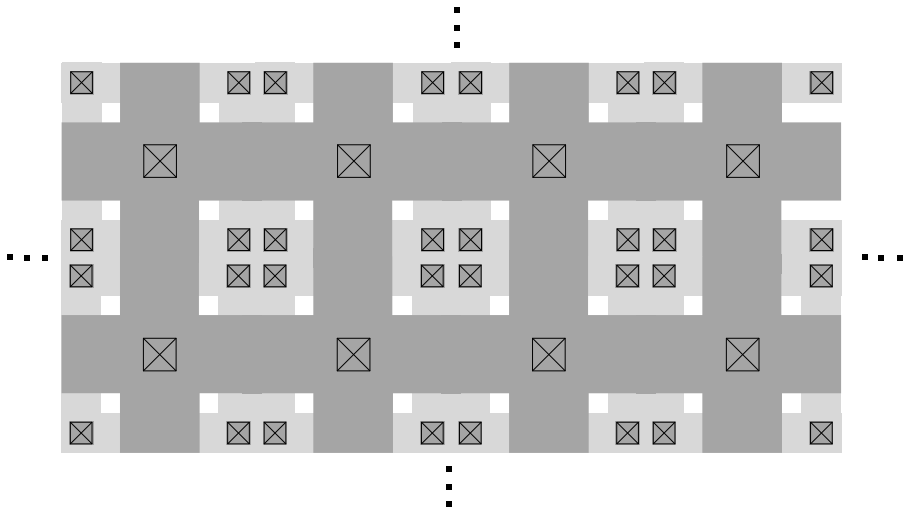
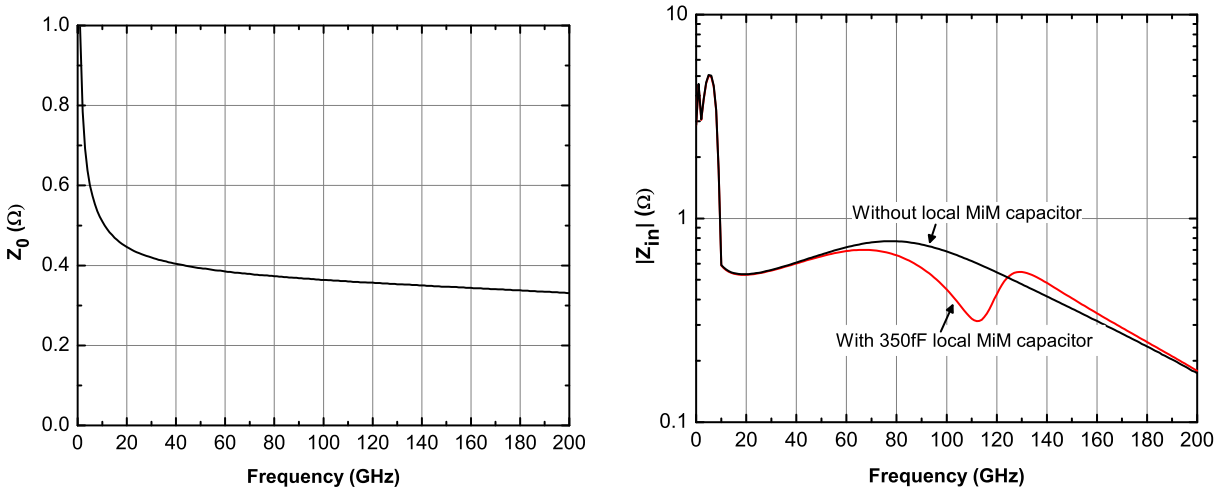


Figure 2.29: Top view of the supply-ground mesh stack.



(a) Simulated characteristic impedance of the $W \times L = 100\,\mu\text{m} \times 200\,\mu\text{m}$ section of mesh.

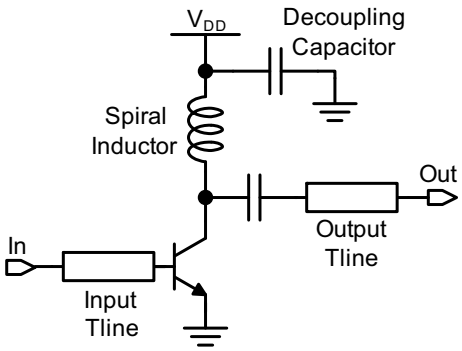
(b) Simulated input impedance of the decoupling network of figure 2.25.

Figure 2.30: Decoupling capability of the mesh structure.

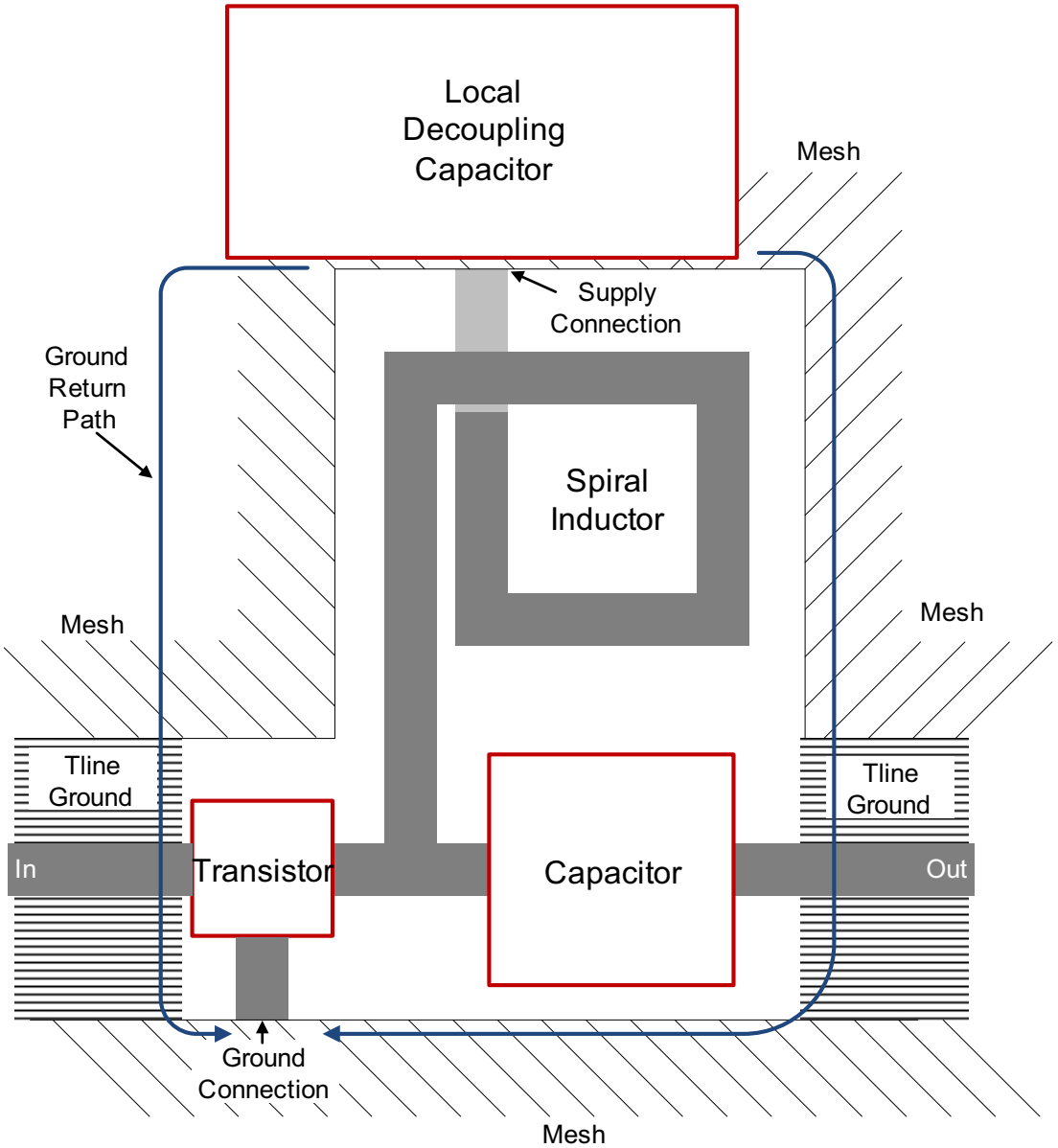
at lower frequencies.

Figure 2.30 reproduces the simulated input impedance Z_{in} of the whole decoupling network of figure 2.25 when 1 nH bondwire inductors are assumed. The impedance of the network is below 4 Ω across a very wide frequency range, from DC to 200 GHz. The very low frequency decoupling is handled by the external capacitors until they resonate, a point at which the lumped on-chip capacitors kick-in. The on-chip capacitors will also resonate along with the bondwire inductors, causing the 4 Ω resonance. At higher frequencies, the decoupling will be handled exclusively by the mesh. An impedance even lower than 4 Ω at the resonance is also possible by increasing the size on the on-chip capacitors. Also shown in figure 2.30 is the behavior of Z_{in} when a local 350 fF decoupling MiM capacitor is included. The capacitor has minimal impact and if necessary, it can even be omitted since the mesh structure provides adequate decoupling.

As an example for the above methodologies, figure 2.31b shows a physical layout sketch of the schematic of figure 2.31a. The circuit is completely surrounded by the supply/ground mesh structure, rendering the corresponding connections very easy. The possible ground return paths are also marked in the sketch, where it can be seen that they can be quite long. Therefore, if they are not made by very wide conductors, such as those found in the mesh, the inductance contribution of the return path can be significant. Even in this simple example, the layout of the return path is already complicated and it will not be exactly known until the layout of a whole block is complete, complicating (if not making impossible) its exact simulation.



(a) Schematic.



(b) Layout.

Figure 2.31: Example of physical layout of a single-ended circuit.

2.6. Summary

This chapter presented some crucial particularities of the theory behind high frequency circuit design. This theory, along with the measured and simulated transistor data of sections 2.1.3 and 2.2 and the simulated passive components of section 2.3 will be used to guide the design of circuits in the following chapters. The physical design rules of sections 2.4 and 2.5 have proven extremely valuable and are employed in every circuit block presented in this thesis.

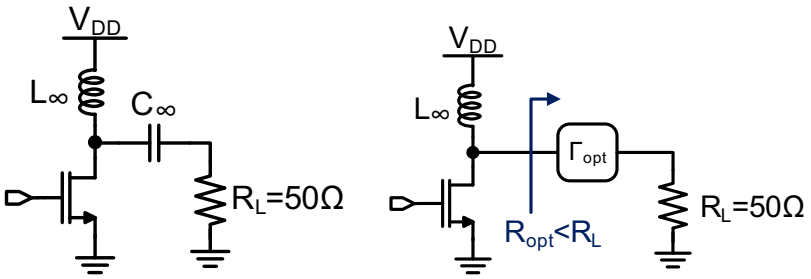
3 Design of a 45 nm SOI CMOS Stacked-Cascode Power Amplifier

The ever-increasing demand for low-cost portable communication devices pushes for higher integration of wireless transceivers in deeply-scaled silicon technologies. Given the overwhelming digital content of a mobile platform, ideally, the RF components should be realized with topologies that allow for their seamless scaling into 22-nm and 14-nm CMOS technologies. The power amplifier (PA) remains one of the most challenging circuit blocks to implement in nanoscale CMOS due to the strict requirements for output power, efficiency and linearity imposed by wireless communication standards.

As will be analytically presented in section 3.1, the low breakdown voltage of nanoscale FETs limits the maximum drain voltage swing and the maximum achievable output power. In order to circumvent this problem, a typical approach is to increase the device size and use a reactive matching network to transform the load resistance to a value significantly lower than 50 Ω. Nevertheless, due to the typically low-Q passive components that can be manufactured in nanoscale CMOS, and because of the high impedance transformation ratio involved, most of the additional output power that would be gained by increasing the device size is wasted in resistive losses in the matching networks, resulting in poor efficiency.

In this chapter, a purely digital, scalable solution for PAs that takes advantage of the significant f_T/f_{max} improvement in pFETs as a result of strain engineering will be presented. The proposed Class-D power amplifier, features a stacked, cascode CMOS inverter output stage which facilitates high voltage operation while employing only thin-oxide devices in the 45 nm SOI CMOS process.

The chapter is organized as follows. Section 3.1 reviews the basic power amplifier design theory and introduces the most important design challenges, while section 3.2 revisits the common methods employed to address them and their corresponding drawbacks. Section 3.3 introduces the proposed stacked cascode CMOS power amplifier topology and sections 3.4 and 3.5 present the design methodology necessary to improve its frequency of operation and efficiency. Finally, sections 3.7 and 3.8 present the implementation details and measurement



(a) Simple power amplifier. (b) Power amplifier with matching to R_{opt}.

Figure 3.1: Basic power amplifier design.

results, respectively.

3.1. Power Amplifier Design Theory Synopsis and Motivation

3.1.1. Output Power and R_{opt}

The design of power amplifiers generally differs from that of the small-signal amplifiers presented in chapter 2. When the maximum output power is considered, the transistor needs to be treated at least as a simple nonlinear device, whose output voltage is limited within certain ranges. This leads to design equations than differ from those of linear amplifiers presented in chapter 2.

Consider the simple amplifier of figure 3.1a where L_∞ and C_∞ are an ideal, infinitely large inductor and capacitor respectively (also known as "RF choke" inductor and "big fat" capacitor). The amplifying device is an nFET, but it could be replaced by any semiconductor device.

The maximum power is delivered to the load when the swing at the drain of the FET is maximum, since C_∞ acts as a short circuit. Due to the presence of L_∞, the theoretically maximum drain swing is $2V_{\text{DD}}$ [88, 89]. However, in practice, the semiconductor device imposes limitations due to its non-ideal behavior at the two edges of the voltage excursion.

This is better illustrated in figure 3.2 where the drain current of a 45 nm nFET is plotted versus V_{DS} for different values of V_{GS}, along with the load line imposed by R_L. The slope of the load line is set by R_L and its absolute location is defined by its midpoint M, located at $V_{\text{DS}} = V_{\text{DD}}$. On the upper side, the maximum V_{DS}, V_{max}, is limited by the breakdown region of the device that marks the onset of breakdown phenomena [51]. Apart from increasing the output conductance, these deleterious effects cause reliability problems, ultimately leading to the catastrophic, irreversible breakdown. V_{brk} generally depends of the frequency of operation

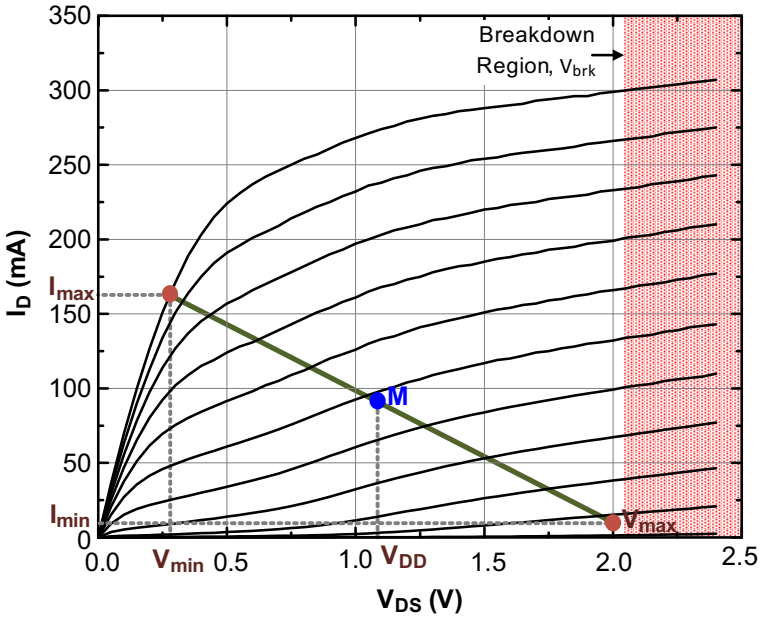


Figure 3.2: Power amplifier load-line.

and in CMOS, when operating in the GHz range, it is located at *twice* the maximum supply voltage for digital circuits.

The minimum V_{DS}, V_{min}, depends on a variety of factors that are under the control of the circuit designer, especially on the transistor width. If linear operation of the PA is required, then $V_{\mathrm{min}} = V_{\mathrm{Dsat}}$, i.e, the transistor should not enter the triode region. In non-linear PAs, where the transistors operate as switches, $V_{\mathrm{min}} = I_{\mathrm{max}} R_{\mathrm{on}}$ where R_{on} is the "on" resistance of the switch. This implies that the transistor needs to be as wide as possible is order to minimize V_{min} and maximize the output swing.

From figure 3.2 and the boundary condition imposed by the inductor, the following two equations can be derived:

$$\frac{V_{\mathrm{max}} - V_{\mathrm{min}}}{I_{\mathrm{max}} - I_{\mathrm{min}}} = -R_L \tag{3.1}$$

$$\frac{V_{\mathrm{max}} + V_{\mathrm{min}}}{2} = V_{\mathrm{DD}} \tag{3.2}$$

In the limit, $V_{\mathrm{max}} = V_{\mathrm{brk}}$ and $I_{\mathrm{min}} = 0$. Solving equations (3.1) and (3.2) for V_{DD} and I_{max} yields:

$$I_{\mathrm{max}} = \frac{V_{\mathrm{brk}} - V_{\mathrm{min}}}{R_L} \simeq \frac{V_{\mathrm{brk}}}{R_L} \tag{3.3}$$

$$V_{\mathrm{DD}} = \frac{V_{\mathrm{brk}} + V_{\mathrm{min}}}{2} \simeq \frac{V_{\mathrm{brk}}}{2} \tag{3.4}$$

where the approximation that $V_{\min} \simeq 0$ is valid especially for nonlinear PAs. From the above equations, it can be concluded that the PA supply voltage V_{DD} must be set to an appropriate value, such that when the maximum V_{DS} is limited by $V_{\min}$ to $2V_{\mathrm{DD}} - V_{\min}$, it will be sufficiently low to avoid entering into the breakdown region of the device. The maximum power at the load, known also as the saturated output power (P_{sat}), becomes:

$$P_L = P_{\mathrm{sat}} = \frac{[(V_{\mathrm{brk}} - V_{\min})/2]^2}{2R_L} = \frac{(V_{\mathrm{DD}} - V_{\min})^2}{2R_L} \simeq \frac{V_{\mathrm{DD}}^2}{2R_L} \tag{3.5}$$

where R_L is typically 50Ω.

Since V_{brk} and thus V_{DD} is set by the semiconductor technology, the only method to increase the output power beyond the limit set by the above equation is to *decrease* the load resistance. Although R_L is almost always fixed to $50\,\Omega$, load reduction can be achieved by invoking an impedance transformation network, as depicted in figure 3.1b. The new resistance that the network has to transform to can be calculated by [88]:

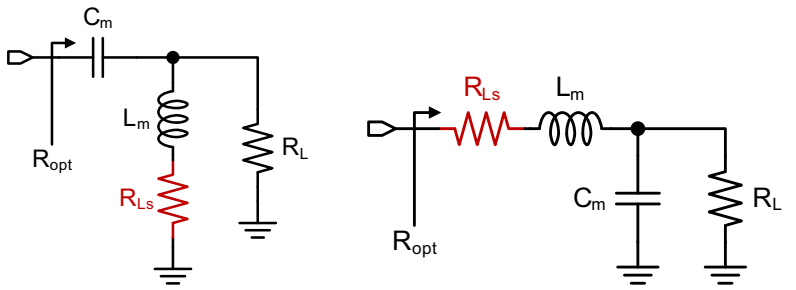
$$R_{\mathrm{opt}} = \frac{(V_{\mathrm{DD}} - V_{\min})^2}{2P_L} \simeq \frac{V_{\mathrm{DD}}^2}{2P_L} \tag{3.6}$$

After R_{opt} is selected, the current $I_{\max}$ that the transistor needs to provide can be calculated by equation (3.3):

$$I_{\max} = \frac{V_{\mathrm{brk}} - V_{\min}}{R_{\mathrm{opt}}} = \frac{2V_{\mathrm{DD}} - V_{\min}}{R_{\mathrm{opt}}} \simeq \frac{2V_{\mathrm{DD}}}{R_{\mathrm{opt}}} \tag{3.7}$$

In nanoscale CMOS, where the transistor gate length is scaled down to a few tens of nanometers, V_{brk} decreases in every technology node. For example, in the employed 45 nm SOI CMOS process, V_{DD} is set to 1.1 V, i.e. to the maximum digital supply plus 10%. If 23 dBm output power is required, R_{opt} must be only $2\,\Omega$! and the transistor needs to provide a maximum current of 1 A. Such a small R_{opt} and huge current will easily be overwhelmed by resistive parasitics, generating significant $I \cdot R$ losses that will limit the efficiency. In addition, transferring such large currents through the integrated circuit metal wires would generate electromigration reliability issues, mandating the use of very wide metals, rendering the design of passive components challenging and their area occupation significant. In deeper scaled CMOS nodes, such as 28 nm and 22 nm, R_{opt} will get even smaller for the same output power, further exacerbating the problem.

Figure 3.3 illustrates the two possible L-match networks that are able to step-down R_L to R_{opt}. The inductors suffer from relatively low quality factors, introducing R_{Ls} in figure 3.3, that will dissipate a portion of the power that the transistor sends towards the load. For the circuit of 3.3a, the efficiency, defined as the ratio of power delivered to the load, over the



(a) Series L, shunt C step-down L-match.

(b) Series C, shunt L step-down L-match.

Figure 3.3: Step-down L-match topologies.

total power provided at the input of the network is calculated by [10, 90][1]:

$$\eta = \frac{P_L}{P_{\text{in}}} = \frac{Q_L}{Q_L + Q_M} \tag{3.8}$$

where Q_L is the quality factor of the inductor and

$$Q_M = \sqrt{\frac{R_L}{R_{\text{opt}}} - 1} \tag{3.9}$$

This equation shows that both the quality factor of the inductor and the transformation ratio R_L/R_{opt} are important: the smaller the Q_L and R_{opt} are, the lower the efficiency of the matching network. Figure 3.4 shows the efficiency versus the power delivered to the load P_L for the matching networks of figure 3.3 when V_{DD} is fixed to 1 V, and impedance transformation needs to be performed to increase the output power. According to the EM simulations of chapter 2, $Q_L = 15$ is a reasonable value for inductors below 50 GHz. The efficiency suffers more at high output powers with η at only 63% when $P_L = 23$ dBm and $Q_L = 15$.

On top of the above efficiency reduction, one must also account for the finite efficiency of DC-to-RF power conversion in the nFET device, which will be the subject of the following section. It thus becomes evident that a different design approach must be pursued to increase the output power without R_{opt} becoming excessively small is necessary in nanoscale CMOS.

3.1.2. Amplifier Class of Operation

The theory presented in the previous section did not assume a certain waveform shape for the transistor voltage and current. This shape defines the class of operation of the power

[1]The analysis of the network of figure 3.3b is more involved but its efficiency can be approximately calculated by the same equations, provided that $Q_L > 10$.

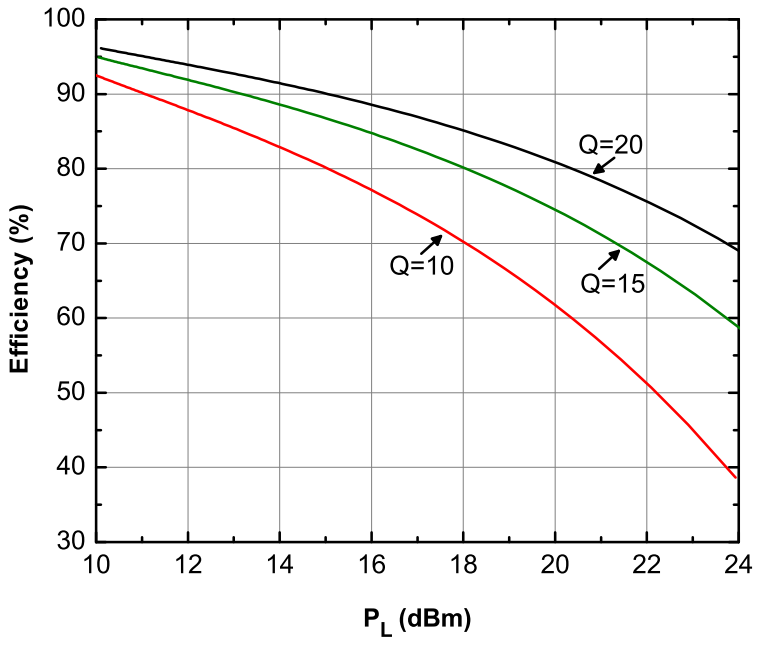


Figure 3.4: Matching network efficiency versus output power for different values of Q_L.

amplifier and depends on the input voltage as well as on how the PA is biased. The class of operation is tightly linked to the efficiency and linearity. The efficiency of a PA, or *drain efficiency*, is defined as:

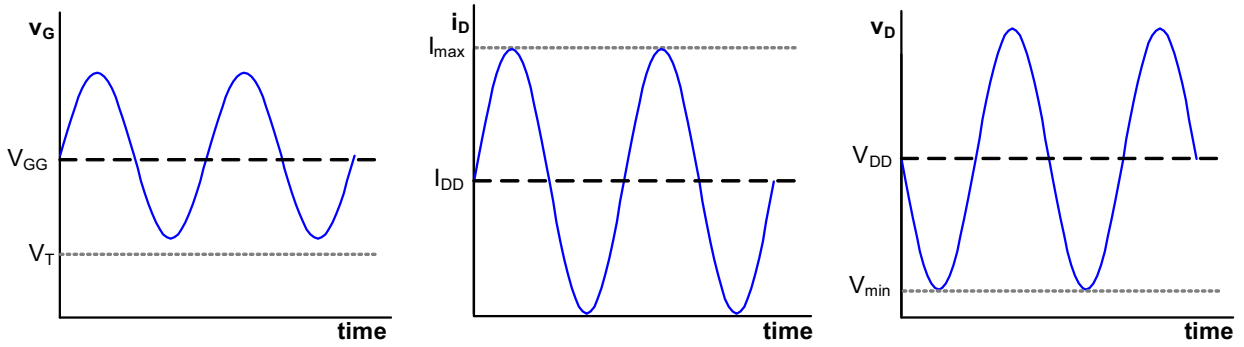
$$\eta_{\text{drain}} = \frac{P_L}{P_{\text{DC}}} \tag{3.10}$$

where P_{DC} is the power dissipated from the DC power supply. η_{drain} does not account for the input power required to drive the transistor, which can be considerable at high frequencies. This is captured by the *Power Added Efficiency* (PAE):

$$\text{PAE} = \frac{P_L - P_{in}}{P_{\text{DC}}} \tag{3.11}$$

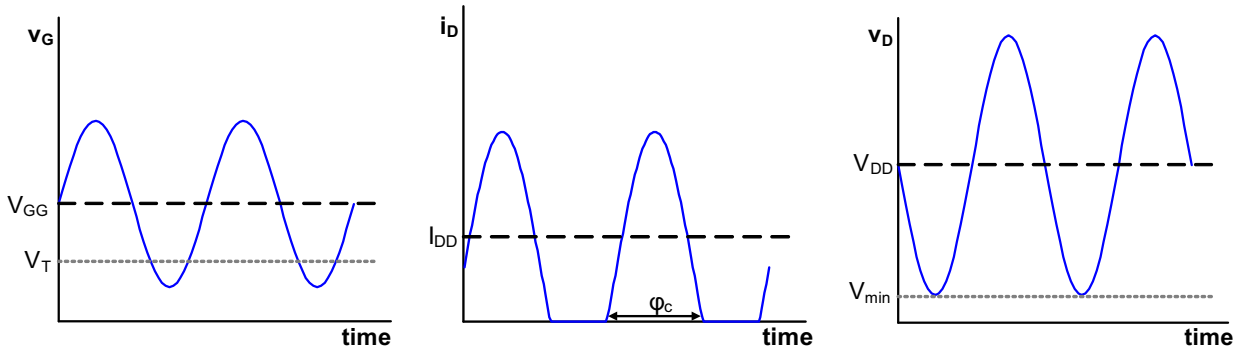
Consider the case that the amplifier of figure 3.1a is biased as linear, class-A amplifier. Its input and output waveforms are shown in figure 3.5. The gate bias V_{GG} and voltage swing are set accordingly so that v_G does not fall below the transistor threshold voltage V_T, avoiding turning off the device. As a result, the output current (fig. 3.5b) swings around I_{DD} and never goes to zero, which would cause positive clipping of the output voltage v_D (fig.3.5c). Similarly, the maximum current is set such that $V_{\text{DD}} - I_{\max} R_L > V_{\text{Dsat}}$, thus avoiding negative clipping in v_D.

Although the class-A PA amplifies the input signal with minimal distortion, it suffers from low efficiency. Specifically, since there is always current flowing through the device when there is voltage across it, it will consume power, $v_D i_D$, across the whole conduction period.



(a) Input voltage. (b) Output current. (c) Output voltage.

Figure 3.5: Class-A voltage and current waveforms.



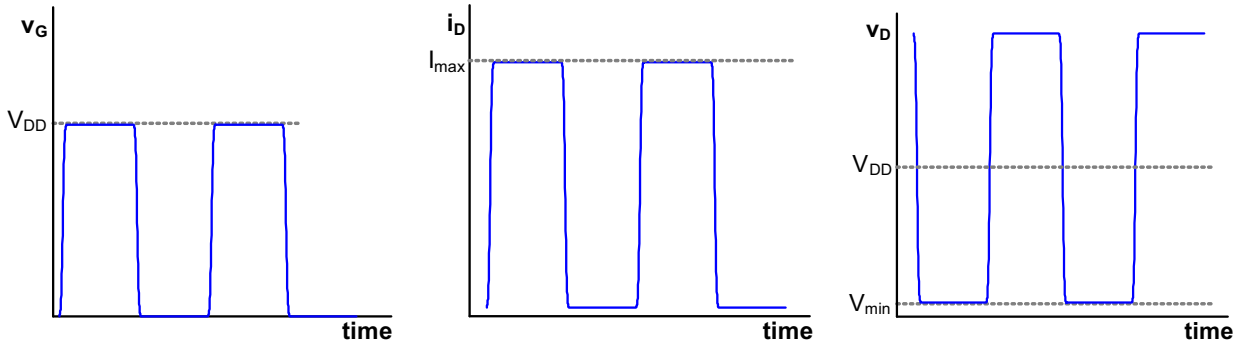
(a) Input voltage. (b) Output current. (c) Output voltage.

Figure 3.6: Class-AB voltage and current waveforms.

This limits the maximum theoretical drain efficiency of a class-A amplifier to 50% [88, 89].

The efficiency of the class-A amplifier can be improved, by trading it with linearity in the class-AB amplifier, whose waveforms are illustrated in figure 3.6. In contrast to the class-A stage, the bias voltage V_{GG} is set such that v_G falls below the threshold voltage for a certain portion of the cycle. This results in the clipped current waveform of figure 3.6b which apart from the fundamental component ω_0 of the input voltage, it also contains the harmonics at $2\omega_0, 3\omega_0, \ldots$. This results in harmonic distortion which is usually dealt with an output filter that presents a short-circuit at frequencies higher than ω_0, thus shunting the higher harmonics of the voltage. Obviously, there is no "free-lunch" in distortion, and if the input signal is modulated by a non-constant envelope, as is always the case in modern communication systems, there will still be intermodulation distortion even if the filter is present. Nevertheless, since there no current flowing through the device when v_D is maximum, the power consumption of the device is reduced and the efficiency is considerably improved compared to the class-A case.

Depending on the portion of the period that the device is on, also known as the conduction



(a) Input voltage. (b) Output current. (c) Output voltage.

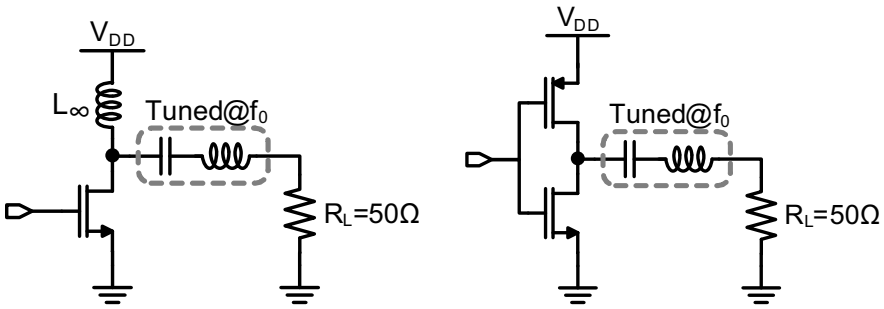
Figure 3.7: Class-D voltage and current waveforms.

angle ϕ_c, the efficiency of the class-AB PA can vary from 50%, when $\phi_c = 360°$ to 78.5% when $\phi_c = 180°$ [88, 89], with the distortion increasing along with the efficiency. $\phi_c = 360°$ is the limit when the class-AB amplifier becomes class-A whereas when $\phi_c = 180°$ is known as class-B operation.

The idea of minimizing the current-voltage overlap is taken to extreme by the class-D amplifier, whose waveforms are shown in figure 3.7 [91]. In this case, the input voltage is not a sinusoid, but a digital pulse train that swings from 0 to the maximum allowed input voltage, usually V_{DD}. This essentially forces the transistor to operate like a switch, causing voltage and current pulses at the output, as shown in figures 3.7b and 3.7c. The transistor will, ideally, have either zero voltage across its terminals, or zero current through it, pushing the theoretical maximum efficiency to 100%.

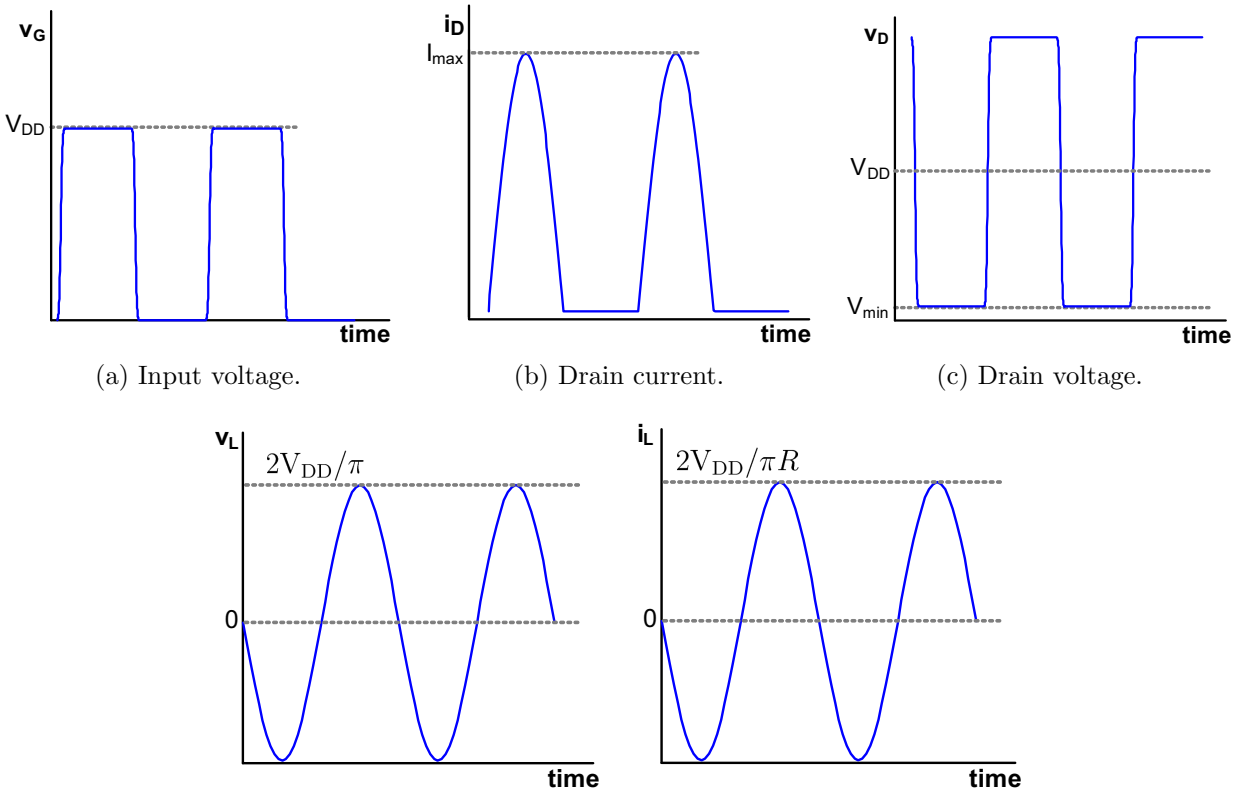
No output filter has been assumed in the waveforms of figure 3.7. Therefore, the harmonics of both the voltage and of the current are able to "see" the load R_L resulting in pulse waveforms on the load as well as the transistor drain. In applications where this is not acceptable, a series L-C filter is introduced at the output, as shown in figure 3.8, allowing only the fundamental component of the current to flow into the load. The resulting amplifier is usually referred to as the voltage-mode class-D [91] and the corresponding waveforms are illustrated in figure 3.9. Since the series L-C filter presents an open-circuit at the odd harmonics, the voltage waveform is still a pulse, thus retaining the 100% maximum theoretical efficiency. In fact, more complicated waveform shaping is also possible by presenting an open-circuit in the odd harmonics and a short-circuit in the even harmonics, a property that is exploited by the class-F amplifier [88, 89], which is the "cousin" of the class-D.

It should be noted that the above drain efficiencies of the various amplifier classes are derived based on theoretical consideration for the voltage-current overlap on an ideal device. In practice, the parasitic capacitances and the R_{on} of the transistor, which were ignored,



(a) nFET-only implementation. (b) CMOS implementation.

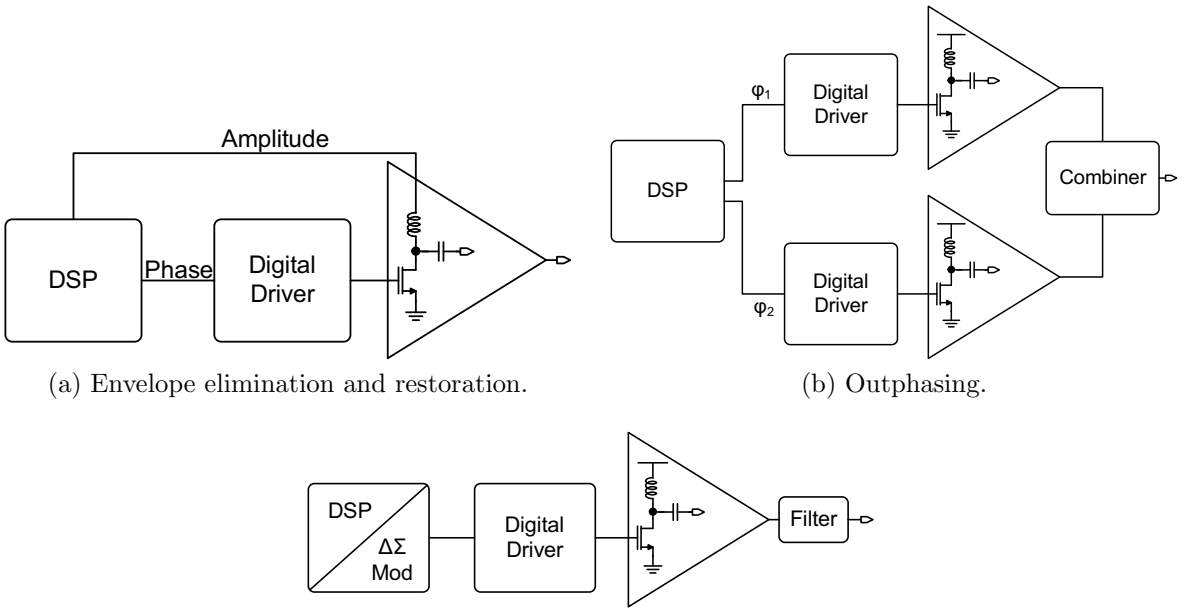
Figure 3.8: Voltage-mode class-D power amplifier with series L-C output filter.



(a) Input voltage. (b) Drain current. (c) Drain voltage.

(d) Load voltage. (e) Load current.

Figure 3.9: Voltage-mode class-D voltage and current waveforms.



(a) Envelope elimination and restoration.

(b) Outphasing.

(c) Pulse width modulation.

Figure 3.10: Linearization methods for non-linear PAs.

will limit the efficiency to lower numbers than the theoretical maximum. This problem is exacerbated at higher frequencies where the power dissipated in the capacitive parasitics increases and dominates the efficiency.

The input waveforms of figures 3.7a and 3.9a for the class-D PA indicate that its operation is highly non-linear and can only process constant-envelope input signals. In most cases, this in not acceptable as the input signal has a variable envelope, necessitating the use of linear amplification. Using non-linear components to achieve linear amplification (LINC) is an old topic that has received considerable research attention recently. Although there are several challenges in LINC systems that are still impeding their widespread adoption, they have a huge potential in overcoming the efficiency problems that plague linear PAs, especially in CMOS. This is true not only because they employ high-efficiency switch-mode PAs (like the class-D), but also because many LINC architectures can offer high *average* efficiency when processing signals with high peak-to average ratios. The drain efficiency numbers quoted for the class-A and class-AB PAs correspond to the efficiency when they deliver the maximum power, P_{sat}, to the load. When the input signal is of variable envelope, there will be points in time where lower output power is required. At lower output powers the efficiency of linear PAs is much smaller than the peak, and as a result, the average efficiency when processing a variable envelope signal will also be much lower.

Figure 3.10 illustrates three well-established methods of linear amplification with non-linear amplifiers. Figure 3.10a shows the envelope elimination and restoration (EER) trans-

mitter architecture. In this case, the baseband signal processor (DSP) decomposes the transmitter signal into polar components (amplitude and phase). Subsequently, the phase modulated sinusoidal signal at the carrier frequency is amplified by a digital preamplifier and drives the switch-mode PA input. The amplitude signal is processed separately by a high efficiency DC-DC converter and controls the power supply of the PA. When the input signal amplitude is low, the power supply of the PA is reduced, generating the corresponding amplitude at the output, and maintaining the overall efficiency of the system.

Figure 3.10b reproduces the block diagram of the outphasing transmitter architecture. The DSP generates two constant amplitude, phase modulated signals that are amplified by two non-linear PAs. When the two phase modulated signals are summed at the output, the non-constant amplitude signal is generated. Lastly, figure 3.10c shows the block diagram of a pulse-width modulated transmitter. In this case, the DSP includes a $\Delta\Sigma$-modulator that encodes the amplitude information of the signal in the duty cycle of the digital signal. After being amplified, the signal is band-pass filtered at the output and the non-constant envelope is generated.

The above architectures are especially appealing for implementation in CMOS technologies. Not only because the CMOS transistors can better operate as switches and can produce high efficiency class-D PAs, but also because the massive DSP capabilities available in CMOS allow for the implementation of superior LINC transmitters by correcting their various non-idealities. Therefore, class-D mode of operation is selected for the PA in this work. No specific LINC architecture is implemented and the primary focus will be on improving the output power and efficiency of the PA, but the proposed system will be suitable for potential use in all three LINC transmitter architectures.

3.2. Increasing the R_{opt} - Prior Art

Two methods can be devised from equation (3.6) to alleviate the problems associated with low R_{opt}. The first approach, which is currently the most widespread, is to employ power combining. The second involves the use of devices or circuit techniques that are able to increase V_{brk} and thus allow the use of higher power supply.

3.2.1. Power Combining

If the desired power level P_L leads to a prohibitively low R_{opt}, P_L can be split among N smaller PAs whose outputs are summed through a power combiner. Since the power required from each smaller PA is ideally P_L/N, the new R'_{opt} each PA has to be presented

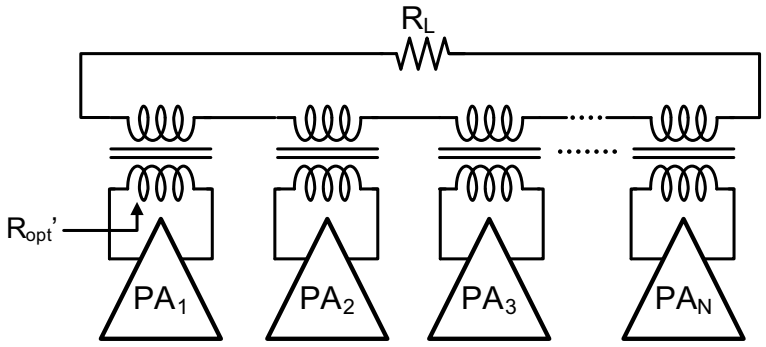


Figure 3.11: Series transformer power combining.

with becomes:

$$R'_{\text{opt}} = \frac{N(V_{\text{DD}} - V_{\text{min}})^2}{2P_L} = NR_{\text{opt}} \tag{3.12}$$

Consequently, the R_{opt} of the individual PA is N times higher.

Several methods of power combining have been derived over the last few years. A popular method involves the use of series transformers [92, 93] as illustrated in figure 3.11. Other methods involve using 90° hybrids, in the balanced amplifier [41], $\lambda/4$ transmission lines [94], or even multiple antennas and free space combining [95].

There are disadvantages associated with the power combining approach. At lower frequencies, the size of the power combiner becomes excessively large, occupying a significant amount of silicon area. This is in contrast with the design requirements of nanoscale CMOS where the price of silicon is increasing almost exponentially and the chip size needs to be minimized. Moreover, the passive power combiner will also introduce losses, which usually increase with the number of input ports, i.e. the number of PAs whose outputs are to be combined. This will reduce the benefit of combining and result only in modest improvements the the overall system efficiency.

3.2.2. Thick Oxide and Extended Drain Devices

In almost every CMOS process, the manufacturer provides devices with thicker oxides, whose primary purpose is to implement interface circuits for 2.5 V and 3.3 V digital logic. These thicker oxide devices feature much higher breakdown voltages and have been proposed for use in power amplifiers, e.g. in [96,97]. However, the thick oxide devices must have longer gates and therefore suffer from considerably lower f_T/f_{max} than the regular, thin oxide FETs in the same process and as a result, they exhibit higher R_{on} and parasitic capacitances. This would result in lower efficiency PAs due to the increased losses in these parasitics. In order to partially alleviate this problem, the output stage of the PA is usually a cascode whose common source transistor is a thin oxide device while the common-gate uses the thicker

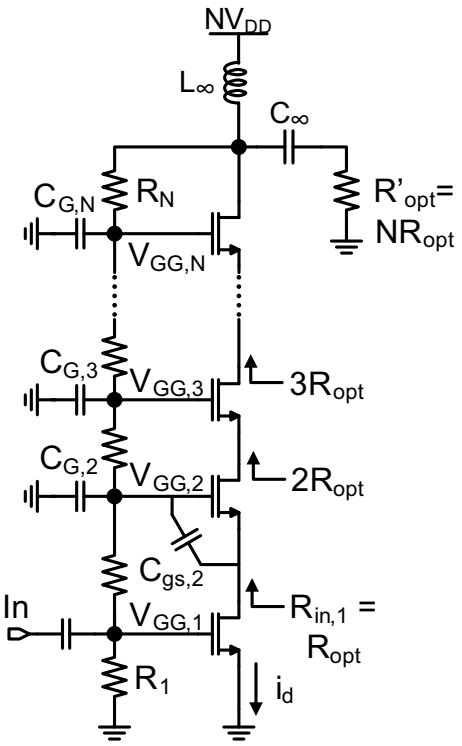


Figure 3.12: nFET stacked-cascode power amplifier.

oxide.

A similar approach is to employ various extended-drain FET (EDFET) [98–100] structures. These are usually thin oxide devices that feature higher breakdown voltages due to different drain engineering than regular FETs. Nevertheless, as with the thick oxide devices, they suffer from lower f_T/f_{max} that limits the efficiency and maximum frequency of operation of the PA. Furthermore, EDFETs are usually non-standard devices and are not offered by all CMOS foundries.

3.2.3. nFET-only Stacked-Cascode

The breakdown voltage can also be increased using only standard, thin-oxide devices along with circuit techniques. A method to do so is to employ stacking of N devices in the stacked-cascode configuration, as shown in figure 3.12. This topology was known as the "beanstalk" amplifier with bipolar transistors as early as 1964 [101], and was proposed for GaAs MESFET power amplifiers in 1992 by M. Shifrin in [102]. The N stacked standard, thin oxide FETs have a breakdown voltage of NV_{brk} and can thus be biased from $V'_{\mathrm{DD}} = NV_{\mathrm{DD}}$. The drain amplitude becomes $N(V_{\mathrm{DD}} - V_{\mathrm{min}})$, increasing the output power to NP_L and R_{opt} to $R'_{\mathrm{opt}} = NR_{\mathrm{opt}}$. If the same output power P_L as that of the single device is required from the stacked PA, then $R'_{\mathrm{opt}} = N^2 R_{\mathrm{opt}}$

There are some important design considerations that need to be taken into account when designing the stacked-cascode stage of figure 3.12. The resistors $R_1, \ldots, R_N$ of the resistive divider need to set the transistor gate voltage at:

$$V_{G,i} = V_{\text{GG,i}} + (i-1)V_{\text{DD}} \tag{3.13}$$

where $i = 1, \ldots, N$. This condition will and ensure that

$$V_{S,i} = (i-1)V_{\text{DD}} \tag{3.14}$$

$$V_{D,i} = i \cdot V_{\text{DD}} \tag{3.15}$$

$$V_{GS,i} = V_{\text{GG,i}} \tag{3.16}$$

I.e., the supply voltage NV_{DD} will be split *equally* among the N transistors of the stack, guarantying that they will not break down. The gate bias voltages $V_{GG,i}$ are set by the resistive divider, formed by resistors $R_1, R_2, \ldots R_N$ in figure 3.12, which are ideally infinitely large. These voltage determine the conduction angle of the PA when operating in class-AB mode and in the simplest case they can all be set equal $V_{GS,i} = V_{GG}$, or ever be slightly varied as proposed in [103].

A second design consideration is related to the dynamic reliability. When the output voltage swings between NV_{min} and $2NV_{\text{DD}} - NV_{\text{min}}$, it is necessary that it splits equally between the N transistors, in a similar fashion to the supply voltage. However, the input resistance $R_{\text{in,i}}$ of the common gate transistors, when looking into their source is approximately $1/g_m$, which is close to zero. As the AC current i_d travels along the stack towards the output, it gets multiplied with $R_{\text{in,i}}$ at every intermediate node and generates the voltage $v_{s,i} = R_{\text{in,i}} i_d$ which is also very small, since $R_{\text{in,i}} = 1/g_m$. Without some design intervention, this would cause all the output swing to develop at the drain of the Nth transistor, while its source voltage is fixed at $(N-1)V_{\text{DD}}$. This would result in a maximum v_{DS} of $2NV_{\text{DD}} - NV_{\text{min}} - (N-1)V_{\text{DD}} = (N-1)V_{\text{DD}} - NV_{\text{min}}$ which would definitely cause it to break down.

In general, to avoid this adverse situation and ensure equal splitting of the output swing among the N transistors, the following condition needs to be satisfied [103]:

$$R_{\text{in,i}} = \frac{i-1}{N} R'_{\text{opt}} = (i-1)R_{\text{opt}} \tag{3.17}$$

where $i = 2, \ldots, N$. I.e. the output load R'_{opt} must also be split equally among the transistors of the stack. Similar considerations hold for the gate voltage of the common-gate FETs. If the gates are decoupled to ground with large capacitors, as in traditional cascode design, the

v_{GD} voltage will become very large when the output swings up, causing the Nth transistor to break down.

Both these problems can be solved by the method adopted in [102–104]. It involves the introduction of the capacitors $C_{G,i}$ at the gates of common-gate nFETs in figure 3.12. The capacitors will increase their input resistance as follows [104]:

$$R_{\mathrm{in,i}} \simeq \frac{1}{g_{m,i}}\left(1 + \frac{C_{gs,i}}{C_{G,i}}\right) \tag{3.18}$$

where $C_{gs,i}$ is the gate-source capacitance of the ith transistor. By carefully adjusting the values of $C_{G,2}, \ldots, C_{G,N}$, the input resistances can be manipulated in order to satisfy equation (3.17) and ensure reliable drain swings in every transistor of the stack. Furthermore, the capacitive divider formed between $C_{gs,i}$ and $C_{G,i}$ will transfer a portion of the source swing to the gate, thus simultaneously reducing the maximum v_{GD} voltage and making the overall dynamic operation of the circuit reliable.

Lastly, when the output conductances of the transistors are ignored, the small signal voltage gain of the circuit can be calculated with [103]:

$$A_v = \frac{g_{m,1}R_L}{\left(1 + \frac{j\omega C_{gs,2}}{g_{m,2}}\right)\left(1 + \frac{j\omega C_{gs,3}}{g_{m,3}}\right)\cdots\left(1 + \frac{j\omega C_{gs,N}}{g_{m,N}}\right)} \tag{3.19}$$

At low frequencies, the gain is approximately $A_v = g_{m,1}R_L$ and decreases with increasing frequency. In the limiting case, when the input peak-to-peak amplitude is V_{DD}, it is necessary for the amplifier to have enough voltage gain to generate output peak-to-peak amplitude of $N(V_{\mathrm{DD}} - V_{\mathrm{min}})$. Therefore the gain must satisfy the following condition:

$$A_v > \frac{N(V_{\mathrm{DD}} - V_{\mathrm{min}})}{V_{\mathrm{DD}}} \simeq N \tag{3.20}$$

The original stacked-cascode amplifier of [102] was implemented using GaAs transistors. GaAs is a semi-insulating substrate and in contrast to silicon, it does not have drain-bulk and source-bulk junctions. If the circuit of figure 3.12 is implemented in bulk silicon, it will suffer from bulk junction breakdown since the silicon substrate is kept at 0 V and the drain voltage will be a multiple of V_{DD}. This can be partially alleviated in a triple-well process, where the deep n-well can be employed to isolate the local transistor substrate and connect it to its source. Nevertheless, there are still limits associated with the deep n-well breakdown voltage, which in nanoscale CMOS is usually 2-3 times the maximum digital V_{DD}. Furthermore, the large deep n-well to substrate junction capacitance will have to be continuously charged and discharged, degrading the PA efficiency.

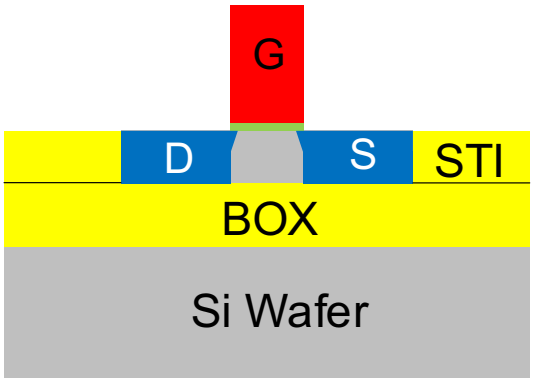


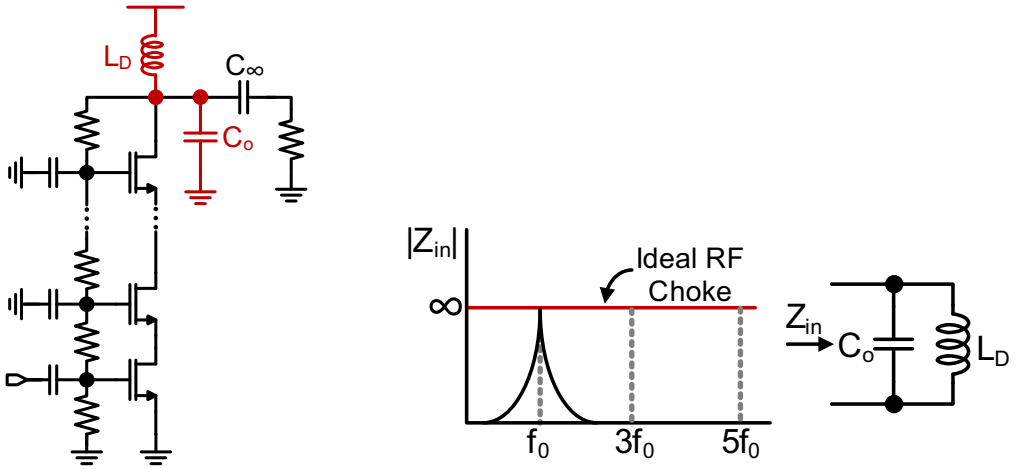
Figure 3.13: Cross-section of an SOI FET.

These problems can be circumvented in a silicon-on-insulator (SOI) CMOS process as was originally proposed in [103]. The cross section of an SOI transistor is shown in figure 3.13. The transistor is isolated from the silicon wafer by the buried oxide (BOX) and from neighboring FETs by the shallow trenches (STI). Therefore, there is no immediate junction between the source and drain regions and the substrate, alleviating the breakdown problem completely. Moreover, the BOX and STI are relatively thick oxides with a breakdown voltage higher than 15 V, thus allowing the stacking of several devices.

3.3. The CMOS Stacked Cascode Power Amplifier

The stacked nFET cascode elegantly solves the low R_{opt} problem, while avoiding the use of slower thick oxide or extended drain FETs. It has been applied with substantial success in 130 nm SOI CMOS in [103] and 45 nm SOI CMOS in [105]. However, it still requires an RF choke inductor to provide the power supply to the drain of the top transistor in the stack. As illustrated in figure 3.14a, in class-AB PAs, the RF choke can be replaced by an inductor that is designed to resonate in parallel with the output capacitance, thus operating as a resonant DC-feed. Nevertheless, if class-D operation is desired to maximize the efficiency, it is impossible to use this technique since the $L_D - C_o$ tank will resonate only at a single frequency, causing a short-circuit at the odd harmonics of the voltage, as depicted in figure 3.14b. If the odd harmonics vanish, the v_{DS} waveform turns into a sinusoid, instead of a pulse, forcing the amplifier to operate in class-AB mode. Furthermore, it is very difficult to implement an on-chip, broadband RF choke as it requires a large inductor that would appear as an AC open at the fundamental frequency. Such an inductor would self-resonate before the third harmonic, apart from occupying significant silicon area and introducing considerable additional resistive parasitics.

Class-D operation *without* an RF choke inductor can be achieved with the circuit of



(a) Schematic

(b) Input impedance of the parallel $L_D - C_o$ combination compared to the ideal RF choke.

Figure 3.14: nFET-only stacked cascode without the RF choke. L_D and C_o have to resonate at the frequency of interest f_0.

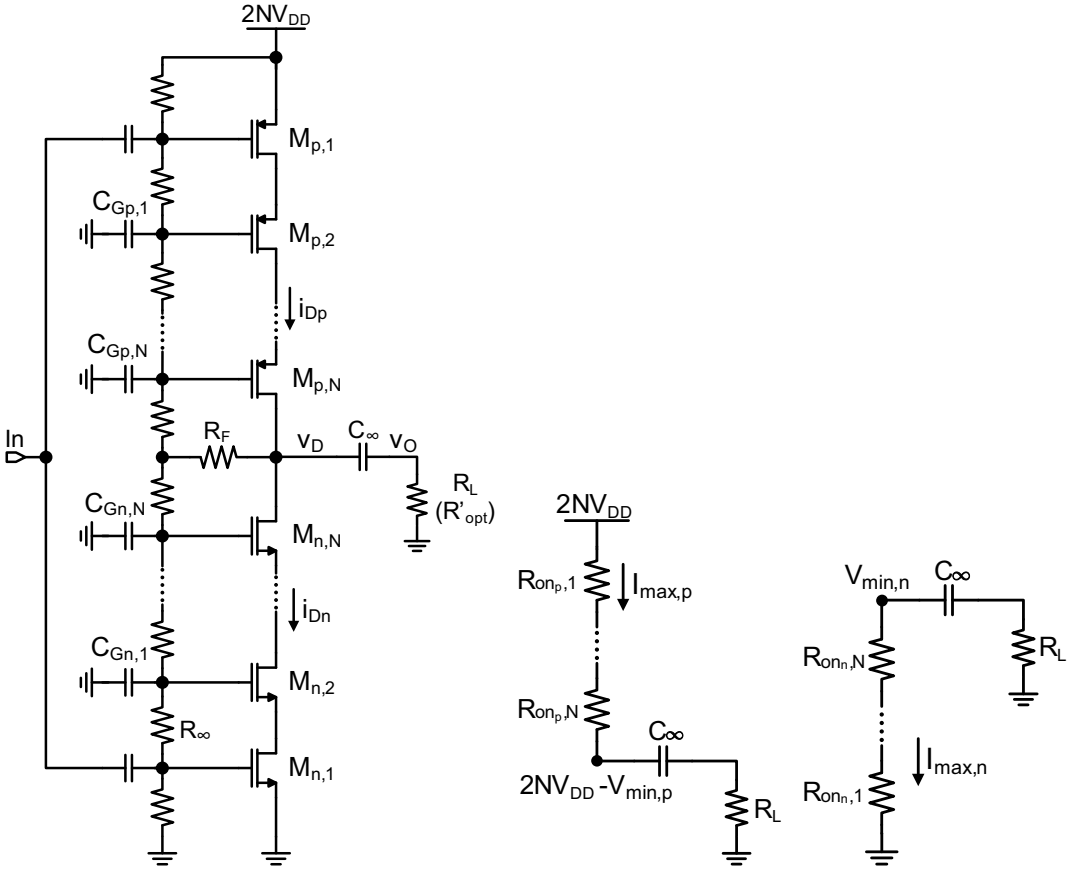
figure 3.15a. The proposed CMOS stacked-cascode PA is an extension of the simple CMOS inverter class-D PA [91, 100, 106] and replaces the choke with N pFETs, on top of the N devices of the nFET stack. As with the CMOS inverter driven by a periodic signal, during each half of the conduction cycle, either the nFET or the pFET stack is on, while the opposite stack is off, ideally behaving as an *open circuit*, thus replacing the operation of the choke.

The output waveforms of the CMOS stacked-cascode PA in class-D mode are shown in figure 3.16. During the half of the conduction cycle shown in figure 3.15b, the nFET stack is off, while the pFET stack is on, pushing current $I_{\text{max,p}}$ to the load. The common drain node voltage v_D is driven to $2NV_{\text{DD}} - V_{\text{min,p}}$ where $V_{\text{min,p}}$ is the voltage drop on the "on" resistances of the pFETs. Similarly, during the other half of the cycle, illustrated in figure 3.15c, the nFET stack in on, pulling current $I_{\text{max,n}}$ from the load, driving v_D to $V_{\text{min,n}}$. The capacitor C_∞ will remove the DC component of v_D and the output node v_O will swing symmetrically around 0, as depicted in 3.16c, with peak-to-peak amplitude $2V_o = 2NV_{\text{DD}} - V_{\text{min,p}} - V_{\text{min,n}}$.

The currents $I_{\text{max,p}}$ and $I_{\text{max,n}}$ can be calculated by considering the fact that they need to alternatively sustain the voltage V_o on the load. Therefore:

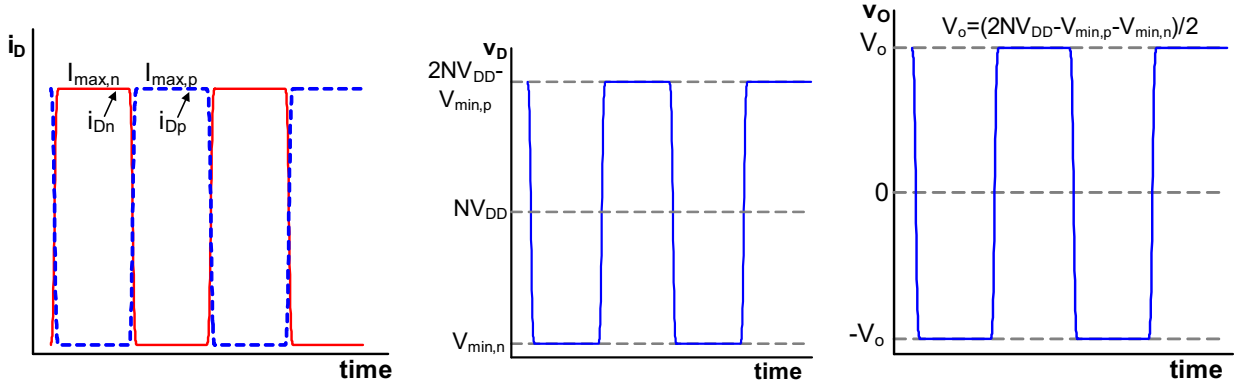
$$I_{\text{max,p}} = \frac{NV_{\text{DD}}}{R_L + R_{\text{on}_\text{p}}} \tag{3.21}$$

$$I_{\text{max,n}} = \frac{NV_{\text{DD}}}{R_L + R_{\text{on}_\text{n}}} \tag{3.22}$$



(a) Circuit schematic. (b) pFETs on. (c) nFETs on.

Figure 3.15: CMOS stacked-cascode power amplifier.



(a) Drain currents. (b) Drain voltage. (c) Output voltage.

Figure 3.16: CMOS PA voltage and current waveforms when operating in class-D.

where

$$R_{\mathrm{on_p}} = \sum_{i=1}^{N} R_{\mathrm{on_p},i} \tag{3.23}$$

$$R_{\mathrm{on_n}} = \sum_{i=1}^{N} R_{\mathrm{on_n},i} \tag{3.24}$$

are the total resistances of the pFET and nFET stacks. The minimum voltages thus become:

$$V_{\mathrm{min,p}} = I_{\mathrm{max,p}} R_{\mathrm{on_p}} = \frac{N V_{\mathrm{DD}} R_{\mathrm{on_p}}}{R_L + R_{\mathrm{on_p}}} \tag{3.25}$$

$$V_{\mathrm{min,n}} = I_{\mathrm{max,n}} R_{\mathrm{on_n}} = \frac{N V_{\mathrm{DD}} R_{\mathrm{on_n}}}{R_L + R_{\mathrm{on_n}}} \tag{3.26}$$

If no series L-C filter is used at the output, the voltage delivered to the load is a square pulse (fig. 3.16c). The RMS amplitude of a periodic pulse train centered around zero is half its peak-to-peak amplitude and, therefore, the output power can be calculated by:

$$P_{L,\mathrm{pulse}} = \frac{(2NV_{\mathrm{DD}} - V_{\mathrm{min,p}} - V_{\mathrm{min,n}})^2}{4R_L} = \frac{N^2 V_{\mathrm{DD}}^2 R_L}{4} \Big[\frac{1}{R_L + R_{\mathrm{on_n}}} + \frac{1}{R_L + R_{\mathrm{on_p}}} \Big]^2 \tag{3.27}$$

If the pFETs and nFETs are sized for equal driving strength, i.e. $R_{\mathrm{on}} = R_{\mathrm{on_n}} = R_{\mathrm{on_p}}$ then

$$P_{L,\mathrm{pulse}} = \frac{N^2 V_{\mathrm{DD}}^2 R_L}{(R_L + R_{\mathrm{on}})^2} \simeq \frac{N^2 V_{\mathrm{DD}}^2}{R_L} \tag{3.28}$$

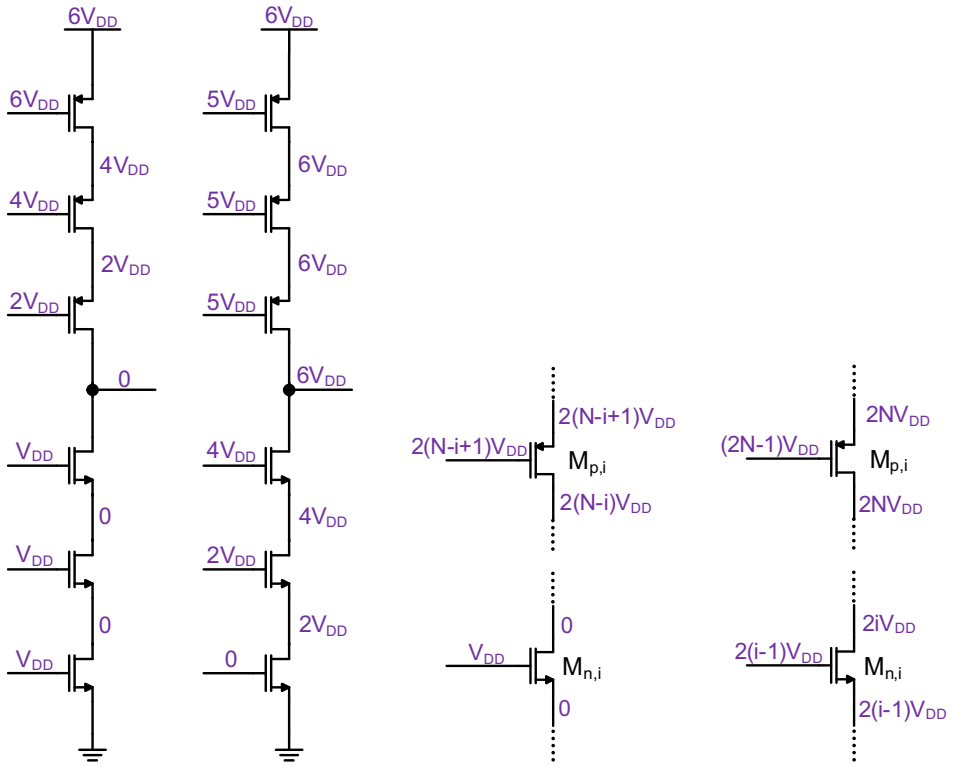
In most cases, either a series L-C filter is included at the output, or the load is tuned (as would happen with an antenna) and only the fundamental component of the pulse reaches it. The RMS amplitude of this component is $\frac{\sqrt{2}}{\pi}$ times the peak-to-peak amplitude of the pulse, and the power delivered to the load at the fundamental becomes:

$$P_{L,\mathrm{fund}} = \frac{2N^2 V_{\mathrm{DD}}^2 R_L}{\pi^2} \Big[\frac{1}{R_L + R_{\mathrm{on_n}}} + \frac{1}{R_L + R_{\mathrm{on_p}}} \Big]^2 \tag{3.29}$$

and if $R_{\mathrm{on}} = R_{\mathrm{on_n}} = R_{\mathrm{on_p}}$:

$$P_{L,\mathrm{fund}} = \frac{8}{\pi^2} \frac{N^2 V_{\mathrm{DD}}^2 R_L}{(R_L + R_{\mathrm{on}})^2} \simeq \frac{8}{\pi^2} \frac{N^2 V_{\mathrm{DD}}^2}{R_L} \tag{3.30}$$

Therefore, the power at the fundamental is $8/\pi^2$=0.9 dB smaller than the power delivered by the pulse. In the remainder of this chapter, P_L and P_{sat} will always refer to $P_{L,\mathrm{fund}}$.



(a) $N = 3$, Output low. (b) $N = 3$, Output high. (c) ith nFET, Output low. (d) ith nFET, Output high.

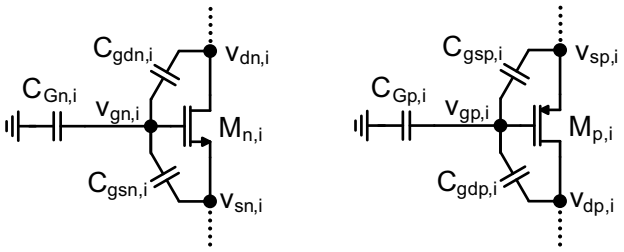
Figure 3.17: Voltage distribution in a stacked-CMOS cascode with $N = 3$ and in the general, ith, case in the two switching states.

For a certain power level P_L, R_{opt} can be calculated by:

$$R_{\text{opt}} \simeq \frac{8}{\pi^2} \frac{N^2 V_{\text{DD}}^2}{P_L} \tag{3.31}$$

There are some additional considerations related to the dynamic operation of the CMOS stacked cascode amplifier. Equations (3.17) and (3.18) for sizing the gate capacitors, are based on the assumption that the circuit is operating as a linear amplifier. This is approximately true for the nFET-only stack operating in class-AB mode, but becomes invalid for class-D operation.

To calculate the capacitors $C_{Gn,i}$ and $C_{Gp,i}$ of figure 3.15a, the ideal voltage distribution along the stack during the two switching conditions of the amplifier needs to be considered. To simplify the derivation, the assumption is made that $V_{\text{min,n}} = V_{\text{min,p}} \simeq 0$ and that the gate voltage levels that turn on and off the transistors are V_{DD} and 0. Figures 3.17a and 3.17b illustrate the voltage distribution that ensures that no transistor is subject to a terminal voltage difference higher than $2V_{\text{DD}}$ when $N = 3$.



(a) nFET. (b) pFET.

Figure 3.18: Equivalent circuits for the calculation of $C_{Gn,i}$ and $C_{Gp,i}$.

The voltage distribution for the general, ith transistor case is shown in figures 3.17c and 3.17d. The voltage values in these figures can be used to calculate the necessary DC components and voltage swings. Specifically, the gate voltage of each FET needs to have a DC component, that is the average between the two switching states:

$$V_{Gn,i} = \left(i - \frac{1}{2}\right)V_{\mathrm{DD}} \tag{3.32}$$

$$V_{Gp,i} = \left(2N - i + \frac{1}{2}\right)V_{\mathrm{DD}} \tag{3.33}$$

The signal amplitude at each gate can be calculated by subtracting the gate voltages between the two switching states and dividing by two:

$$v_{gn,i} = \frac{2i-3}{2}V_{\mathrm{DD}} \tag{3.34}$$

$$v_{gp,i} = \frac{2i-3}{2}V_{\mathrm{DD}} \tag{3.35}$$

Similarly, the source and drain swings become:

$$v_{sn,i} = (i-1)V_{\mathrm{DD}} \tag{3.36}$$

$$v_{sp,i} = (i-1)V_{\mathrm{DD}} \tag{3.37}$$

$$v_{dn,i} = iV_{\mathrm{DD}} \tag{3.38}$$

$$v_{dp,i} = iV_{\mathrm{DD}} \tag{3.39}$$

The gate DC voltages are generated by the resistive divider in figure 3.15a, while the swing is generated by the capacitors $C_{Gn,i}$ and $C_{Gp,i}$, that form capacitive dividers with C_{gs} and C_{gd}. These dividers introduce a portion of the source and drain swing at the gate, ensuring that the transistor will turn off when its V_{DS} is $2V_{\mathrm{DD}}$, retaining the corresponding voltages. The values of the capacitors can be calculated by employing the simplified equivalent circuits

of figure 3.18. The gate voltage $v_{gn,i}$ needs to satisfy the KCL:

$$v_{gn,i}C_{Gn,i} + (v_{gn,i} - v_{sn,i})C_{gsn,i} + (v_{gn,i} - v_{dn,i})C_{gdn,i} = 0 \tag{3.40}$$

Substituting equations (3.34), (3.36) and (3.38) and solving for $C_{Gn,i}$ yields:

$$C_{Gn,i} = \frac{C_{gsn,i} + 3C_{gdn,i}}{2i - 3} \tag{3.41}$$

for $i = 2, \ldots, N$. And similarly for the pFET:

$$C_{Gp,i} = \frac{C_{gsp,i} + 3C_{gdp,i}}{2i - 3} \tag{3.42}$$

Therefore, $C_{Gn,i}$ and $C_{Gp,i}$ need to be progressively smaller in order to facilitate the higher swing required at the transistor gates.

Lastly, special arrangements are necessary to set the DC voltage level at the output. This point is a high-impedance node since it is connected directly to the drains of the two stacks. Therefore, the DC voltage is ill-defined and a mechanism is necessary to set it to the mid-rail, NV_{DD}. In the circuit of figure 3.15a, this is achieved by self-biasing through the feedback resistor R_F. This resistor essentially forces the inverter to operate in the middle of its voltage transfer characteristic and is chosen, along with the other resistors of the resistive divider, to be large enough so that it does not impede with the AC operation of the circuit. A different technique to set the output bias with a separate feedback loop can be found in [107].

A CMOS stacked cascode was previously proposed as a low-frequency driver in [108, 109] in bulk CMOS. However, the dynamic bias method employed would, unlike the capacitive dividers, limit the frequency of operation and the number of devices that can be stacked.

3.4. Loss Mechanisms

There are four important, interrelated loss mechanisms associated with the stacked CMOS cascode. They are similar to the ones encountered in simple inverter-based class-D PAs [106, 110, 111], but some additional considerations are necessary.

3.4.1. Shoot-Through Current

As in every inverter, the pFETs and nFETs have finite turn-on and turn-off times. When a transition happens at the input of the inverter, there will be a short period of time when both transistors are on, causing a large shoot-through current spike from V_{DD} to ground.

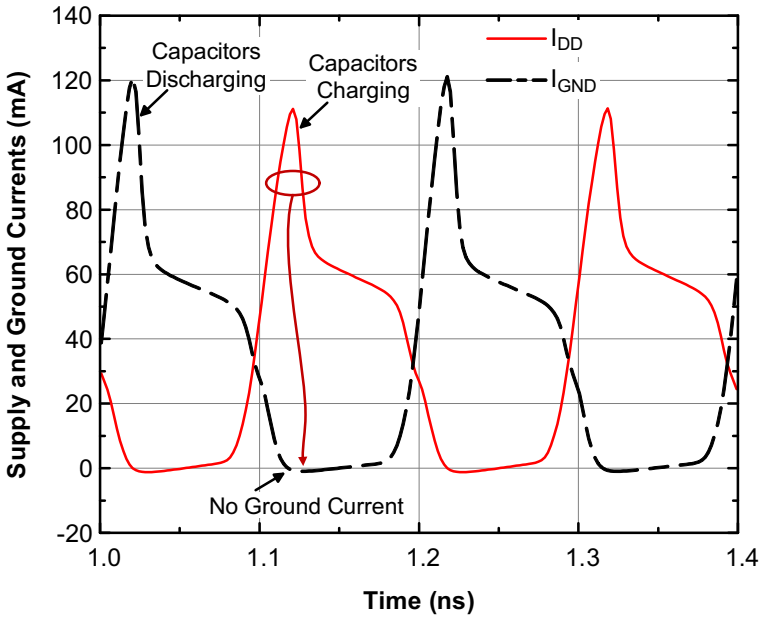


Figure 3.19: Transient simulations of the supply and ground currents with a 5 GHz sinusoidal input.

Simulation indicated that these spikes manifest when the amplifier is driven with very sharp-edged pulses, which are unrealistic for the frequencies of interest (>1 GHz). For example, figure 3.19, reproduces the simulated supply and ground currents in a CMOS stacked cascode with $N = 3$, when driven by a sinusoid at 5 GHz. No significant shoot-through current can be observed.

3.4.2. Resistive Losses

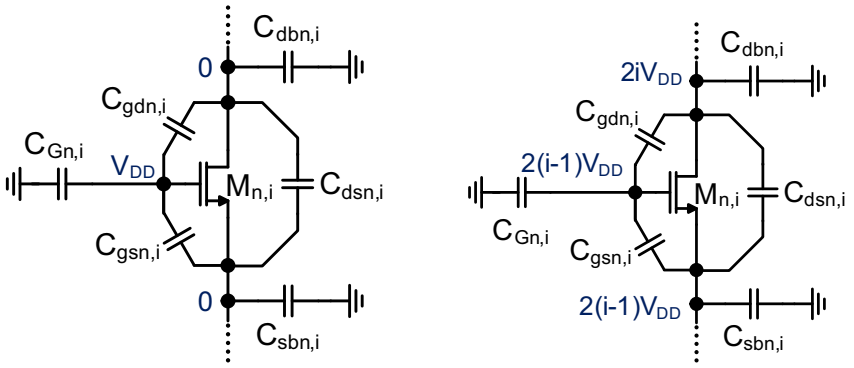
The transistors in the class-D amplifier ideally operate in the triode region for most of their conduction cycle. As illustrated in figures 3.15b and 3.15c, the currents $I_{\mathrm{max,p}}$ and $I_{\mathrm{max,n}}$ that flow to the load also cause $I \cdot R$ losses on the "on" resistances of the transistors. This loss can be calculated by:

$$P_R = \frac{1}{2} I_{\mathrm{max,p}}^2 R_{\mathrm{on_p}} + \frac{1}{2} I_{\mathrm{max,n}}^2 R_{\mathrm{on_n}} = \frac{N^2 V_{\mathrm{DD}}^2}{2} \left[\frac{R_{\mathrm{on_p}}}{(R_L + R_{\mathrm{on_p}})^2} + \frac{R_{\mathrm{on_n}}}{(R_L + R_{\mathrm{on_n}})^2} \right] \tag{3.43}$$

If $R_{\mathrm{on_p}} = R_{\mathrm{on_n}} = R_{\mathrm{on}}$ by sizing for equal driving strength, then

$$P_R = \frac{N^2 V_{\mathrm{DD}}^2 R_{\mathrm{on}}}{(R_L + R_{\mathrm{on}})^2} \tag{3.44}$$

The above equations are based on the assumption that R_{on} in independent of V_{DS}. In nanoscale CMOS, this is not exactly the case and R_{on} will exhibit some dependence due to



(a) Output low. (b) Output high.

Figure 3.20: Voltages across the capacitors of the ith nFET during the two switching states.

short channel effects. A good approximation is to use R_{on} when $V_{\text{DS}} = 0$, since the voltage drop across the transistors is very small in properly designed amplifiers.

The resistive losses can be minimized by increasing the transistor width, thus reducing R_{on_p} and R_{on_n}. However, this increases the parasitic capacitances, which will introduce loss at high frequencies.

3.4.3. Capacitive Losses

The most significant source of power loss for every inverter operating at high frequencies is the dynamic power dissipation due to the parasitic capacitances. When a capacitor is charged and discharged to different voltage levels through a switch, it will result in energy loss in the parasitic resistance of the switch, which depends only on the amount of charge that is pumped in and out of the capacitor, and not on the resistance of the switch itself [112].

By employing Tellegen's theorem, and when the resistive losses are ignored, it can be shown that in a circuit with two switching states the capacitive losses can be calculated by [113]:

$$P_C = f \sum_{\text{capacitors}} q_i \Delta v_i = f \sum_{\text{capacitors}} C_i \Delta v_i^2 \tag{3.45}$$

where Δv_i is the difference of voltages across the capacitor among the two switching states, q_i is the difference in charge stored in the capacitor between the two states and C_i is the capacitor value, when assumed constant with voltage (i.e. a linear capacitor).

Figure 3.20 illustrates the capacitors associated with the ith nFET in the stack, along with the voltage levels across their terminals during the two switching states. By employing equation (3.45), the loss becomes:

$$P_{Cn,i} = \left[C_{gsn,i} + 9C_{gdn,i} + (2i-3)^2 C_{Gn,i} + 4i^2 C_{dbn,i} + 4(i-1)^2 C_{sbn,i} + 4C_{dsn,i}\right] f V_{\text{DD}}^2 \tag{3.46}$$

Substituting $C_{Gn,i}$ from equation (3.41) yields:

$$P_{Cn,i} = 2\big[(i-1)C_{gsn,i} + 3iC_{gdn,i} + 2i^2C_{dbn,i} + 2(i-1)^2C_{sbn,i} + 2C_{dsn,i}\big]fV_{\mathrm{DD}}^2 \tag{3.47}$$

and a similar expression holds for the ith pFET:

$$P_{Cp,i} = 2\big[(i-1)C_{gsp,i} + 3iC_{gdp,i} + 2i^2C_{dbp,i} + 2(i-1)^2C_{sbp,i} + 2C_{dsp,i}\big]fV_{\mathrm{DD}}^2 \tag{3.48}$$

The transistors with $i = 1$ require separate consideration because there is no C_G capacitance:

$$P_{Cn,1} = \big[C_{gsn,1} + 9C_{gdn,1} + 4(C_{dsn,1} + C_{dbn,1})\big]fV_{\mathrm{DD}}^2 \tag{3.49}$$

$$P_{Cp,1} = \big[C_{gsp,1} + 9C_{gdp,1} + 4(C_{dsp,1} + C_{dbp,1})\big]fV_{\mathrm{DD}}^2 \tag{3.50}$$

The total dynamic power dissipations then becomes:

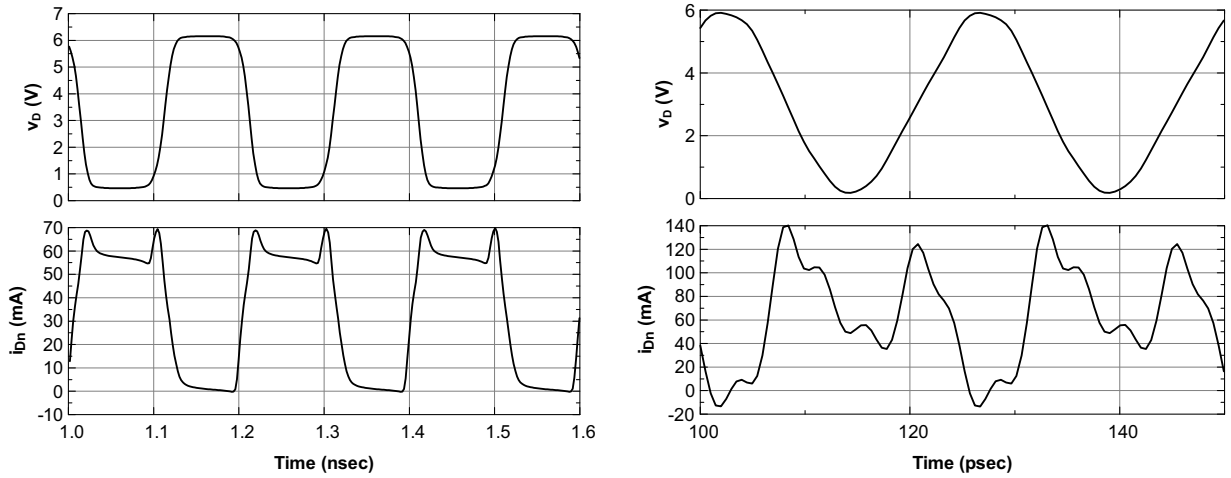
$$P_C = \sum_{i=1}^{N}(P_{Cn,i} + P_{Cp,i}) \tag{3.51}$$

The above equations indicate that the parasitic capacitors of transistors closer to the output node, i.e. with increasing i, will *contribute more* to P_C because they are charged and discharged to higher voltages. Moreover, unlike C_{gs} and C_{gd}, the contribution of C_{db} and C_{sb} increases with i^2 and will dominate the loss unless an SOI process that minimizes them is employed. These facts will have significant implications in the transistor sizing, as will be shown in the following section.

It is important to note that the above derivation ignores the fact that the transistors will have finite on resistance, and as a result, a voltage $I_{\max}R_{\mathrm{on}}$ will build up on them when conducting. This will result in slightly different charging and discharging voltages in figure 3.20, and thus in slightly different capacitive loss, coupling the resistive and capacitive losses. However, $I_{\max}R_{\mathrm{on}}$ will be small in order to maximize the output power and the above approximation is valid. Furthermore, it was assumed that the capacitance value during the on and off states remains unchanged. This is not entirely true, as the intrinsic C_{gs} and C_{gd} depend on the bias voltages. During approximate calculations the average of the on and off values can be used.

3.4.4. Non-Ideal Switching - Overlap

During the derivation of equation (3.43) for the resistive losses, the transistors were assumed to be operating in triode for 100% of their cycle. However, due to the presence of the



(a) At 5 GHz. (b) At 40 GHz.

Figure 3.21: Simulated drain voltage and current through the nFET stack when driven by a large amplitude sinusoid.

parasitic capacitors, the terminal voltages will change with finite speed during a transition, and the transistors will spend a portion of their conduction period in their saturation region, operating as current sources, and charging or discharging the capacitors with finite current. The loss during this time interval is not dictated by equation (3.43), but by the voltage-current product along the current source.

Initially, this non-ideality manifests itself by increasing the rise and fall time of the pulses, which keeps degrading with increasing frequency. Eventually, the saturation period and the rise and fall times will occupy most of the switching period and the output voltage will resemble a sinusoid (or a slew-rate limited amplifier), rather than a pulse. An equivalent view of this problem is in terms of the harmonic content of the signal. The parasitic capacitors will low-pass filter the odd harmonics that constitute the pulse and the output voltage will be progressively more sinusoidal at increasing frequency.

The net effect is that the amplifier will eventually operate in class-AB mode, with voltage-current overlap for most of its conduction period, resulting in higher loss and efficiency degradation. Furthermore, when the output voltage resembles more a sinusoid than a pulse, the output power will also be lower since the fundamental component will have a lower magnitude.

Figure 3.21a reproduces a transient simulation of the output voltage v_{D} and of the current through the nFET stack i_{Dn} of a CMOS stacked cascode with $N = 3$ at 5 GHz. Both the output voltage and current have a pulse shape with minimal overlap due to finite pulse rise and fall times (the two spikes in the current are due to charging and discharging of the parasitic capacitors), indicating class-D operation. On the contrary, figure 3.21b reproduces the

same quantities simulated at 40 GHz. In this case, the rise and fall times extend throughout the period and the capacitors are never fully charged or discharged. As a result, the almost sinusoidal output voltage exhibits extended periods of overlap with the current, which is dominated by large capacitor charging and discharging transients.

3.5. Design Considerations

The analysis presented in the previous sections needs to be linked with the optimal transistor size selection that maximizes the efficiency for a certain output power level.

Equations (3.43) and (3.51) reveal a frequency dependent trade-off when selecting the transistor width, since P_R and P_{Cd} are affected the opposite way by it. Larger widths decrease R_{on}, and thus P_R, but introduce higher parasitic capacitances, which result in increased P_{Cd}. Therefore, at each frequency f, there will be an optimum size that balances between capacitive and resistive losses. Finding this optimum is not trivial due to the differences in the behavior of each transistor based on its location in the stack. Transistors that are closer to the output will contribute more to the capacitive loss, and thus need to be sized differently than those closer to the input. Similarly, pFETs have lower mobilities than nFETs, and higher R_{on}, and their optimum size is different than the nFETs.

In older CMOS process nodes ($\geq$65 nm), pFETs had almost half the mobility of the nFET and sizing an inverter for equal driving strengths would result in the pFETs having twice the size for the nFETs, leading to twice as much parasitic capacitance. This would limit the maximum frequency of operation and degrade the efficiency. However, in deep nanoscale CMOS, aggressive application of strain has improved the hole mobility and the pFETs are only slightly slower than the nFETs. In IBM's 45 nm SOI CMOS, sizing for equal driving strength results in $W_p/W_n = 1.25 : 1$, which is also confirmed by the $\mathrm{f_T}$ measurements presented in chapter 2.

In any case, a systematic method is necessary to estimate the optimum transistor width of the stack. This can be achieved by observing that the "on" resistance of each FET, as well as its parasitic capacitances, can be expressed as a function of width:

$$\begin{aligned}
R_{\mathrm{on_p},i} &= R'_{\mathrm{on_p}}/W_{\mathrm{p},i}\ , \quad R_{\mathrm{on_n},i} = R'_{\mathrm{on_n}}/W_{\mathrm{n},i} \\
C_{gsp,i} &= C'_{gsp}W_{\mathrm{p},i}\ , \quad C_{gsn,i} = C'_{gsn}W_{\mathrm{n},i} \\
C_{gdp,i} &= C'_{gdp}W_{\mathrm{p},i}\ , \quad C_{gdn,i} = C'_{gdn}W_{\mathrm{n},i} \\
C_{dsp,i} &= C'_{dsp}W_{\mathrm{p},i}\ , \quad C_{dsn,i} = C'_{dsn}W_{\mathrm{n},i} \\
C_{dbp,i} &= C'_{dbp}W_{\mathrm{p},i}\ , \quad C_{dbn,i} = C'_{dbn}W_{\mathrm{n},i} \\
C_{sbp,i} &= C'_{sbp}W_{\mathrm{p},i}\ , \quad C_{sbn,i} = C'_{sbn}W_{\mathrm{n},i}
\end{aligned} \tag{3.52}$$

Table 3.1: 45 nm SOI CMOS parameters

	pFET			**nFET**		
R'_{on}	$365\,\Omega \cdot \mu m$			$275\,\Omega \cdot \mu m$		
	Out high $V_{\text{SG}} = 1.1\,\text{V}$ $V_{\text{DS}} = 0\,\text{V}$	**Out low** $V_{\text{SG}} = 0\,\text{V}$ $V_{\text{DS}} = 2\,\text{V}$	**Avg.**	**Out high** $V_{\text{GS}} = 0\,\text{V}$ $V_{\text{DS}} = 2\,\text{V}$	**Out low** $V_{\text{GS}} = 1.1\,\text{V}$ $V_{\text{DS}} = 0\,\text{V}$	**Avg.**
C'_{gs} (fF/μm)	0.49	0.36	0.43	0.45	0.59	0.52
C'_{gd} (fF/μm)	0.46	0.26	0.36	0.37	0.56	0.47
C'_{ds} (fF/μm)	0.05	0.11	0.08	0.12	0.02	0.07
C'_{db} (fF/μm)	0.15	0.15	0.15	0.1	0.1	0.1
C'_{sb} (fF/μm)	0.18	0.18	0.18	0.1	0.1	0.1

where, for minimum gate length devices, R'_{on_p}, R'_{on_n}, C'_{gsp}, C'_{gsn}, C'_{gdp}, C'_{gdn}, C'_{dsp}, C'_{dsn}, C'_{dbp}, C'_{dbn}, C'_{sbp}, C'_{sbn} depend exclusively on the CMOS process parameters.

The output power, resistive and capacitive losses, as calculated from equations (3.29), (3.43) and (3.51), respectively, can also be expressed as a function of the transistor width. When the overlap and shoot-through current losses are ignored, the drain efficiency of the amplifier then becomes:[2]

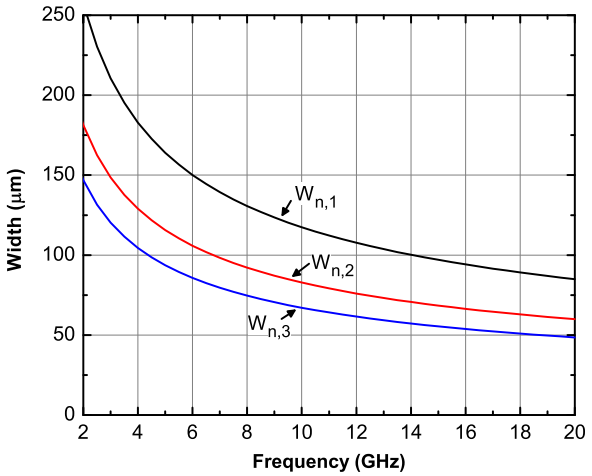
$$\eta_{\text{drain}} = \frac{P_L}{P_L + P_R + P_{Cd}} = f(W_{\text{p},i}, W_{\text{n},i}) \tag{3.53}$$

The transistor widths can be selected such that the efficiency is maximized. Although $\eta_{\text{drain}} = f(W_{\text{p},i}, W_{\text{n},i})$ shows strong nonlinear dependence on $W_{\text{p},i}$ and $W_{\text{n},i}$, the function can be shown to be *concave* and thus exhibits a single, global maximum, which can be found by employing a numerical optimizer, or a nonlinear equation system solver.

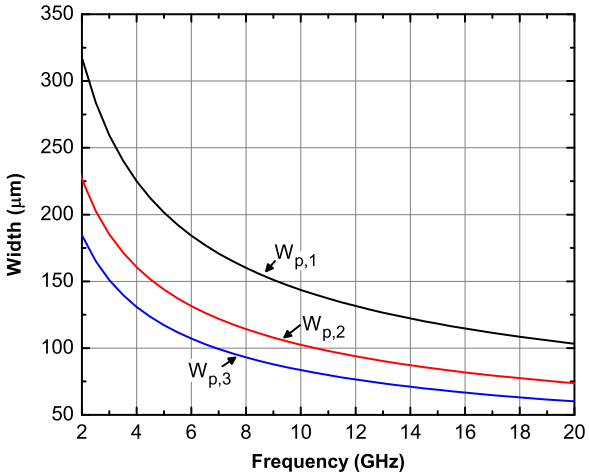
Obviously, equation (3.53) as well as (3.29), (3.43) and (3.51) are derived after several approximations, and the optimization result itself will be only an approximation. This is especially true at higher frequencies, when the amplifier operates in class-AB mode, and the overlap losses will be more significant than the resistive. However, the optimization result can still be used as a starting point for further refinement in the circuit simulator.

The proposed design methodology was employed for the design of a stacked-cascode CMOS amplifier with $N = 3$ and V_{DD}=1.1 V, that drives a 50 Ω load in the 45 nm SOI process. The theoretical maximum power for this amplifier is calculated from equation (3.30) to be 22.5 dBm. Thin oxide devices with minimum gate length are employed, whose approximate R'_{on} and capacitance per micron of width are summarized in Table 3.1. The transistor widths that result to maximum drain efficiency versus frequency are illustrated

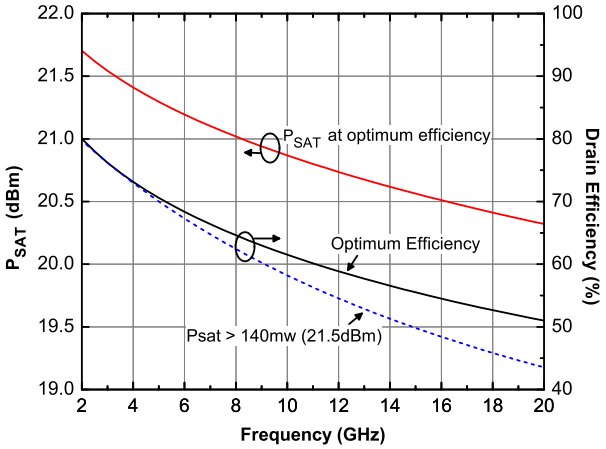
[2] in a multistage design, the PAE can be calculated with a similar formula and adding the capacitive losses of the preceding stages at the denominator



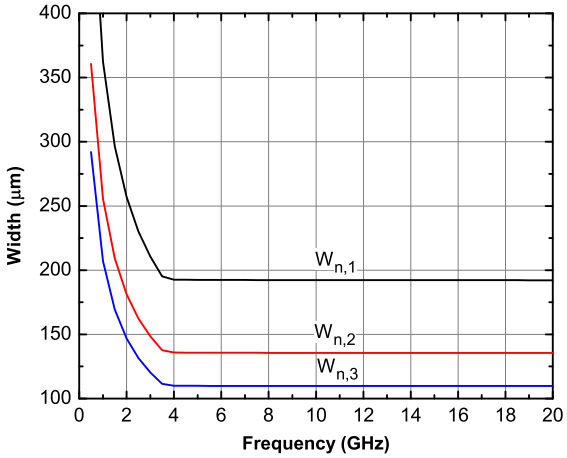
(a) nFET widths.



(b) pFET widths.



(c) Efficiency and output power.



(d) Optimum nFET versus frequency when the P_{sat} >21.5 dBm constraint is imposed.

Figure 3.22: Calculated optimum transistor widths versus frequency.

in figures 3.22a and 3.22b. As expected, they become smaller with frequency, since the capacitive losses become increasingly important at higher frequencies. For the same reason, the optimum width is smaller for transistors closer to the output. Interestingly, the width ratios are independent of frequency:

$$\frac{W_{\text{n,2}}}{W_{\text{n,1}}} \simeq 0.7, \quad \frac{W_{\text{n,3}}}{W_{\text{n,1}}} \simeq 0.57 \tag{3.54}$$

And the ratio of the pFETs and nFETs was found to be independent of frequency and equal to approximately 1.2, the same as the f_T ratios (chapter 2).

Figure 3.22c illustrates the calculated maximum efficiency and corresponding output power at each frequency. A portion of the output power needs to be sacrificed as the

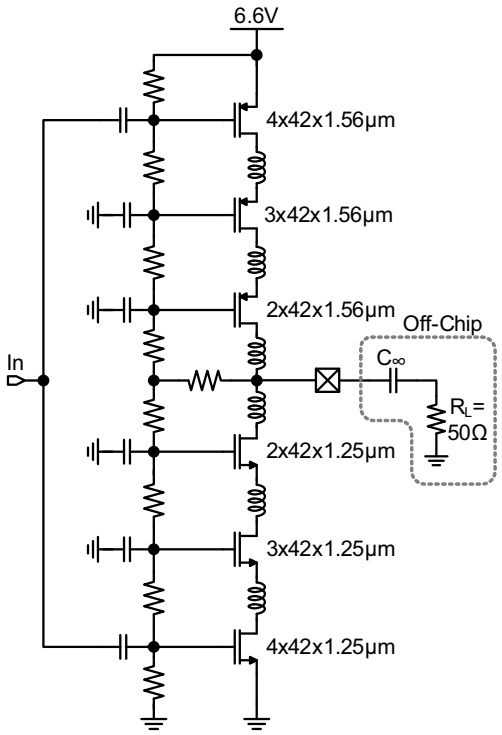


Figure 3.23: Schematic of the implemented three stack cascode CMOS PA.

frequency increases in order to maintain the maximum drain efficiency. Specifically, as the frequency increases from 2 GHz to 20 GHz, the output power at the optimum decreases substantially, by almost 1.5 dBm.

If this output power reduction is intolerable, an alternative design approach is to impose a minimum power constraint in the numerical maximization algorithm when computing the efficiency maximum. Such an example is included in figure 3.22c, where the minimum output power is required to always be higher than 21.5 dBm, slightly lower than the theoretical maximum of 22.5 dBm. In this case the high frequency efficiency is compromised, becoming almost 10% lower than the optimum at 20 GHz, revealing an output power-efficiency trade-off. The corresponding transistor widths for the nFETs are reproduced in figure 3.22d. Their values flatten out after a certain frequency, which is necessary to maintain the minimum output power.

Based on the above analysis, the stacked CMOS amplifier of figure 3.23 was designed. The transistor sizes were selected based on the minimum output power constraint efficiency maximization approach. After further optimization in the circuit simulator, slight adjustments compared to the calculated values were necessary. The number of fingers of the nFET and pFET with the same i was kept constant and the width was adjusted by increasing the pFET finger width to 1.56 μm.

A transformation network to a load impedance lower than 50 Ω was avoided in this proof-of-concept implementation, in order to to experimentally investigate the behavior of the circuit over a wide range of frequencies. As a result, the PA was loaded directly by the 50 Ω load. For the same reason, no series L-C filter was included in the output.

An inductive broadbanding technique that makes use of the series intermediate inductors [114,115] of figure 3.23 was employed to minimize the effects of the parasitic capacitances at higher frequencies. The parasitic capacitors will partially charge through the inductors, which will reduce the dynamic loss associated with them. This also helps to reduce the filtering of the higher order harmonics, retaining the shape of the output pulse, and preventing the power amplifier from early entering into class-AB mode. A technique that involves the use of shunt inductors to ground in the intermediate nodes, causing a parallel resonance was proposed in [116].

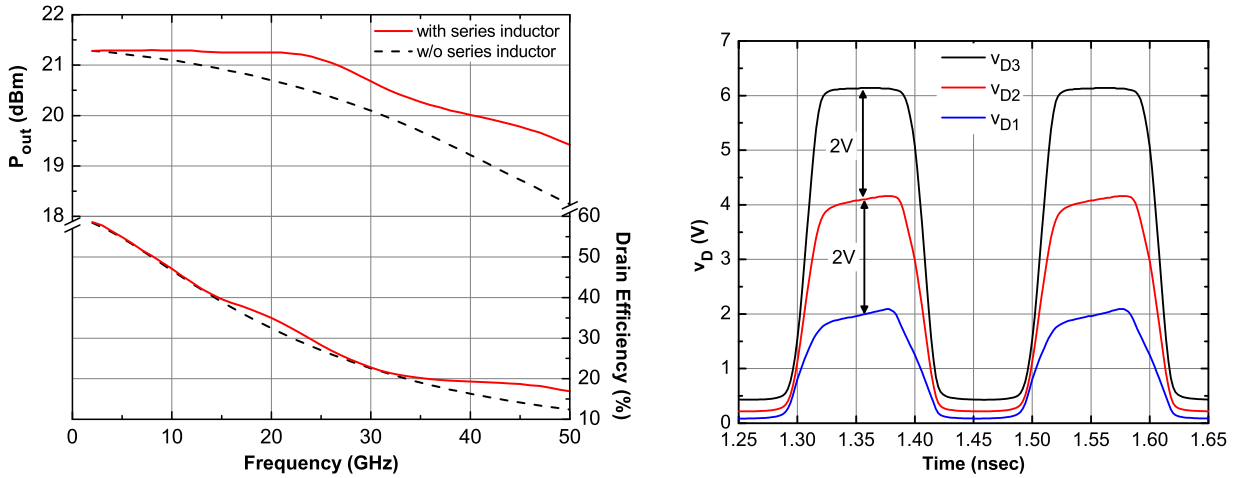
In the circuit of figure 3.23, the series inductors where chosen so that they resonate with the equivalent intermediate capacitors at approximately 40 GHz. The inductance values required to achieve this are small, and were implemented with short transmission lines. This approach also helps to increase the distance between neighboring transistors in the stack, allowing for the heat generated during dynamic operation to dissipate.

Figure 3.24a reproduces the simulated impact of the series inductors on the output power and efficiency of the PA. The efficiency improves from 12% to 17% around 45 GHz. The output power increases over a wide frequency range, becoming almost 1 dB higher at 45 GHz. This substantial power increase at very high frequencies is explained because the transistors would otherwise not have enough time to fully charge and discharge the parasitic capacitors, limiting the output swing. Figure 3.24b illustrates the simulated drain voltages of the nFETs. Due to the proper sizing of the gate capacitors, no voltage is allowed to exceed 2 V, ensuring that the transistors will not breakdown.

3.6. Impact of Process Scaling

The analysis of the previous section indicates that there can be a compelling efficiency benefit in the stacked CMOS cascode amplifier from process scaling. In deeper sub-micron nodes, such as 22 and 16-nm, the transistor ON resistance will improve due to the application of higher levels of strain to their channel, while the parasitic capacitances are going to improve due to the scaling process itself.

Nevertheless, CMOS scaling is always associated with a lower breakdown voltage, which will limit the maximum V_{DS} that can be applied to each transistor and will mandate the stacking of more devices to achieve the same output power level. This will add to the parasitic capacitances of the stack and will, in turn, trim the efficiency benefit of scaling.



(a) Efficiency and output power improvement with and without the series inductors.

(b) Voltage distribution along the drain of the nFETs in the stack at 10 GHz.

Figure 3.24: Simulated performance of the designed 3-stack CMOS power amplifier.

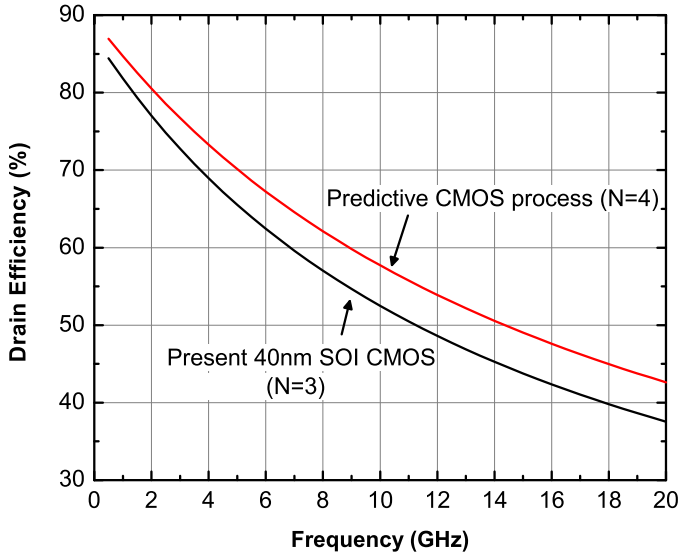


Figure 3.25: Drain efficiency comparison between the 45 nm SOI CMOS stacked CMOS amplifier with $N = 3$ and a $N = 4$ stacked amplifier in the predictive CMOS process.

To quantify this scaling- lower V_{DD} - efficiency trade-off, a predictive CMOS process that features the parasitics of table 3.1, divided by $\sqrt{2}$, but the maximum allowed V_{DS} of each transistor is 0.8 V is considered. In order to achieve 21.5 dBmoutput power with this process, the number of stacked transistors needs to be increased to $N = 4$ (8 in total). Subsequently, the optimum transistor widths that maximize the efficiency under the constraint of a minimum output power of 21.5 dBmare calculated with the procedure described in the previous section, along with the resulting drain efficiency. The results are illustrated in figure 3.25 and compared with the 45 nmdesign with $N = 3$ of the previous section. Although the benefit of scaling is shrunken by the fact that four stacked devices are required, the drain efficiency is still improved by approximately 5%.

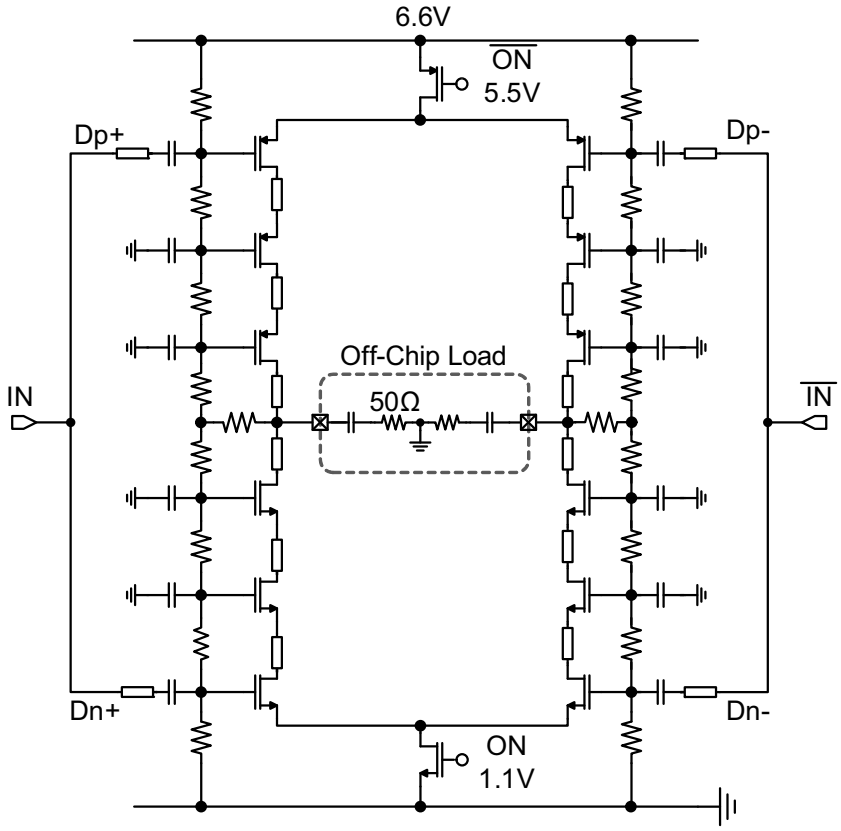
3.7. Power Amplifier System Design

A differential version of the CMOS stacked-cascode amplifier of figure 3.23 was also designed, as shown in figure 3.26a, and was integrated along with a CMOS-only driver in the system of figure 3.26b. The differential configuration allows for the the inclusion of common-mode switches that are able to turn on and off the PA, with minimal impact on performance. The ON/OFF switch can be employed for output amplitude modulation through duty cycle control, rendering the PA capable of operating as an EER transmitter (fig. 3.10a).

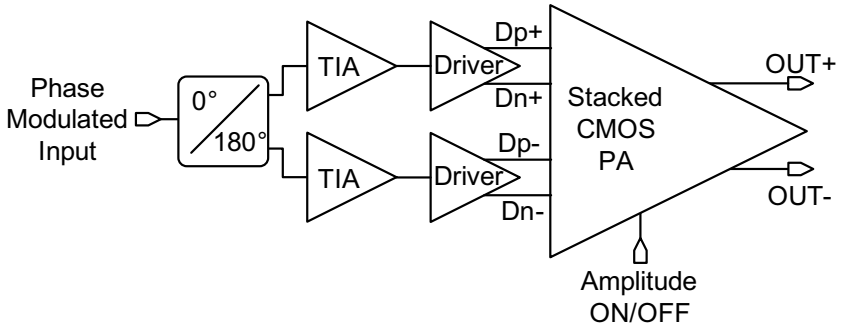
In the block diagram of figure 3.26b, a pair of broadband CMOS Trans-Impedance Amplifiers (TIAs) amplifies the input signal and provides broadband input matching. Following the TIAs, the driver of figure 3.26c, consisting of cascaded, scaled CMOS inverters, is employed to provide adequate rail-to-rail voltage swing to saturate the stacked, cascode CMOS output stage. Two techniques were applied to maximize the bandwidth of the inverters in the driver chain. First, the pFET to nFET gate-width ratio was set to 1.25:1, the same as in the output stage, which corresponds to sizing the devices for equal driving strength. Second, series peaking inductors and transmission lines were employed in the CMOS inverter chain as well as at the TIA. In addition to the broad bandwidth and unlike the tuned nFET-only drivers, this CMOS inverter based driver is less sensitive to poorly modeled layout parasitics, as well as to device model inaccuracies. The TIA and CMOS driver chain were designed by Alex Tomkins.

3.8. Measurement Results

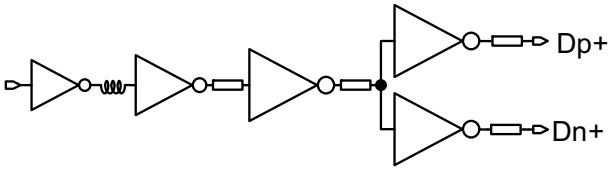
The power amplifier of figure 3.26b was fabricated in the 45 nm SOI CMOS process along with a breakout of the single-ended stacked-cascode CMOS output stage 3.23. The



(a) Schematic of the differential stacked-cascode CMOS output stage.

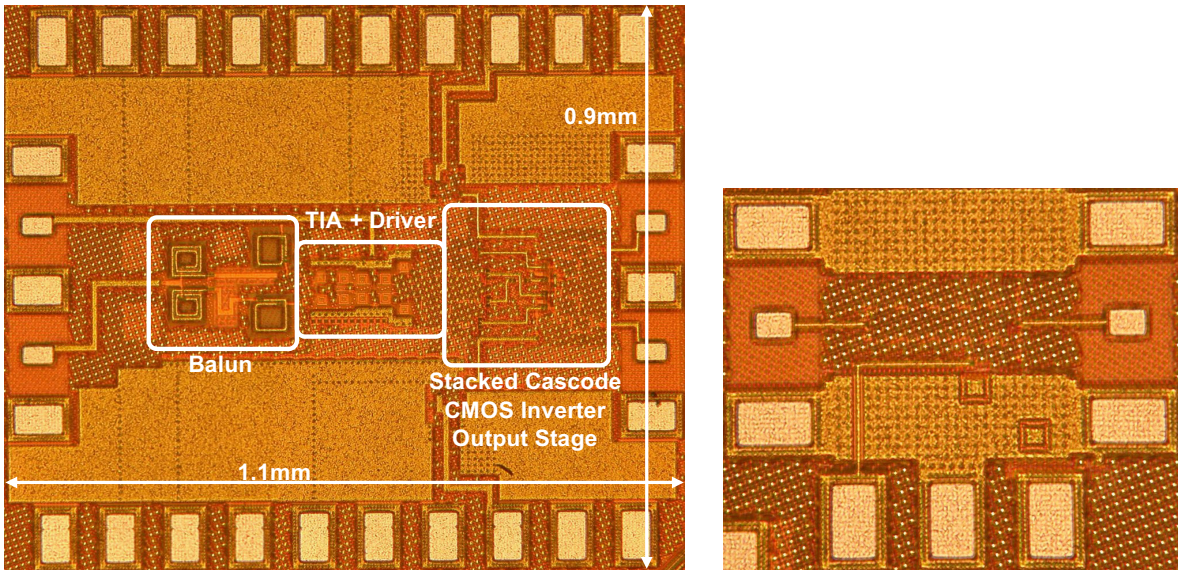


(b) Overall block diagram of the power amplifier.



(c) Block diagram of the CMOS driver.

Figure 3.26: Details of the implemented PA.



(a) Power Amplifier.

(b) Stacked CMOS output stage breakout.

Figure 3.27: Die microphotographs.

die microphotographs of the two chips are shown in figure 3.27

Interestingly, the output stage is completely filled by dummy metal tiles in both cases. This was possible since other than the short lines that connect the common-gate transistors, no other inductor is involved.

3.8.1. Stacked Cascode CMOS Breakout

The breakout was characterized both using small-signal S-parameters and with large signal time domain and output power measurements. Figure 3.28 illustrates the measured S_{21} versus frequency for three supply voltages. The circuit has significant power gain at lower frequencies, which decreases with increasing frequency due to the parasitic capacitors. The low frequency roll-off is due to the series input capacitors in the gates of the input transistors.

Figure 3.29 reproduces the measured saturated output power at the fundamental harmonic, and the associated drain efficiency versus frequency for three supply voltages. The input source employed in the measurement had enough power to saturate the amplifier up to approximately 20 GHz. At 6.6 V supply, the output power is approximately 21.4 dBm whereas the low frequency drain efficiency is 57% and exhibits a 1/f drop as expected for a class-D amplifier dominated by the capacitive losses. When the supply voltage is reduced to 6.0 V, the efficiency increases to 70% while the output power drops by 1 dB.

The output voltage of the amplifier in time domain is shown in figure 3.30 when it is

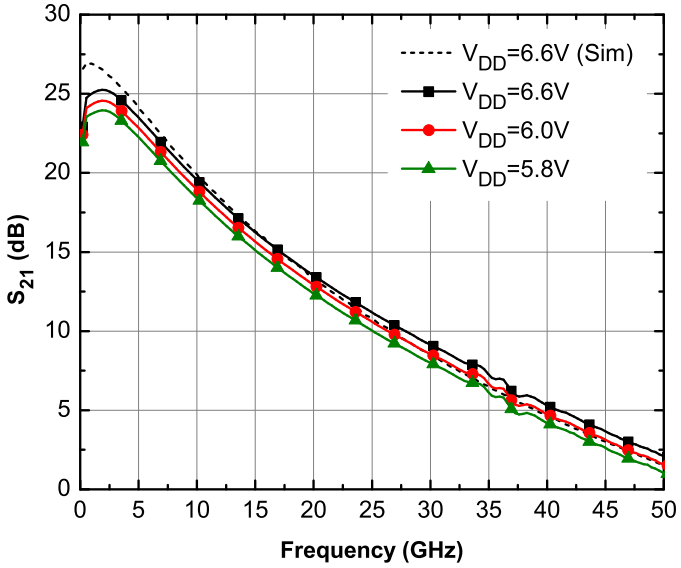
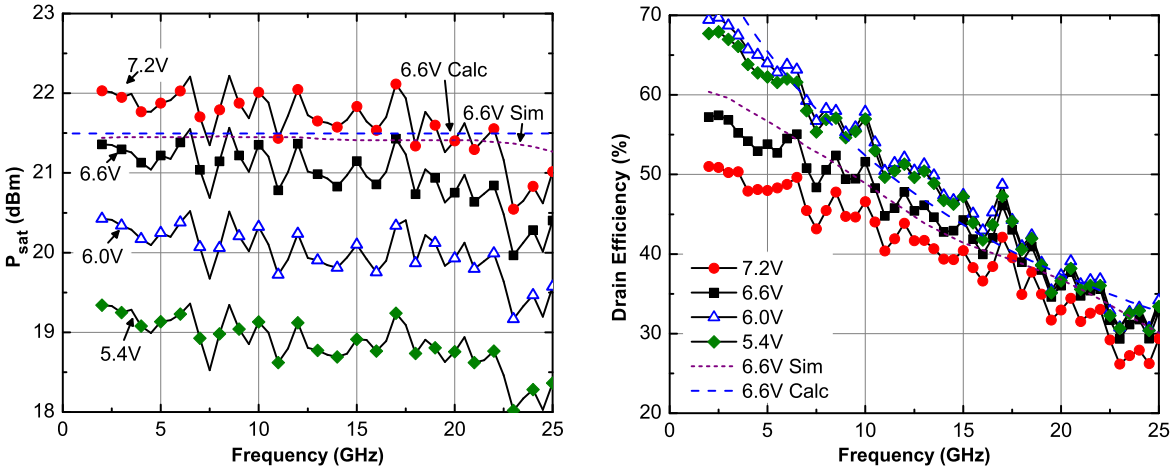


Figure 3.28: Measured S_{21} versus frequency of the CMOS stacked-cascode breakout versus frequency for three supply voltages.



(a) Saturated output power.

(b) Drain Efficiency at P_{sat}.

Figure 3.29: Measured P_{sat} and drain efficiency versus frequency of the CMOS stacked-cascode breakout versus frequency for three supply voltages.

driven by a sinusoid at 2 and 10 GHz. The output signal even at 10 GHz, i.e. 20 Gb/s, has a pulse shape, of 6 V amplitude, demonstrating class-D operation. To further verify the broadband operation of the circuit, a $1\,V_{pp}$, 12 Gb/s $2^7 - 1$ PRBS signal was applied at the input at the stacked-CMOS cascode breakout to assess its performance as a broadband driver. The output eye diagram is shown in figure 3.31. A 20 dB attenuator was included in the setup and without any further attempt for deembedding, the eye amplitude is 5.4 V from a 6.6 V power supply. Although the circuit cannot process the low frequency components of the PRBS sequence due its roll-off (figure 3.28), this result demonstrates the potential of the CMOS stacked-cascode as a broadband CMOS driver for wireline and fiber communications.

Lastly, to investigate the reliability of the circuit, the output power was monitored for over two hours of operation at 2 GHz. A low frequency was selected for this experiment as it is expected that any large voltage swings would degrade the devices more at lower frequencies. The results are reproduces in figure 3.32 and no noticeable reduction in output power was observed.

3.8.2. Differential Power Amplifier

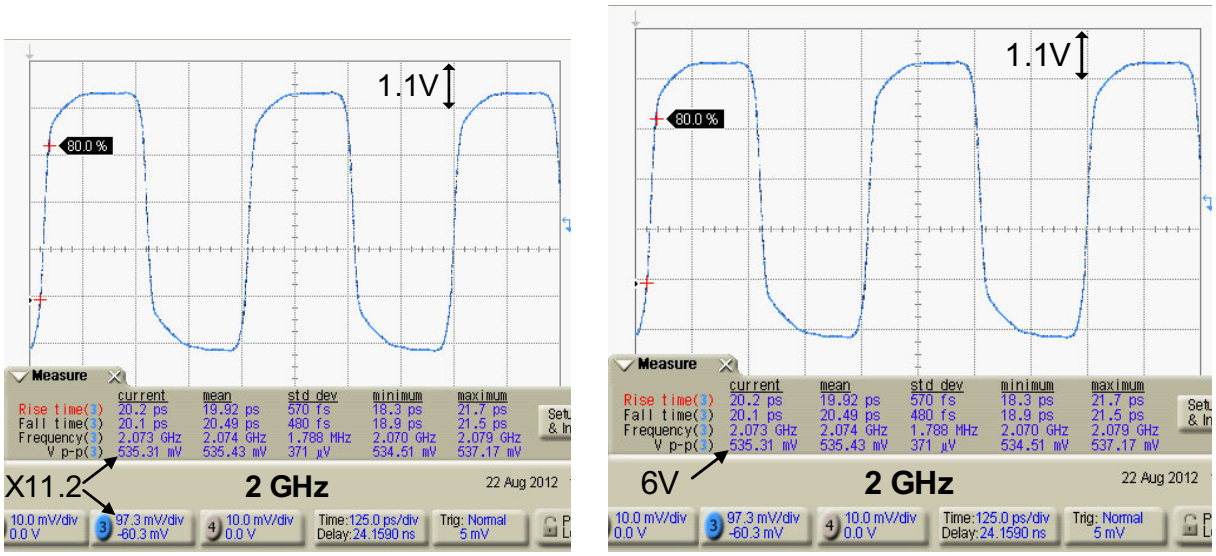
The measured differential output power at the fundamental, drain efficiency and PAE of the CMOS PA are illustrated in figure 3.33. The supply voltage of the output stage is kept at 6.6 V while the driver's is swept. The output power at low frequencies is as high as 24 dBm, which corresponds to a differential swing of 12 V on the load. The output power drops to 22 dBm at 45 GHz, demonstrating the very wide bandwidth achieved, with only minimal use of passive components.

The drain efficiency accounts for the power consumption of the output stage alone whereas the PAE refers to the whole PA, including the drivers. The PAE decreases from 55% at low frequencies, down to 15% at 45 GHz. The efficiency follows a 1/f slope up to approximately 30 GHz and then flattens out due to the action of the inductive broadbanding, as expected also from the simulation of figure 3.24a.

3.9. Summary

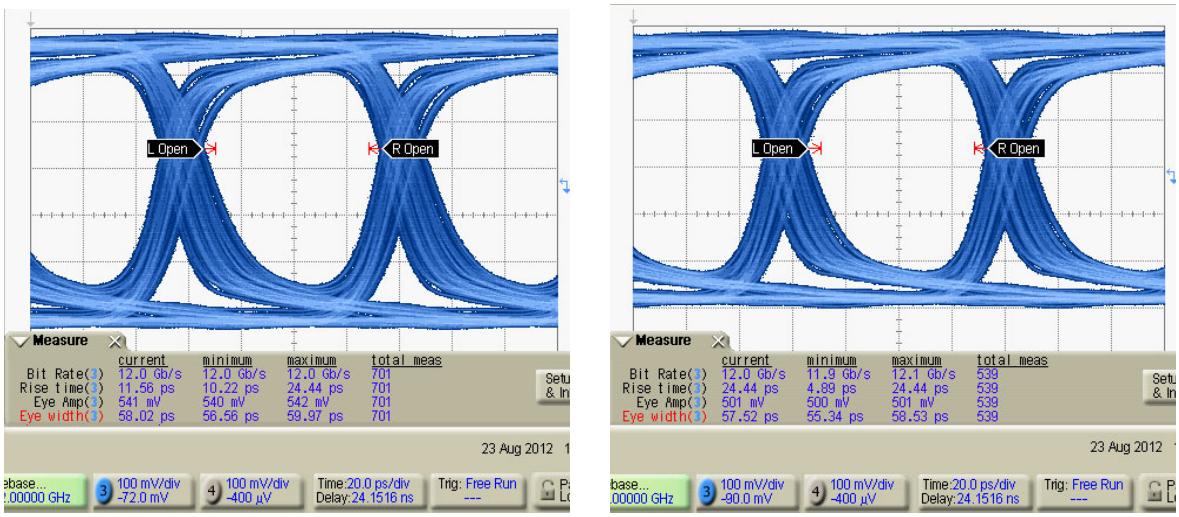
A novel power amplifier solution suitable for nanoscale SOI CMOS has been presented. The PA output stage features transistor stacking to increase the output swing and avoid matching to a very low R_{opt}, along with replacing of the RF choke inductor with a pFET stack. This allows for class-D operation without the necessity of any large, bulky inductors that would occupy silicon area and introduce additional losses.

Table 3.2 compares the presented power amplifier with recently published work. Due to these techniques, the proposed amplifier achieves one of the highest reported efficiencies over



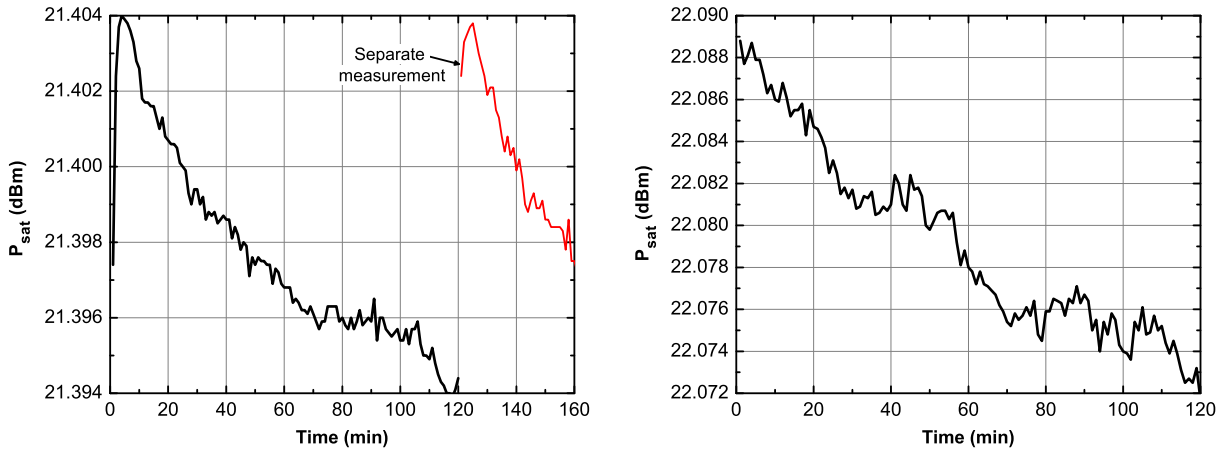
(a) f_{in}=2 GHz. (b) f_{in}=10 GHz.

Figure 3.30: Measured output voltage of the CMOS stacked-cascode breakout when driven by a sinusoidal input signal. V_{DD} = 6.6 V and a 20 dB attenuator was included.



(a) V_{DD} =6.6 V. (b) V_{DD} =6.0 V.

Figure 3.31: Output eye diagrams when the input of the CMOS stacked-cascode is a 12 Gb/s $2^7 - 1$ PRBS signal. A 20 dB attenuator was included in the setup.



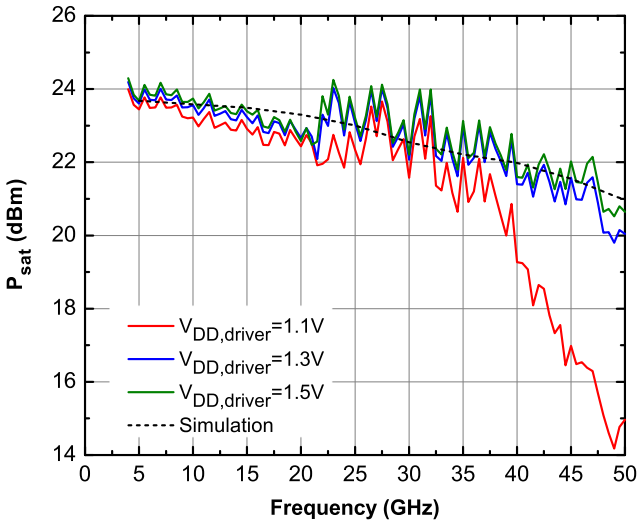
(a) V_{DD} =6.6 V. (b) V_{DD} =7.7 V.

Figure 3.32: Output power versus time of the CMOS stacked-cascode when operating at f_{in} = 2 GHz.

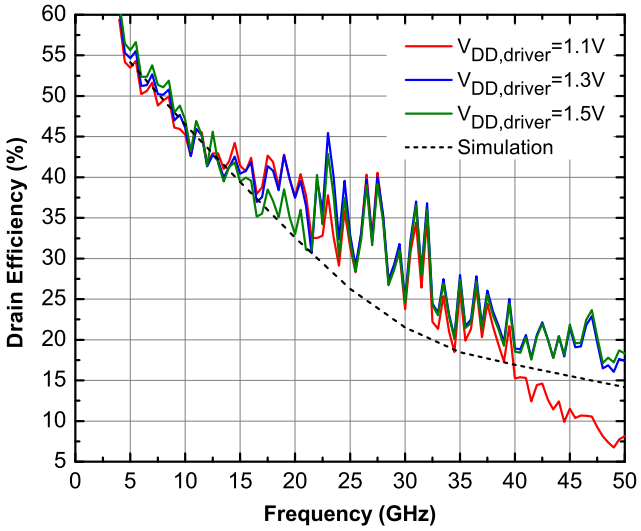
Table 3.2: PA Performance Comparison

	This Work	[100]	[117]	[118]	[119]	[105]
Technology	45 nm SOI	65 nm	32 nm	130 nm	65 nm	45 nm SOI
Frequency	4-50 GHz	2.1 GHz	2.4 GHz	18 GHz	24 GHz	45 GHz
P_{sat}	23.7 dBm @ 5 GHz 23 dBm @ 20 GHz 22 dBm @ 45 GHz	23.4 dBm	25.3 dBm	10.9 dBm	18 dBm	18.2 dBm
PAE	45% @ 5 GHz 30% @ 20 GHz 15.8% @ 45 GHz		35%	23.5%	29%	23%
P_{in}@P_{sat}	<0 dBm			-5 dBm		11.7 dBm

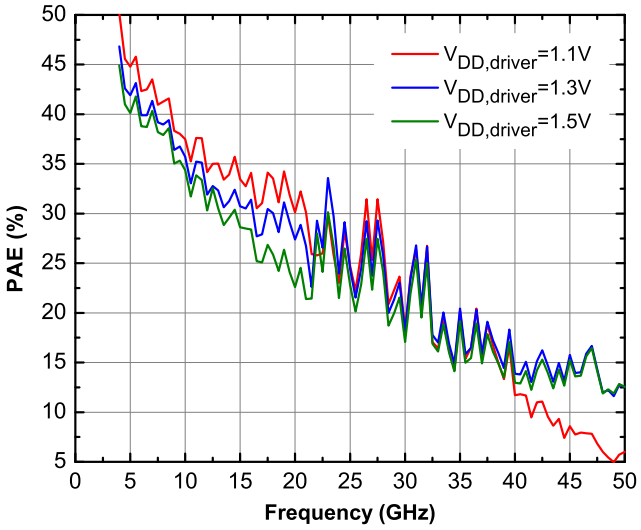
a very wide frequency range.



(a) Saturated output power.



(b) Drain Efficiency at P_{sat}.



(c) Power Added Efficiency at P_{sat}.

Figure 3.33: Measured results for differential power amplifier. The supply voltage of the output stage was kept at 6.6 V while the driver's was swept.

4 Design of a SiGe BiCMOS Transceiver for Point-to-Point Links

The continuous expansion of data-hungry applications in modern cellphones has led to the development of new wireless standards, such as WiMAX, LTE and LTE-advanced, that provide unprecedented data rates to the user. This traffic will be aggregated at the base-station, which needs to feature a data link of adequate bandwidth to the backbone network. The technology for such links already exists, and they can be easily implemented using fiber optics. Nevertheless, there are cases where installing a fiber connection at the base-station level is not feasible. For example, the majority of commercial buildings do not have fiber installations and it can be economically impractical to install one. Similarly, most metropolitan areas do not possess last-mile fiber network infrastructure, and trenching to create it is extremely expensive, or even banned by regulation to minimize disruption. These limitations have pushed forward for the development of cheaper, wireless multi-Gb/s alternatives, where traffic from one or several base-stations can be aggregated at a central location, that already features a fiber connection, and from there onwards to the backbone network.

In 2004, the 71-76 GHz and 81-86 GHz bands, collectively referred to as the E-band, were allocated for this application, with no channelization restrictions across the 5-GHz bandwidths and no particular modulation requirements. There are several advantages for using the E-band for this application. First, the point-to-point nature of the link permits the use of the highly directive antennas, which combined with the low atmospheric attenuation [120], relaxes the required transmit power and thus reduces the design constraints on the power amplifier efficiency and output power. Second, the wide bandwidth available permits high-data-rate transmission by utilizing a simple modulation scheme such as QPSK. Lastly, antennas with very high directivity are easily realizable in this frequency range with apertures less than 1 ft [121].

To this day, several low-cost mm-wave data communication solutions in silicon have been proposed (for example in [6, 122, 123]). The vast majority of them target the 57-64 GHz

band and try to address the stringent channelization and modulation requirements of the popular standards of that band, limiting the maximum achievable data rate. Other solutions proposed low power, 60-GHz radios employing simple direct modulation schemes such as BPSK [124], OOK [125, 126] or upconverted QPSK [127] modulation, and avoiding the 2-GHz channelization requirements imposed by the IEEE802.15c 60-GHz radio communication standard. However, due to the low power consumption and low antenna directivity constraints for short-range mobile data communications, the achievable data rate is limited to values lower than 10 Gb/s.

In the E-band, most of the currently available solutions are implemented with III-V semiconductors [121, 128]. In these systems, the various components (e.g., amplifiers, oscillators, multipliers, etc.) are usually fabricated in separate chips and have to be interconnected with waveguides, resulting in bulky, costly systems. More recently (after the work proposed in this chapter was published), separate transmitter and receiver chipsets have been proposed in SiGe [129] and even 65 nm CMOS [130].

In this chapter, a SiGe BiCMOS transceiver for last-mile, point-to-point communications in the 70-80 GHz band will be described. The proposed modulation scheme is direct QPSK, which is applied to the carrier through a novel, direct, large-power, mm-wave modulator. This avoids both wideband linear upconversion and separate mm-wave linear power amplification, thus retaining the transmitted signal constellation quality and minimizing the overall transceiver power dissipation. Along with the careful selection of the quadrature receiver architecture which ensures broadband I-Q balance, the direct modulator renders the transceiver capable of operation with data rates in excess of 18 Gb/s, on par with the most advanced short-range fiber-optics links. A variant of the proposed direct-QPSK modulator has also been implemented as a stand-alone circuit in the 60 GHz band and will also be presented.

The chapter is organized as follows. Sections 4.1 and 4.2 analyze the transceiver design considerations and architecture. Section 4.3 discusses the circuit design aspects for the transceiver implementation. The experimental characterization is reviewed in Section 4.4. Section 4.5 presents the implementation and measurement results of a stand-alone 60 GHz QPSK modulator and section 4.6 provides concluding remarks.

4.1. System Design Considerations

4.1.1. Link Margin

In order to assess the maximum range and the required output power for a wideband point-to-point link at 75 GHz, a transceiver scenario with the characteristics listed in Ta-

Table 4.1: Link margin analysis parameters

Quantity	Symbol	Value
Receiver bandwidth	B	10 GHz
Receiver temperature	T	270° K
Receiver Noise Figure	F	10 dB
Carrier frequency	f	75 GHz
Transmitter antenna gain	G_{TX}	50 dBi
Receiver antenna gain	G_{RX}	50 dBi

ble 4.1 is considered. The system bandwidth is assumed to be 10 GHz to facilitate data rates above 20 Gb/s. A rather pessimistic noise figure (e.g. compared to [87]) of 10 dB is considered, in order to account for the expected performance degradation due to packaging, process, supply and temperature variations. The antenna gain is assumed to be 50 dBi, 7 dB higher than the minimum allowed limit of 43 dBi set by the FCC (Federal Communications Commission).

Using the values of Table 4.1, the receiver noise floor referred to the input is calculated:

$$N_{\mathrm{RX}} = k_B\, T\, B\, F \simeq -64\,\mathrm{dBm} \tag{4.1}$$

where B and F are the receiver bandwidth and noise figure.

The received power can be obtained by employing Friis' transmission equation:

$$P_{\mathrm{RX}} = P_{\mathrm{TX}} G_{\mathrm{TX}} G_{\mathrm{RX}} \Big(\frac{c}{4\pi f d}\Big)^2 \tag{4.2}$$

Where d is the distance between the transmitter and receiver antenna and P_{TX} is the power at the output of the transmitter. The link margin in dB is defined as:

$$LM = P_{\mathrm{RX}} - N_{\mathrm{RX}} - \alpha - 10.5\,\mathrm{dB} \tag{4.3}$$

where α is the atmospheric attenuation due to rain and 10.5 dB correspond to the SNR required by the QPSK demodulator for operation with a Bit Error Rate (BER) better than 10^{-6} [9].

Figure 4.1 illustrates the maximum achievable link distance versus transmitted power for three different atmospheric attenuation levels and with the link margin equal to 10 dB. It can be seen that, provided highly directive antennas are used, an output power level as low as 0 dBm is sufficient for link operation over 500 m even in the case of severe rain attenuation

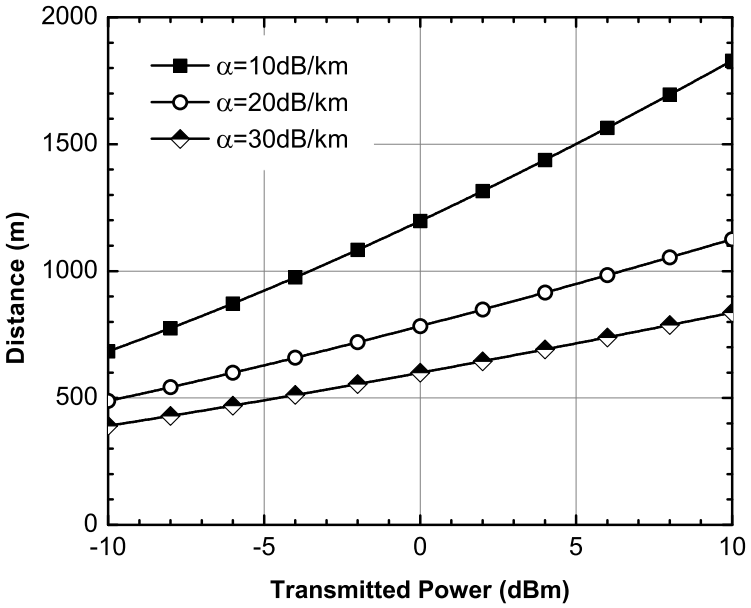


Figure 4.1: Maximum distance versus transmitter power for constant link margin of 10 dB and atmospheric attenuation values of 10, 20 and 30 dB/km.

of 30 dB/km[1]. Since only a moderate output power is required to establish a last-mile, point-to-point link, the transmitter design can be greatly simplified, as will be shown in the following sections.

4.1.2. Modulation Scheme

Spectrally efficient modulation schemes, such as 8-PSK [128] and OFDM-QPSK have been previously proposed for data communication links in the 70-GHz and 80-GHz bands. However, these approaches complicate the design of the transceiver, of the baseband ADCs, and of the digital signal processor, all of which become prohibitively power hungry at data rates exceeding 10 Gb/s per I/Q channel. On the contrary, in [131], a 120 GHz, 10 Gb/s, 2 km wireless link was demonstrated using ASK modulation, which allowed for a significant simplification of not only the mm-wave front-end, but also of the baseband control and data recovery circuitry.

In order to simplify the transceiver architecture and to reduce the overall system power dissipation, a simple modulation scheme such as single carrier, double-sideband QPSK was preferred in this work. Moreover, a direct 70-80 GHz modulator is realized which also provides adequate output power. Since a high-gain waveguide antenna is employed, the transmitted spectrum can be adequately shaped (although not strictly necessary given the low interference probability) with a low-loss waveguide filter placed between the modulator out-

[1]Rain attenuation of 30 dB/km occurs in monsoon-scale rain.

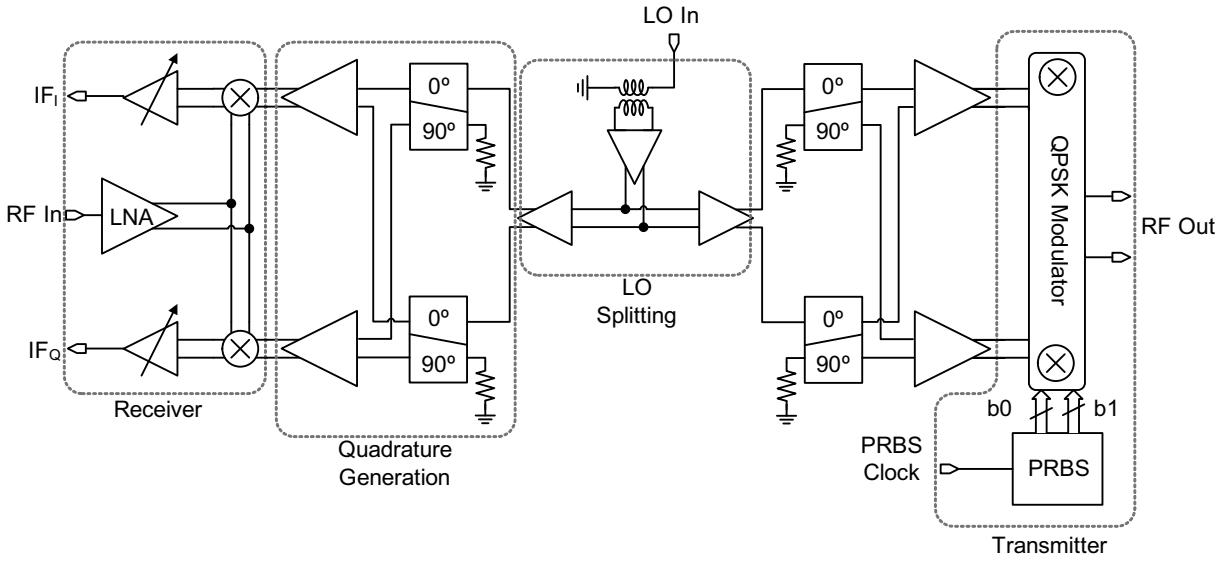


Figure 4.2: Transceiver top-level block diagram.

put and the antenna.

4.2. Transceiver Architecture

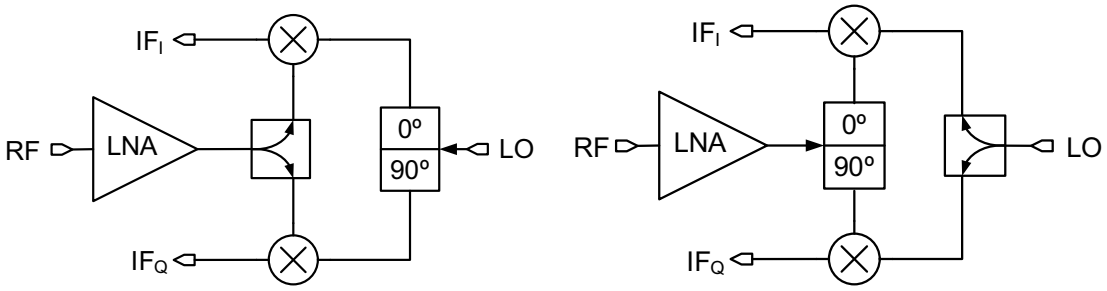
The block diagram of the proposed transceiver is illustrated in figure 4.2. It features (i) a local oscillator distribution tree, (ii) a zero-IF quadrature receiver, (iii) a direct QPSK modulator, and (iv) an on-chip PRBS generator.

The externally-applied LO signal undergoes single-ended-to-differential conversion by an on-chip transformer balun and is split into in-phase (I) and quadrature (Q) parts at both the transmitter and the receiver. The differential, quadrature LO signals are required in the transmitter to form the QPSK modulated carrier and in the receiver for quadrature downconversion, which is necessary for proper demodulation of the I and Q data streams.

The data inputs of the direct QPSK modulator are provided by an on-chip $2^7 - 1$ PRBS generator, operating at up to 20 Gb/s. The PRBS generator is a previous design by Ekaterina Laskin that has been presented in [132].

The zero-IF receiver includes a low noise amplifier (LNA) with active, in-phase, power splitting, two mixers which are driven by the I and Q LO signals, and two variable gain baseband amplifiers, each with a peak gain of over 30 dB.

Two possible zero-IF IQ receiver architectures are shown in figure 4.3 [133]. In the case of figure 4.3a, the LO signals are applied in quadrature to the two mixers while the RF signal at the output of the LNA is split in phase between the two mixers. In the architecture of figure 4.3b, the RF signal is split in quadrature while the LO signals arrive in



(a) I-Q splitting at the LO path. (b) I-Q splitting at the RF path.

Figure 4.3: Quadrature receiver architectures.

phase at the downconvert mixers. In a wideband receiver operating at data rates in excess of 10 Gb/s per I and Q channels, the phase balance of the I-Q splitter at the output of the LNA, figure 4.3b, becomes a difficult-to-satisfy design parameter. Low-loss, wideband, linear mm-wave quadrature splitters are particularly challenging to implement. If this 90-degree splitter does not provide RF signals that are in perfect quadrature over the entire bandwidth of operation, the maximum data rate and receiver sensitivity will be limited by the receiver phase imbalance [134].

Since no channelization is required in this transceiver, and since the LO signal is non-linear and narrow-band, typically centered within the 71-76 GHz or 81-86 GHz bands, the bandwidth requirements on the LO path are relaxed while those for the RF path remain stringent. Therefore, the architecture of figure 4.3a is adopted here along with an active, in-phase, RF signal splitting technique that, at least theoretically, has unlimited bandwidth, as discussed in section 4.3.3.

4.3. Circuit Design

4.3.1. Transmitter

The schematic of the QPSK modulator is depicted in Figure 4.4. It consists of two FET-HBT double-balanced Gilbert cells, driven in quadrature by the LO signal. Each Gilbert cell operates as an independent BPSK modulator, generating two BPSK constellations that are in quadrature with respect to each other. By summing the two constellations in current at the two 25 pH inductive loads, a QPSK constellation is formed. This summation procedure is illustrated graphically in Figure 4.5 and can be described in complex phasor notation by:

$$V_{\text{out}} = [(-1)^{B_I} + j(-1)^{B_Q}]G_m V_{\text{in}} Z_{\text{out}} \tag{4.4}$$

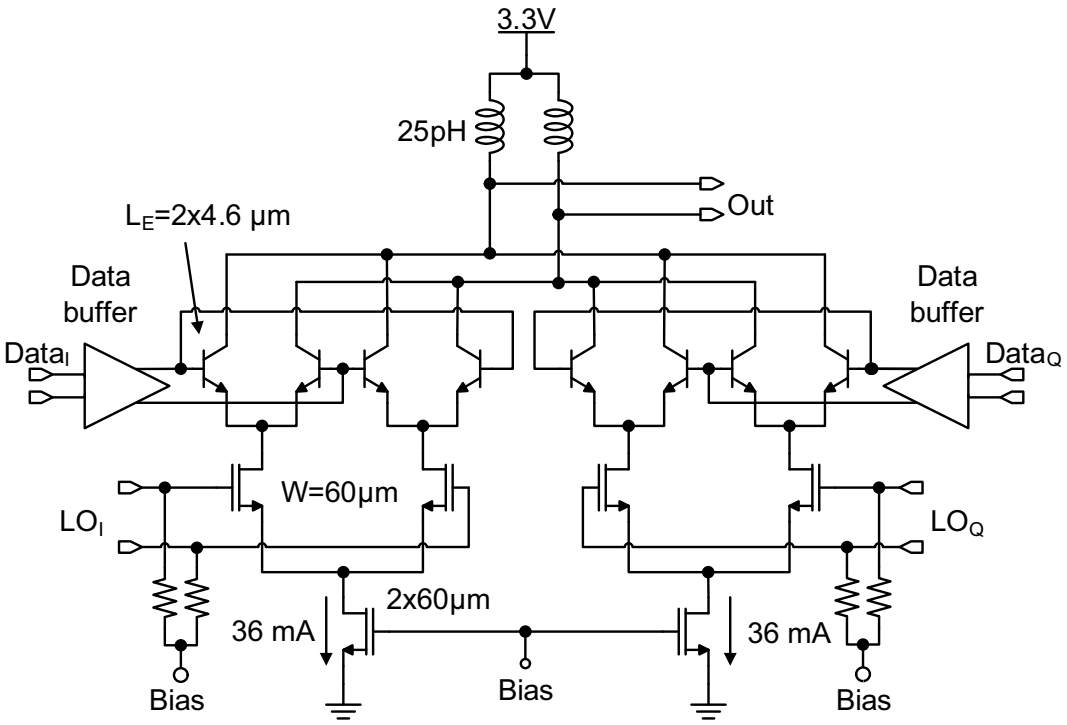


Figure 4.4: Schematic of the QPSK modulator.

B_I and B_Q are the I and Q data bits, assuming values of either 0 or 1, G_m is the effective large signal transconductance of the BiCMOS cascode differential pair and Z_{out} is the load impedance formed by the 25 pH inductors in parallel with the output impedance of the cells and the 50 Ω loads.

The 130-nm FETs on the LO path are biased at 0.3 mA/μmfor improved compression at high LO signal levels. The mixing quads are realized with SiGe HBTs to minimize the CML voltage swing applied at their bases. This scheme maximizes the data rate, the output voltage swing, and the output power of the transmitter.

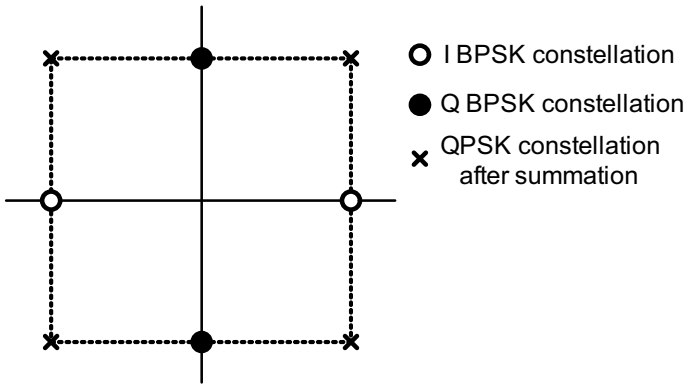


Figure 4.5: QPSK modulation constellation generation by two BPSK sub-constellations.

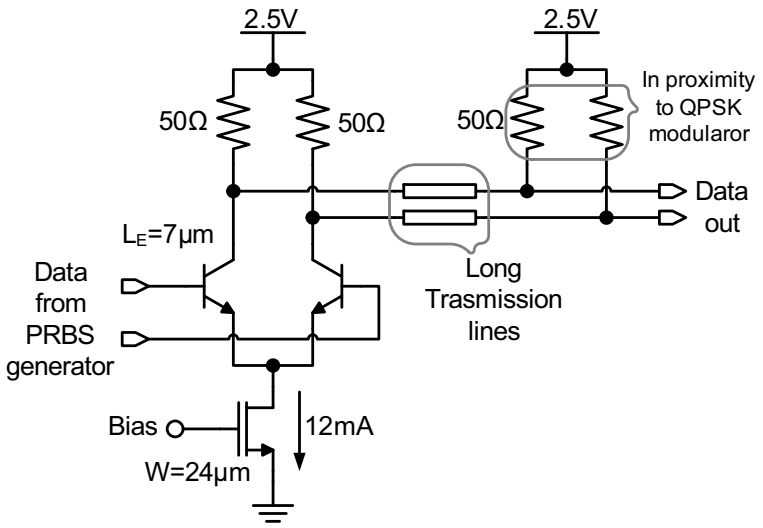


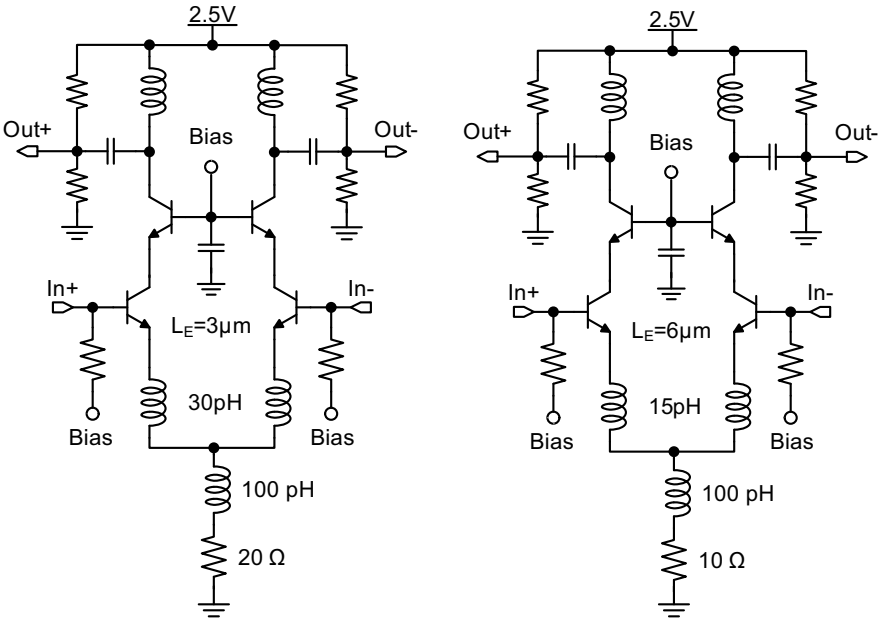
Figure 4.6: Schematic of the QPSK modulator data buffer.

Unlike in a conventional upconverter (e.g. [127]), the LO signal is applied to the bottom FET pair and the data signals are applied to the bases of the Gilbert-cell HBTs. Correct operation is thus ensured even when all transistors, including the FET pair, operate in non-linear, switching mode.

A constant output impedance Z_{out} is assured over all four switching conditions because the total DC current and the number of HBTs that are turned on or off and are connected at the output of the modulator remain constant over all data input states. This guarantees that the amplitude, and thus the transmitted power, of the output signal remains constant over all four transmitted symbols, irrespective of the LO signal amplitude. It allows the entire modulator to be operated with saturated LO signal, making it more robust to temperature, process and supply voltage variation.

Because the pole at the drain of the FET and the emitter of the HBT is pushed beyond the HBT $f_T/2$, the FET-HBT cascode provides over 10 dB of power gain at 80 GHz from a lower supply [135] while having higher stability margin than an HBT-HBT cascode. This is particularly important at large bias currents where an HBT cascode has a tendency to oscillate.

The stability of the QPSK modulator also critically depends on providing a low, non-inductive impedance at the bases of the HBTs which form the two Gilbert cells. This is problematic because those very nodes also have to switch at data rates higher than 10 Gb/s. Two special CML data buffers were employed to drive the Gilbert cells as illustrated in figure 4.4. Their schematic is reproduced in figure 4.6. In order to avoid any excessive inductive loading of the bases of the Gilbert quad transistors because of long interconnect, necessary in the physical layout of the circuit, a technique described in [136] was employed.



(a) Buffer the input of the hybrids.

(b) Buffer at the output of the hybrids.

Figure 4.7: Schematics of the LO buffers.

The data buffers are placed physically apart from the modulator while the resistive load pair is placed close to the modulator, at the far end of the long transmission lines connecting the differential HBT pair in the buffer to the Gilbert quads.

The QPSK modulator consumes 72 mA from 3.3 V power supply while the 4×20 Gb/s, $2^7 - 1$ PRBS generator draws 153 mA from 2.5 V power supply, including all the data and clock buffers. An on-chip PRBS generator solution was preferred, compared to providing the two data streams externally, in order to facilitate high-data-rate on-wafer probing and to minimize the number of high-speed pads required by the transceiver.

4.3.2. LO distribution and I-Q splitting

After single-ended-to-differential conversion, matching to 50 Ω and buffering, the LO signal is split into two equal parts by driving two buffers of identical size with the input buffer, in parallel (figure 4.2). This fan-out-of-two splitting is necessary in order to distribute the LO signal simultaneously to the transmitter and to the receiver.

The schematic of the first three buffers described above is shown in figure 4.7a. They employ a differential HBT-cascode topology with common-mode inductors for common-mode rejection, and common-mode resistors for bias current stability. The emitter degeneration inductors provide a real component to the input impedance of the buffer. Common-mode rejection is critical due to the inevitable common-mode feed-through of the input balun

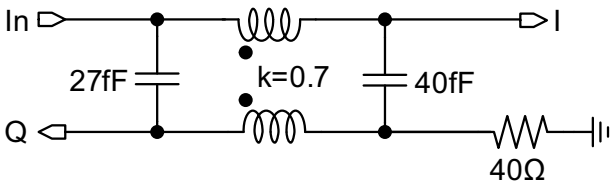


Figure 4.8: Schematic of the lumped quadrature hybrid.

employed for single-ended-to-differential conversion. The resistive dividers at the output of the cascode buffers serve a dual purpose. First, they generate the bias voltage for the subsequent stage. Second, since the HBT-only cascode is a conditionally stable amplifier below 100 GHz, the equivalent resistance generated at their outputs by the divider helps to increase the output conductance and hence stabilize the circuit (i.e. increase g_{22} in equation 2.5).

The quadrature splitting of the LO signal is realized with a lumped 90-degree hybrid whose schematic is shown in figure 4.8 [137, 138]. The hybrid consists of a $17 \times 17\,\mu\text{m}^2$ side-coupled transformer realized in the top metal layer of process, two MiM capacitors, and a $40\,\Omega$ termination resistor.

In order to generate differential quadrature signals, two quadrature hybrids are utilized as illustrated in figure 4.2. The first hybrid splits the 0° LO phase into two signals which are 90° and 180° out of phase with respect to its input. The second hybrid splits the 180° LO phase into two additional signals that are 270° and 360° out of phase with respect to the 0° LO phase. The 90° and 270° phase signals form one differential pair, while those with 180° and 360° phases form the second differential pair.

Following the hybrids, the LO signal is buffered and further amplified by another pair of differential HBT cascodes whose schematic is reproduced in figure 4.7b. These buffers are 2× scaled versions of the input buffers (figure 4.7a), as needed to generate the large output swing to drive the mixer and QPSK modulator.

The simulated gain and output compression point per side of the LO distribution network and modulator are 14 dB and 3.5 dBm respectively.

4.3.3. Quadrature Receiver

The schematic of the quadrature receiver is illustrated in figure 4.9 and was designed along with Sean Nicolson. The three-stage LNA features two common-emitter and a cascode stage. The common-base section of the cascode is composed of two equally-sized transistors, as shown in figure 4.9. The current from the common-emitter section is thus divided equally between the two common-base transistors at the middle, low-impedance node, achieving in-phase power splitting of the RF signal among the two mixers.

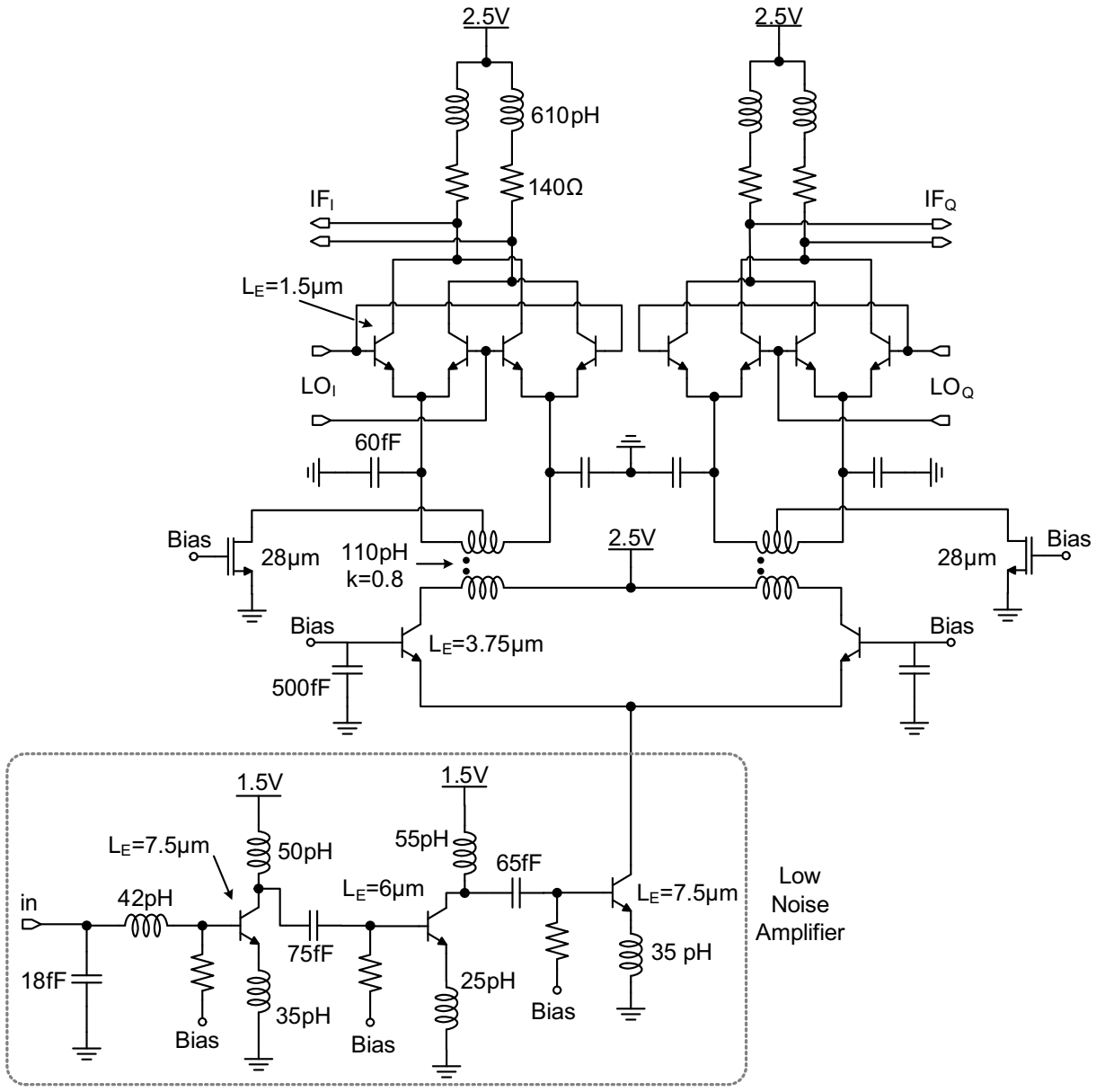


Figure 4.9: Schematics of the LNA and quadrature mixer.

Table 4.2: Power consumption of circuit blocks.

Block	Supply Voltage (V)	Power Consumption (mW)
QPSK Modulator	3.3	238
PRBS generator	2.5	383
Transmitter total		621
LNA and I-Q mixers	1.5 & 2.5	94
Baseband amplifiers	3.3	330
Receiver Total		424
LO Distribution Total	2.5	173
Total		1220

This active power-splitting technique has, in theory, no limit in its bandwidth of operation. If well-matched, the two common-base transistors have identical behavior over frequency and the current from the common-emitter stage will divide equally, regardless of the frequency of the RF signal. The gain of each of the two common-emitter - common-base signal paths is, to a first order, precisely 3 dB lower than that of the corresponding combined cascode stage.

Each of the two common-base transistors of the active cascode power splitter is further loaded with a transformer that converts the signal to differential mode and drives the mixing quad that performs the Zero-IF downconversion. The transformers along with the 60-fF shunt capacitors provide conjugate matching from the output of the cascode to the input of the mixing quads. The simulated LNA gain is 17 dB and the total current drawn by the LNA and the two mixers is 31 mA from 2.5 V, and 11 mA from 1.5 V including the bias circuits.

The IF outputs of the two mixers drive two, five-stage, variable gain amplifiers. The simulated 3-dB bandwidth of the baseband amplifiers is 8 GHz and the group delay varies by less than 9 ps from 0 to 10 GHz. The peak gain is larger than 30 dB, while drawing 50 mA from a 3.3 V power supply.

Table 4.2 lists the contribution of each circuit block to the power consumption of the transceiver. A significant portion of DC power is consumed in the 50 Ω interfaces of the baseband amplifier and in the PRBS generator, which were included solely for testing purposes and can be omitted in the actual system. Additionally, the power consumed by the LO distribution network could also be reduced by replacing the cascode buffers with simple common-emitter amplifiers from 1.2 V, as in the 122 GHz transceiver of chapter 6. However, the LO tree in this work was over-designed for the case that the power provided by the external LO signal source was lower than expected.

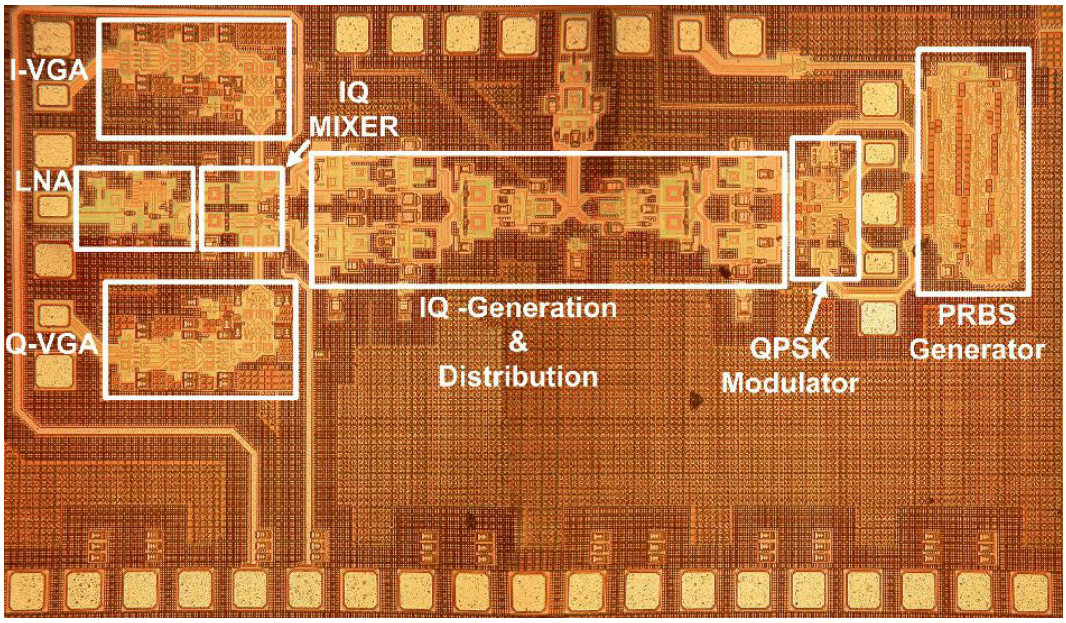


Figure 4.10: Transceiver die microphotograph.

4.4. Measurement Results

The transceiver was manufactured in STMicroelectronics' 130-nm SiGe BiCMOS process. The die photo of the transceiver chip is shown in figure 4.10. It occupies an area of 1.9 mm × 1.1 mm, including all pads. A W-band multiplier whose output power and phase noise are approximately -3 dBm and -90 dBc/Hz at 100 kHz offset, respectively, was employed as the external LO source.

A detailed description of the W-band measurement setups for S-parameters, noise figure, conversion gain and linearity can be found in [2, 139]

4.4.1. Quadrature Hybrid

The simulated and measured phase difference and amplitude imbalance of the two outputs of the hybrid are reproduced in figure 4.11. At 77 GHz, the phase error remains below 4° while the amplitude imbalance is less than 0.5 dB. The discrepancies between measurement and simulation, less than 5° and 1.2 dB in the 60-94 GHz range, are attributed to inaccuracies in transformer modeling, especially in the inter-winding capacitance.

4.4.2. Receiver

The receiver was characterized on-wafer by employing both noise and large signal measurement setups. Figure 4.12a and figure 4.12b show the measured and simulated receiver downconversion gain and double-sideband noise figure versus IF frequency when the LO

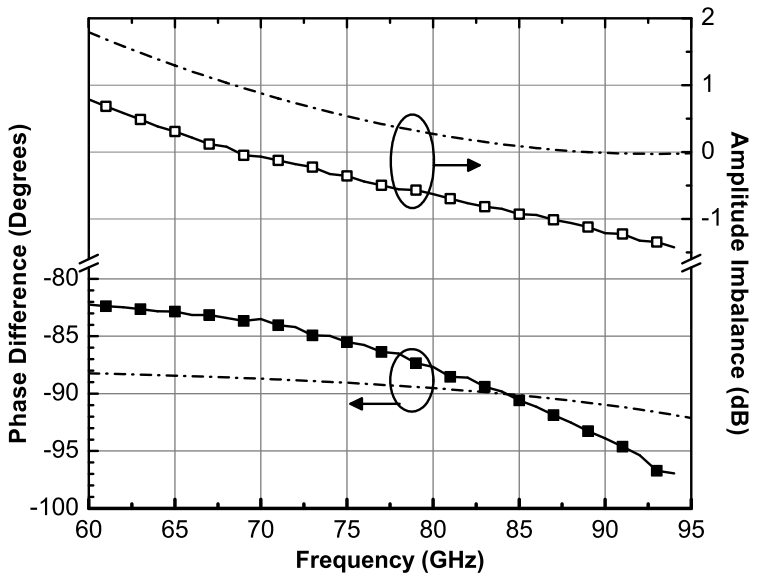


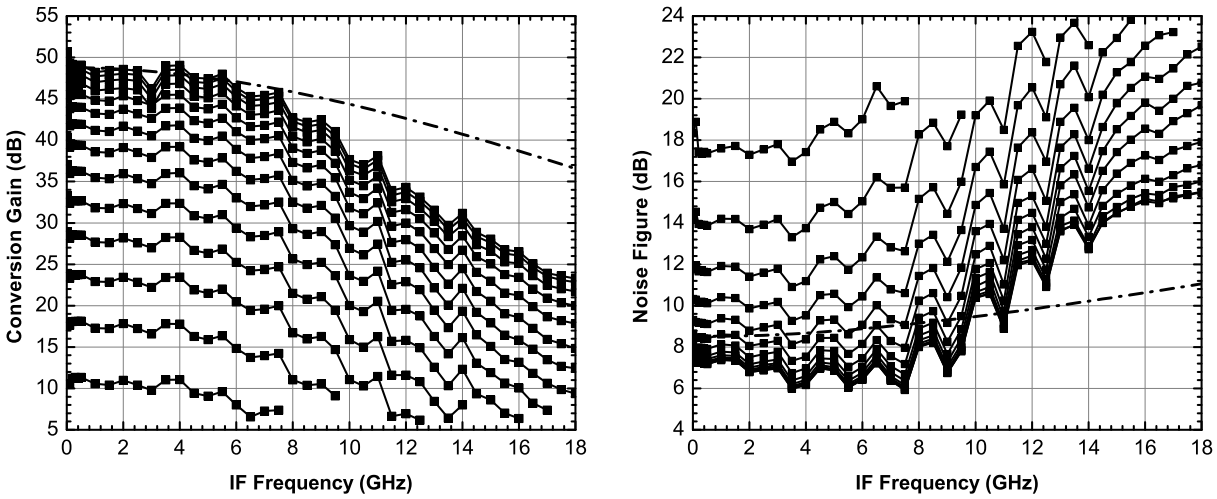
Figure 4.11: Simulated and measured phase difference and amplitude imbalance.

frequency is set to 76 GHz. The receiver gain is adjustable between 10 and 50 dB while the noise figure reaches a minimum value of approximately 7 dB.

The difference between the measured and simulated bandwidth roll-off in both gain and noise figure at high IF frequencies is due to the increased loss of the measurement setup in the two RF sidebands [140]. The lower sideband is subjected to higher loss due to the cut-off frequency of the W-band coaxial-to-waveguide transition, while the upper sideband loss increases due to the naturally increasing loss of the passive components used to connect the W-band noise source to the circuit. Because the gain of the two sidebands cannot be independently measured, it is impossible to accurately deembeded the loss at RF frequencies that are increasingly different from the LO frequency.

Figure 4.13 illustrates the simulated and measured noise figure as a function of receiver gain for the LO frequency at 76 GHz and the IF frequency at 2 GHz. Since a large portion of the downconversion gain is obtained in the front-end consisting of the LNA and the mixer, the receiver noise figure degrades only marginally as the gain is reduced from 50 dB to 30 dB. The discrepancy between the simulated and measured noise figure is attributed to the fact that the transistor model does not include noise correlation between the collector and base noise currents, resulting in pessimistic noise figure values [1].

Figure 4.14a reproduces the measured and simulated amplitude and phase imbalance between the two baseband outputs of the receiver. In this measurement, the baseband amplifier gain was reduced and significant attenuation was introduced with a variable attenuator placed at the input of the receiver, making sure that the receiver was operated in linear mode.



(a) Downconversion gain.

(b) Noise figure.

Figure 4.12: Measured and simulated downconversion gain of noise figure of the receiver with different gain settings in the baseband amplifier. The simulated curves corresponds to the maximum gain setting.

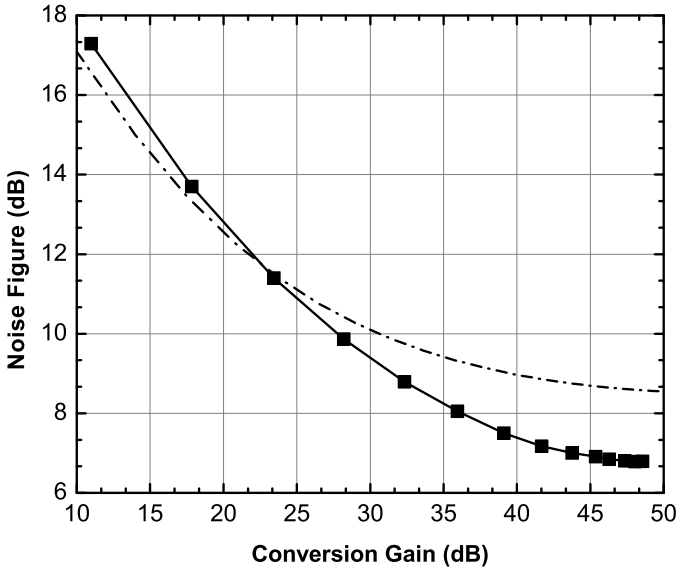
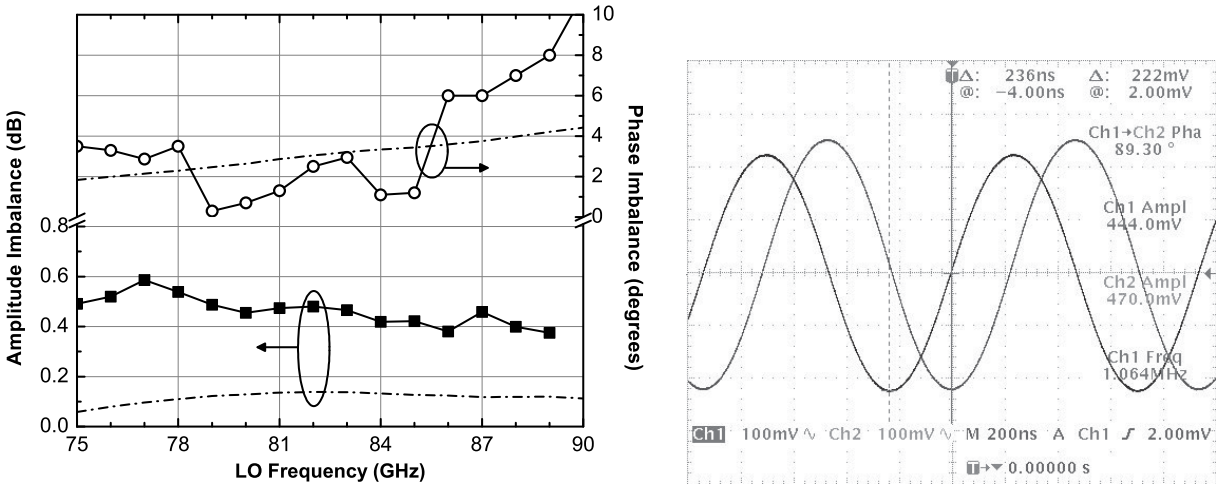


Figure 4.13: Transceiver die micro-photograph.



(a) Measured and simulated I-Q amplitude and phase imbalance of the receiver versus LO frequency. The IF frequency is held constant at 1 MHz.

(b) I and Q IF outputs versus time when f_{LO}=80 GHz and f_{IF}=1 MHz.

Figure 4.14: Measured and simulated receiver I-Q imbalance.

Furthermore, the RF and LO frequencies were simultaneously varied in order to keep the IF frequency constant at 1 MHz during the sweep. This sweep methodology ensures that the frequency variation of the different IF components (i.e. cables, connectors, etc.) does not affect the accuracy of I-Q imbalance measurement. The measured amplitude imbalance is lower than 0.6 dB from 75 to 90 GHz while the phase error remains below 4° from 75 to 85 GHz. Figure 4.14b reproduces the waveforms at the two baseband outputs versus time corresponding to an f_{LO} of 80 GHz.

The slight amplitude imbalance observed between the two baseband outputs is mainly attributed to mismatches between the two baseband amplifiers and can be compensated by separately adjusting the gain of each amplifier.

The receiver IQ imbalance measurement demonstrates the proper, broadband operation of the quadrature hybrid and of the in-phase power splitter in the output stage of the LNA. A large phase imbalance in the quadrature receiver would result in improper QPSK signal demodulation and would require expensive receiver equalization, especially as the bit rate is increased.

Figure 4.15 depicts the simulated and measured downconversion gain versus the RF input signal power. The gain of the baseband amplifiers was once again reduced during the measurement, setting the overall downconversion gain to approximately 20 dB. Under these conditions, the IP1dB is approximately -23 dBm. The gain reduction was necessary because, at maximum gain, the attenuation required at the RF input for the receiver to operate linearly is outside the range of the available mm-wave attenuator. The receiver

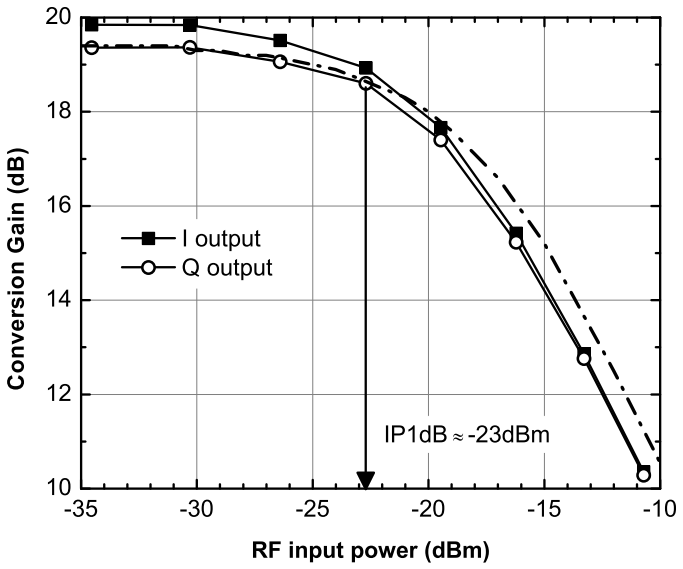


Figure 4.15: Measured and simulated input 1 dB compression point (IP1dB) when the conversion gain is set to 20 dB and the LO and IF frequencies are 76 GHz and 1 MHz respectively.

input compression point at maximum gain was simulated to be -51 dBm.

4.4.3. Transmitter

A transmitter output power of +6 dBm per side was observed in the 60-70 GHz range for an un-modulated carrier.

Figure 4.16 illustrates the measured transmitter output spectrum for an LO signal at 80 GHz with an aggregate QPSK data rate of 400 Mb/s. The lobes of the QPSK spectrum are visible at low bit rates. Image folding from the external double-sideband harmonic mixer used with the spectrum analyzer makes the spectrum more difficult to interpret at higher rates. As a result, at higher bit-rates, the receiver baseband spectrum and baseband eye diagrams were employed instead to verify correct operation.

4.4.4. Loop-back

A loop-back test was conducted with the transmitter connected to the receiver through a high loss (>20 dB) cable and monitoring the I and Q baseband outputs simultaneously with a Bit Error Rate Tester (BERT), an oscilloscope, and a spectrum analyzer. The LO frequency was set to 77 GHz.

First, the I and Q baseband outputs of the receiver were connected to the BERT. Data transmission with a BER smaller than 10^{-12} was verified at both the I and Q outputs for a QPSK bit rate of 5.4 Gb/s (2.7 Gb/s per channel). Figure 4.17a reproduces the measured eye opening for a BER of 10^{-10}, while figure 4.17b illustrates the measured BER versus clock

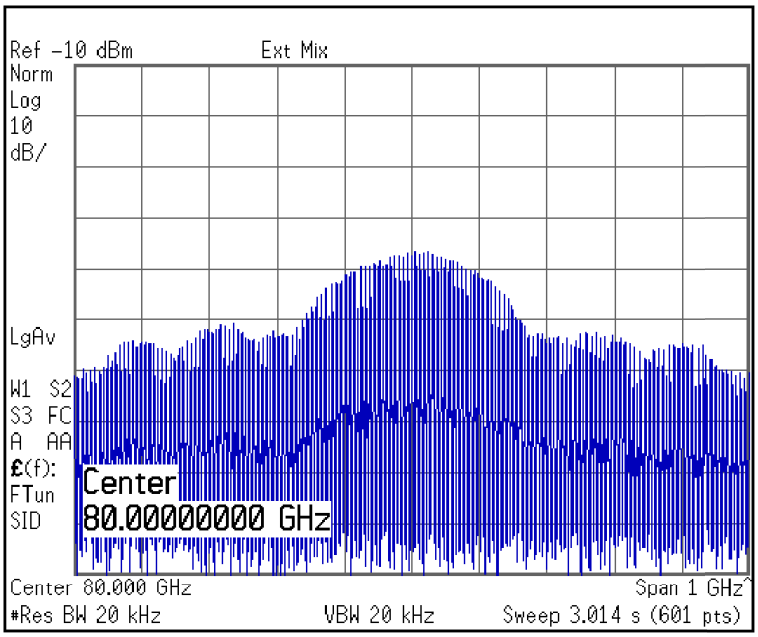


Figure 4.16: Measured 400-Mb/s QPSK spectrum at the 80-GHz output of the transmitter.

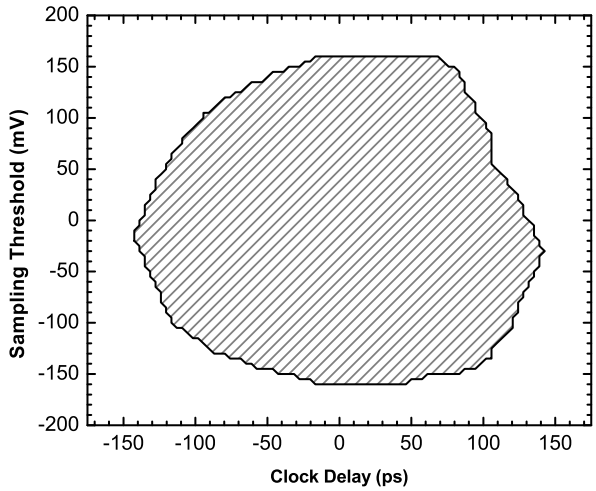
timing.

Figure 4.17c shows the eye diagrams at the I and Q outputs of the receiver at 5.4 Gb/s (2.7 Gb/s per channel) bit rate. It can be seen that the I and Q outputs are almost in perfect quadrature, although the mismatch in probes and test cables has not been calibrated out. Figure 4.17d depicts the received spectrum at the same data rate. The location of the null of the main lobe of the spectrum, as well as the tone spacing within the main lobe, clearly indicate that the received signals are correct $2^7 - 1$ PRBS sequences.

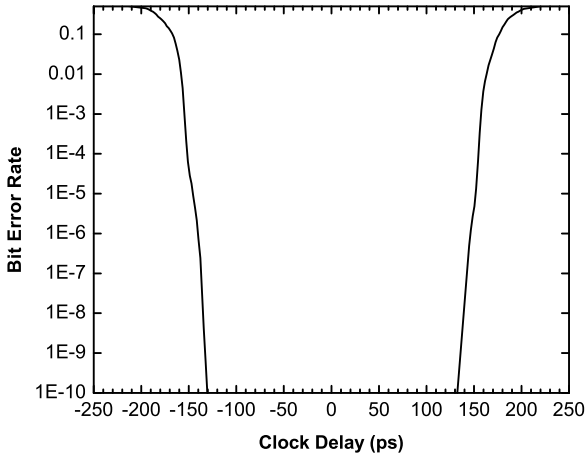
However, 5.4-Gb/s (2.7 Gb/s per channel) is the highest rate at which bit error rate measurements can be conducted with the available BERT. Testing above this rate was limited to inspection of the baseband eye diagrams and spectra, and to capturing the received data sequence using the pattern lock feature of the oscilloscope, and manually verifying that the data sequence corresponds to a $2^7 - 1$ PRBS sequence.

Figure 4.18a reproduces the measured eye diagrams when the bit rate is increased to 16 Gb/s (8 Gb/s per channel), while figure 4.18b shows the measured received spectrum at the same rate. Similarly, figures 4.19a and 4.19b reproduce the received eye diagrams and spectrum respectively at 18 Gb/s (9 Gb/s per channel). Finally, figure 4.20 illustrates the captured data sequences at the I and Q outputs of the receiver, compared with an ideal $2^7 - 1$ PRBS sequence. Error free operation was achieved up to 18 Gb/s QPSK bit rate.

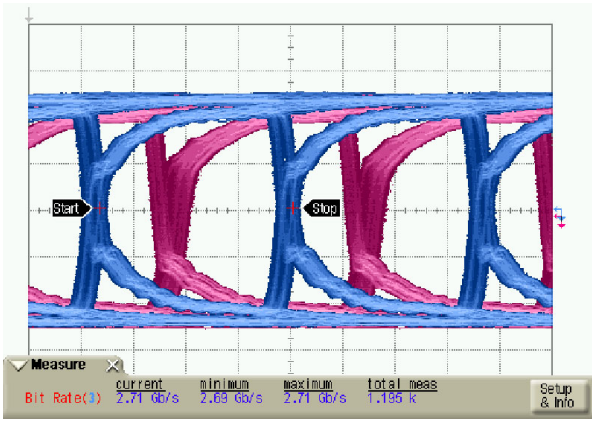
The slight amplitude asymmetry between the amplitude of the I and Q outputs in figure 4.20b are attributed mainly to limitation of the experimental setup. Specifically, The I



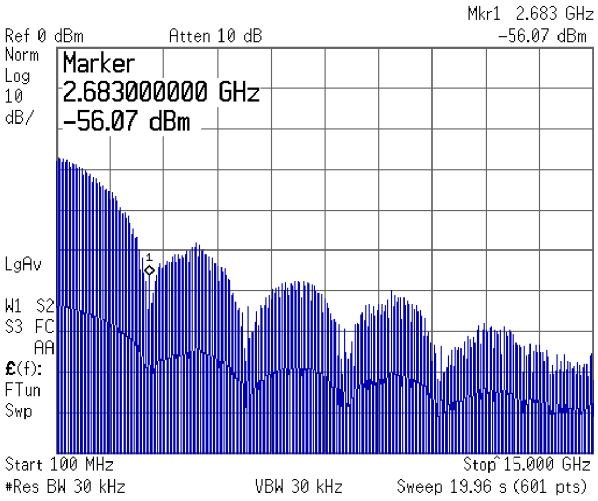
(a) Measured eye opening. The bit error rate (BER) inside the shaded region is better than 10^{-10}.



(b) Measured BER versus clock timing.

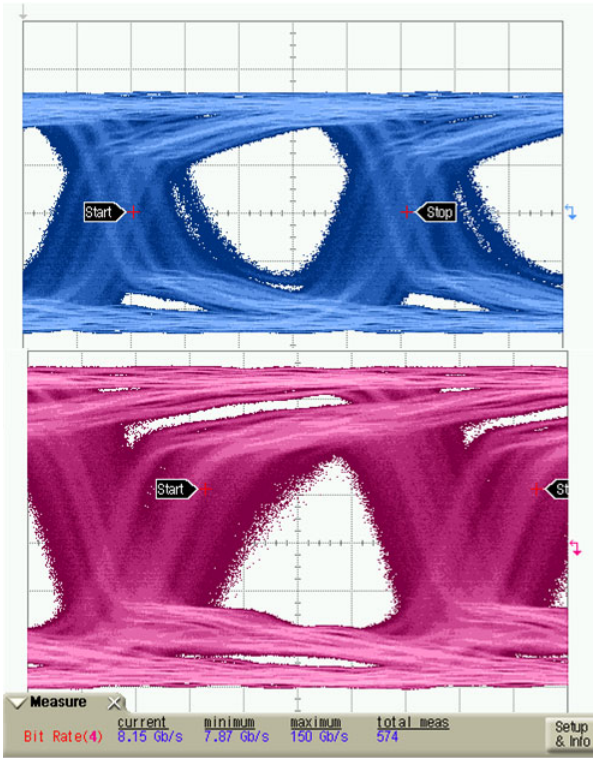


(c) Measured eye diagrams at both the I and Q outputs corresponding to 2.7 Gb/s per I and Q channels.

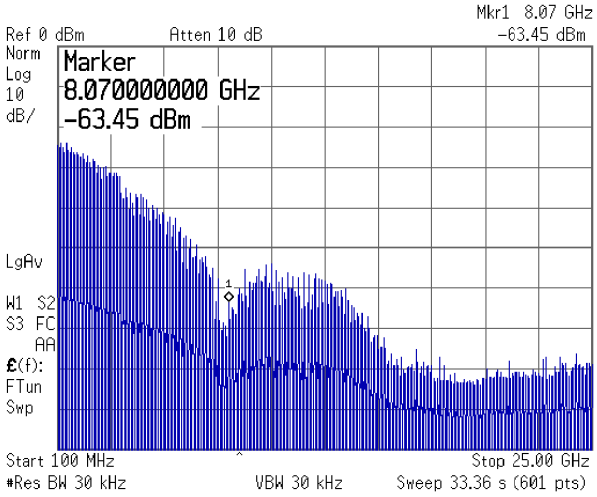


(d) Measured baseband spectrum.

Figure 4.17: Loopback measurements at 5.4-Gb/s QPSK data rate.

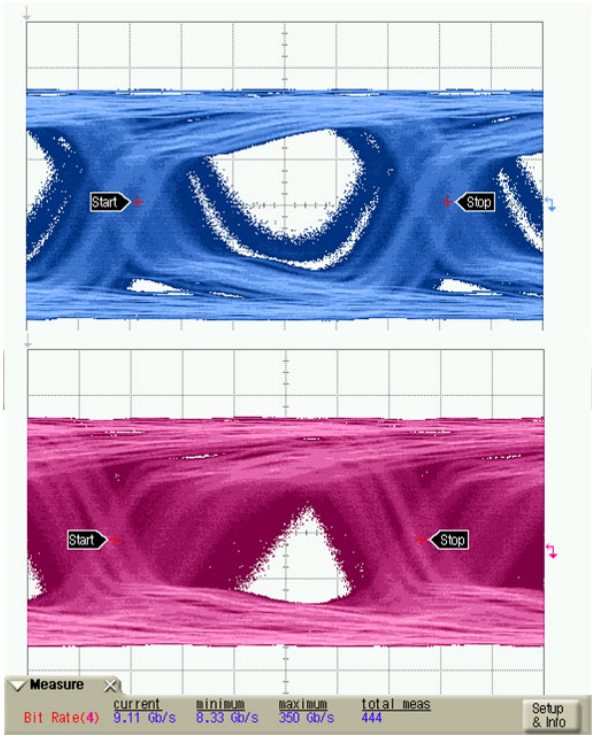


(a) Measured eye diagrams at both the I and Q outputs.

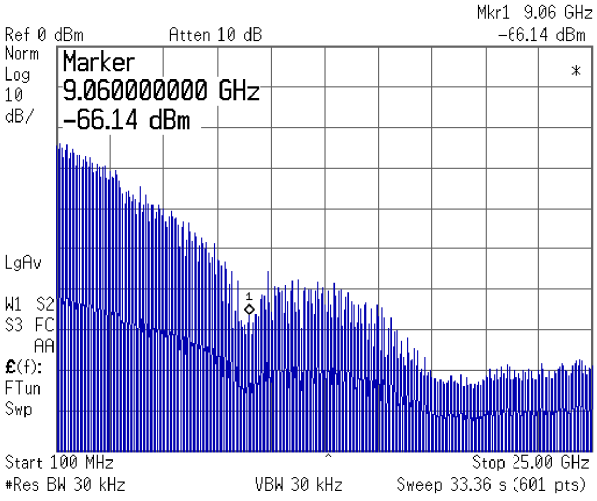


(b) Measured baseband spectrum.

Figure 4.18: Loopback measurements at 16-Gb/s QPSK data rate.

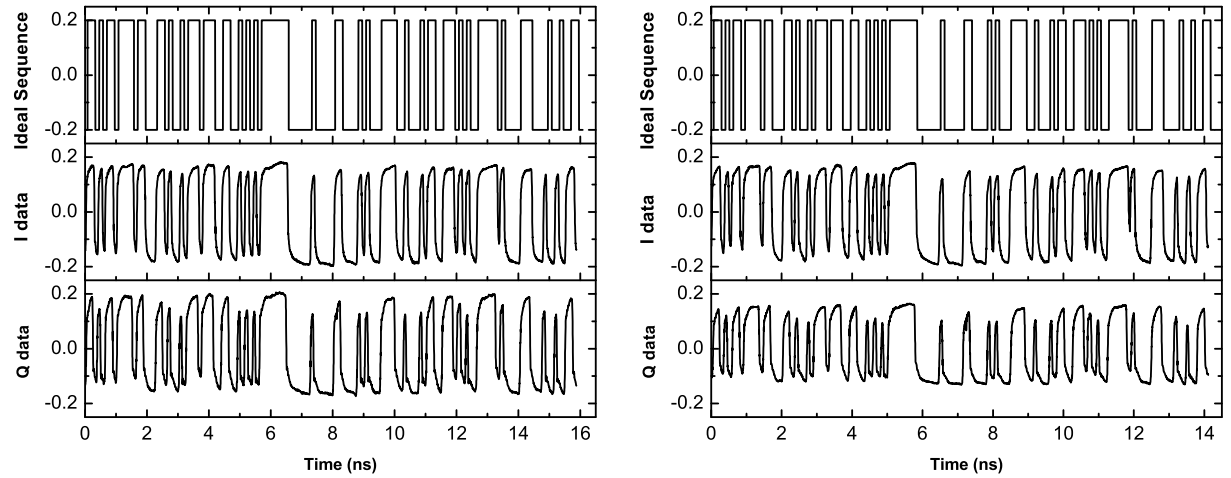


(a) Measured eye diagrams at both the I and Q outputs.



(b) Measured baseband spectrum.

Figure 4.19: Loopback measurements at 18-Gb/s QPSK data rate.



(a) Data sequence at 16-Gb/s QPSK data rate.

(b) Data sequence at 18-Gb/s QPSK data rate.

Figure 4.20: Captured data sequences at the I and Q outputs versus an ideal $2^7 - 1$ bit PRBS sequence.

and Q two output signals are contacted with a wafer-probe that has three signal pins and are connected to the oscilloscope with two different cables. Differences in the contact resistance of the two probe needles as well as the different attenuations of the two cables result in different amplitudes between the two channels, which are more intense in the 18 Gb/scase.

At bit rates above 18 Gb/s, the received eye diagrams begin to close and errors emerge in the received data sequences. Because of these errors, the oscilloscope cannot lock to the received pattern and the number of errors cannot be assessed. It is assumed that the demodulated signal is affected by the channel delay and the frequency response of the receiver. Some form of equalization will be necessary for operation at higher rates. This can be done in an analog fashion, without having to invoke complicated and power consuming data converters, as demonstrated in [141] and in [142] up to 80 Gb/s.

4.5. A 60 GHz Direct QPSK Modulator

Although the E-band QPSK modulator was verified to operate at very high bitrates, its output power and efficiency remained relatively low (2%). In an effort to investigate how to ameliorate these shortcomings, the 60 GHz QPSK modulator of figure 4.21 was designed and fabricated in the same SiGe technology. Several improvements were implemented compared to the circuit of figure 4.4. The current sources at the tail of each BPSK cell were removed and the supply voltage was simultaneously reduced to 1.8 V, allowing to boost the efficiency. Furthermore, the two L-matching networks at the output were replaced with a transformer. The transformer performs an impedance transformation from the 50 Ω load to R_{opt}, thus increasing the output power, along with converting the differential signal to single-ended.

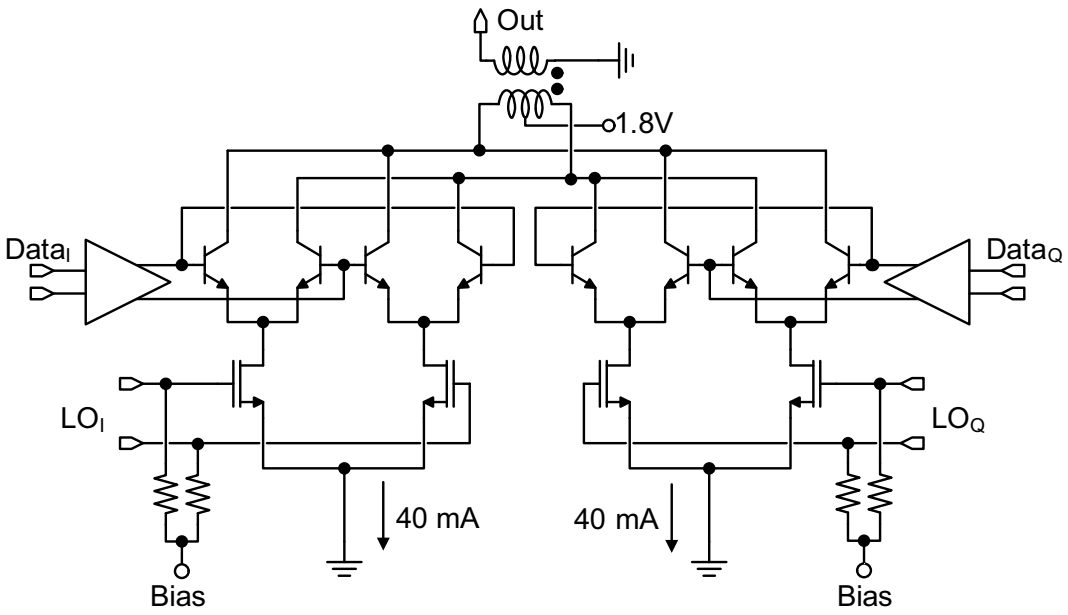


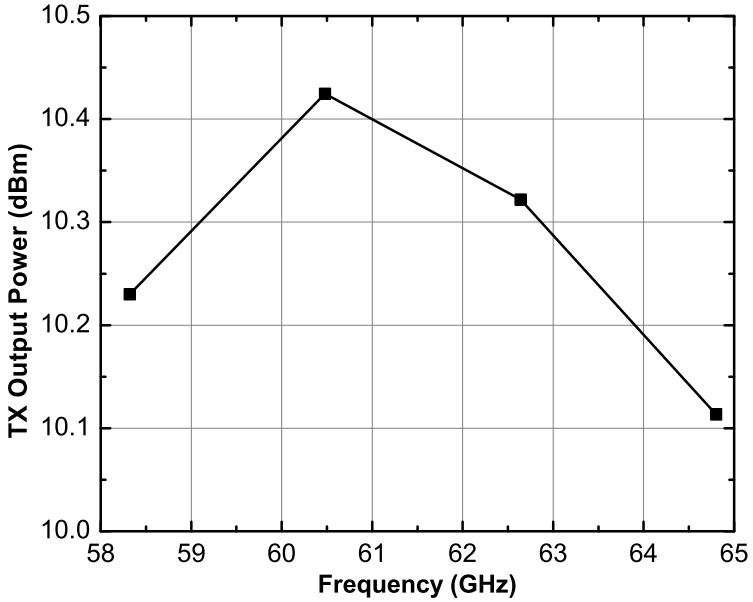
Figure 4.21: Schematic of the 60 GHz QPSK modulator.

The 60 GHz band was selected primarily due to the availability of a single-sideband spectrum analyzer in this frequency range. If this instrument is unavailable, it becomes impossible to test the spectrum at high bitrates, as it will get obscured by double sideband folding like the one in figure 4.22b, and it would be impossible to verify the proper operation of the standalone circuit without a single-sideband receiver.

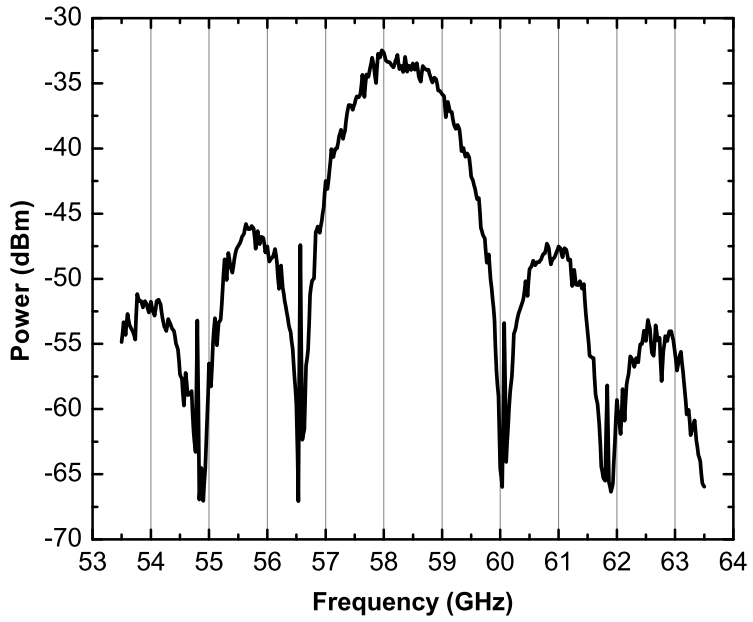
Figure 4.22a illustrates the measured output power of the 60 GHz QPSK modulator across the four channels of the IEEE802.15c 60-GHz radio communication standard. The circuit always operates as a saturated amplifier and as a result, the output power is constant over a very wide frequency range with a drain efficiency of 7%. Figure 4.22b illustrates the output spectrum of the modulator at 1.8 Gb/s aggregate bitrate, indicating that the proper formation of the QPSK modulation. Correct, error free operation was verified up to at least 5.3 Gb/s using BER measurements in [143].

4.6. Summary

A transceiver for last-mile, point-to-point wireless links in the 70-80 GHz band was presented which is capable of data rates up to 18 Gb/s in loop-back configuration. This record high bit rate was made possible by several important transceiver architecture and design choices. First, a simple QPSK modulation scheme was selected which, although somewhat spectrally inefficient, allows for a simplified, saturated power, broadband transmitter design. The transmitter architecture includes a novel direct mm-wave QPSK modulator which can



(a) Output power versus frequency.



(b) QPSK spectrum at 1.8 Gb/s.

Figure 4.22: Measured spectrum and output power of the 60 GHz QPSK modulator.

generate adequate output power for this application, avoiding the use of a broadband linear upconverter.

Furthermore, careful design of the quadrature receiver for precise broadband in-phase and quadrature power splitting of the RF and LO signals between the I and Q mixers resulted in low phase imbalance over a wide frequency range and, as a result, proper reception of record high data-rate signals.

The QPSK modulator circuit was further refined in a 60 GHz version that features higher output power and efficiency, and is able to operate up to at least 5.3 Gb/s.

5 System Design for a mm-wave Distance Sensor

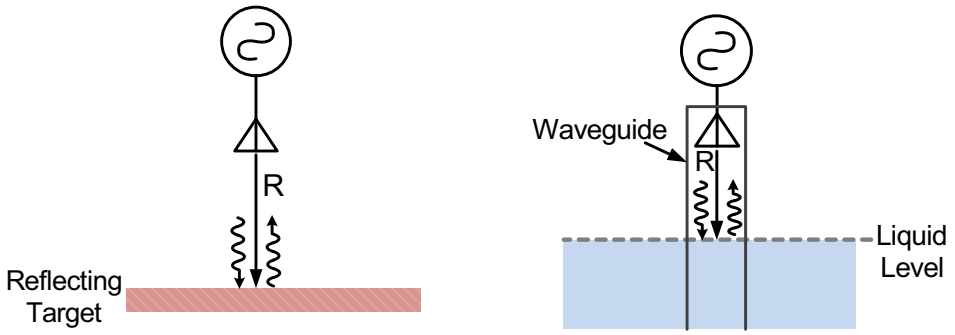
This chapter reviews the system level details associated with the implementation of a mm-wave distance sensor. The existing methods and architectures will be analyzed in terms of the resulting system precision and new improved architectures will be proposed, which will improve on the random and systematic errors of the system.

The chapter proceeds as follows. In sections 5.2 and 5.3 the methods for microwave range-finding will be reviewed and compared. In sections 5.4 and 5.5, the performance limiting factors will be analyzed and the system design parameters will be linked to the maximum attainable accuracy. Finally, self-calibration methodologies will be presented in section 5.6. Appendix 5.A reviews some necessary basic concepts of signal estimation in the presence of noise.

5.1. Application Description

In recent years, regulatory bodies have proposed the use of systems that provide new safety functionalities in cars, such as collision avoidance, automatic cruise control, lane departure warning and blind spot detection. Specific radio bands, at 24 and 77 GHz were allocated for this purpose and mm-wave ICs have been developed. For example, Freescale in [8] and Infineon in [7] reported 77 GHz, 4× SiGe BiCMOS monolithic arrays for MIMO (Multiple In-Multiple Out) radar [144] that are already available in high-end cars.

This work will build upon the gained experience from automotive radar and attempt to address the application of non-contact, precise distance measurement from a large surface. As illustrated in figure 5.1, the surface can be the liquid in a tank for level measurement, a material in an industrial environment whose roughness needs to be monitored, a road under construction or even the blades of a steam turbine [145]. In the automotive area, where the market is pushing for the increasing integration of sensors in the car, such a system could be used for monitoring the roughness of the road and assist in adaptive car suspension. All



(a) Distance measurement from a reflector.

(b) Liquid level measurement with a waveguide

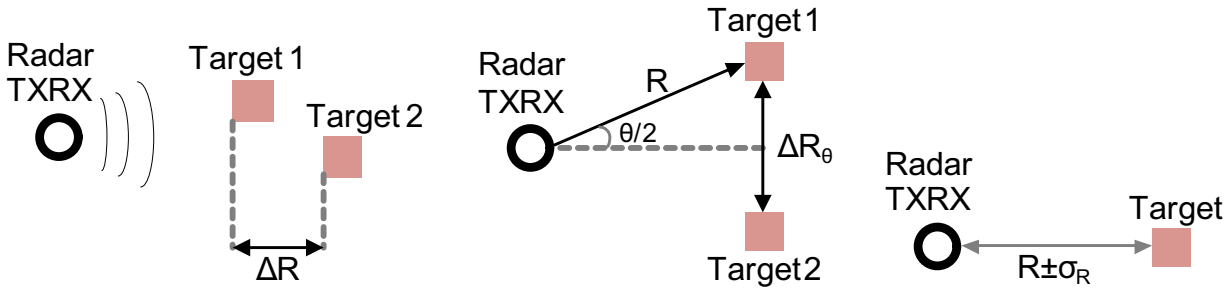
Figure 5.1: Distance measurement system under consideration.

these applications have the common requirement of distance measurement with accuracy better than 0.5 mm, which will be the primary focus of the system.

A number of non-contact distance sensors have been reported in the past, primarily based on lasers, but also on other technologies like ultrasound and microwaves that are less widespread. Lasers offer unparalleled range and accuracy [146, 147] but are very sensitive to low visibility atmospheric conditions such as dust, sand, fog and rain, rendering them unusable in a variety of applications, especially within automotive and industrial environments. Furthermore, an opaque material like mud, smoke or oil that builds up on the opening of the sensor can severely impair its operation. This would be highly undesirable in automotive environments where mud and dust are always present in the car exterior. Likewise with ultrasound, apart from the fact that their operation can be affected by the same environmental conditions as the lasers [148], their non-directive nature can cause nearby objects to interfere with the measurement.

Microwave and mm-wave sensors have the potential to overcome these shortcomings. They are immune to most atmospheric variations and may even penetrate thin materials like paper, wood and plastic. Several such implementations of microwave ranging systems have been proposed [145, 148–154]. However, they are all composed of discrete components and do not account for the difficulties that would arise in nanoscale silicon technologies, especially 1/f noise. Furthermore, the frequency of operation of the reported microwave systems is relatively low, less than 77 GHz, leading to large antenna sizes and low directivity, making them impractical for many applications when compared to laser systems.

At a higher frequency of operation, the transmitter output power decreases due to the inherent lower transistor power gain, which, in combination with the higher free space loss, will inevitably lead to smaller range. However, as will be shown below, this problem can be partially alleviated by employing higher gain antennas and for most of the aforementioned



(a) Range resolution. (b) Angular resolution. (c) Range and accuracy.

Figure 5.2: Important radar performance metrics.

applications, a range of approximately 3 m is adequate, and can be achieved with only moderate output power (<0 dBm).

5.2. Radar Performance Metrics

There are several parameters that quantify the performance of a radar system. The degree to which each parameter needs to be satisfied depends largely on the application. For example, the case of a singe large nearby target has vastly different requirements from that of an airborne radar that has to detect extremely fast moving objects at distances of several hundred kilometers. These differences will be explained below.

5.2.1. Range Resolution

The range resolution of a radar system, graphically illustrated in figure 5.2a, is the minimum *horizontal* distance ΔR between two targets that can be unambiguously separated. It depends on the bandwidth, B, of the radar waveform:

$$\Delta R \geq \frac{c}{2B} \tag{5.1}$$

where c is the speed of light. Interestingly, it is independent of the center frequency of operation, although there is a clear advantage in operating at higher frequencies due to the higher fractional bandwidths available.

The range resolution is especially important in automotive radars where closely spaced cars and motorcycles, bicycles or pedestrians need to be accurately resolved.

5.2.2. Angular Resolution

The angular resolution, depicted in figure 5.2b, represents the minimum *lateral* distance, ΔR_θ at which two targets at the same distance from the radar can be unambiguously separated. This parameter depends on the target distance R and the 3 dB beamwidth θ of the

radar transceiver antenna as follows [155]:

$$\Delta R_\theta \geq 2R \sin \frac{\theta}{2} \tag{5.2}$$

where the assumption is made that the power in the sidelobes of the antenna is low enough to be ignored.

For every antenna, the beamwidth depends on its physical diameter, d, and on the wavelength of operation, λ [156]:

$$\theta = k_\theta \frac{\lambda}{d} \tag{5.3}$$

where k_θ is a proportionality factor dependent on the antenna type. For example, when θ is measured in radians and the antenna is parabolic, $k_\theta = 1.2$. Substituting in (5.2),

$$\Delta R_\theta = 2R \sin \frac{k_\theta \lambda}{2d} = 2R \sin \frac{k_\theta c}{2fd} \tag{5.4}$$

As a result, at higher frequencies, better angular resolution can be obtained with constant antenna size.

The directivity, D of the antenna is also related to the beamwidth. Assuming an elliptical beam spot, D is expressed as follows [156]:

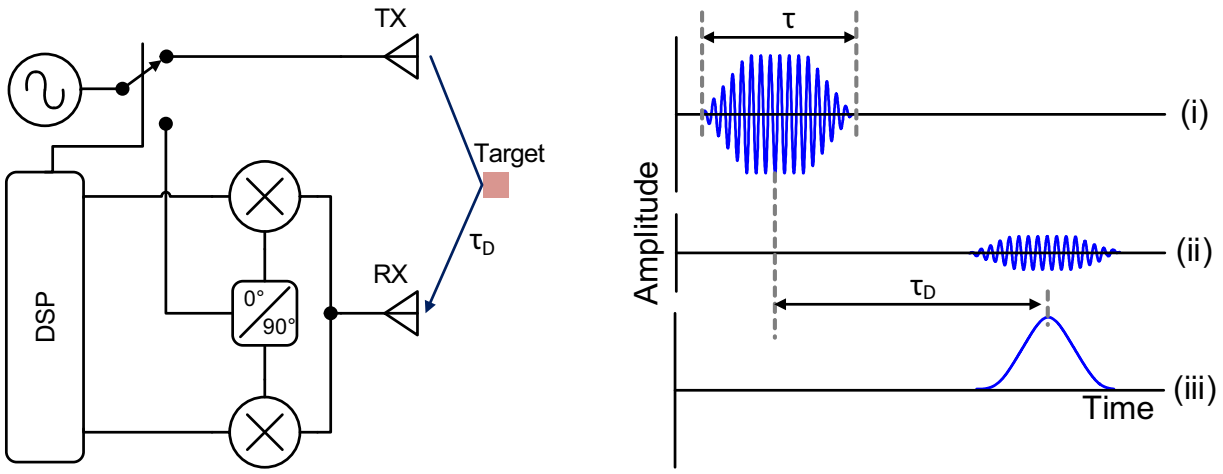
$$D = \frac{16}{\sin \theta_{az} \sin \theta_{el}} \simeq \frac{16}{\theta_{az} \theta_{el}} \tag{5.5}$$

where θ_{az} and θ_{el} is the azimuth and elevation beamwidths in radians. The angles θ_{az} and θ_{el} define the field-of-view (FOV) of the radar, i.e. the maximum angles inside which a target is visible. Assuming a circular spot size, $\theta = \theta_{az} = \theta_{el}$:

$$D = \frac{16}{\theta^2} \tag{5.6}$$

Consequently, the antenna directivity is inversely proportional to the square of the beamwidth, i.e, higher antenna gains will result in narrower beamwidths.

Typically, radar systems that aim for high angular resolution cannot simply use a high gain-low beamwidth antenna to resolve multiple targets. This would severely restrict their field-of-view and limit what the radar can "see". As a result, beam steering with phased arrays or MIMO radar with multiple beams are commonly employed [157].



(a) Generic block diagram.

(b) Waveforms.

Figure 5.3: Pulsed radar.

5.2.3. Range and Accuracy

The range of a radar, R_{max}, is the maximum distance where a target can be unambiguously detected. The range primarily depends on the link margin and the Signal-to-Noise Ratio (SNR) at the output of the receiver and consequently, on the available integration time between successive measurements.

The accuracy of the radar, graphically illustrated in figure 5.2c, is the average absolute error in the detected distance R. It is represented by the standard deviation, σ_R, and depends on the SNR, as well as, on systematic errors such as the non-idealities of the radar waveform. Obviously, due to the increasing free space loss with distance, the accuracy of a radar degrades with increasing distance to the target.

5.3. Radar Waveforms

This section presents the most common signal waveforms employed for detecting distance: pulsed, Frequency Modulated - Continuous Wave (FMCW) and Step Frequency Radar (SFR). Each type of radar essentially requires a different modulation in the transmitter and slightly different receiver architectures. The advantages and disadvantages of each method with regard to the range accuracy and system complexity will be reviewed.

5.3.1. Pulsed Radar

The pulsed radar is probably the most well known type among microwave engineers. It has been used extensively, mostly for military applications (e.g. for airborne radar). The block diagram of a coherent pulsed radar transceiver is illustrated in figure 5.3 [158, 159].

At every measurement cycle, the transmitter generates a pulse of duration τ, modulated at frequency f_c by the oscillator (LO) signal. Immediately after, the LO signal is switched to the receiver. The pulse is scattered by one or multiple targets and its echo arrives at the receiver input after time τ_D, as shown in figure 5.3b(ii). The round trip delay (also known as the time-of-flight, TOF) of the pulse is:

$$\tau_D = \frac{2R}{c} \tag{5.7}$$

where R is the distance to the target and c is the speed of light.

The pulse is coherently downconverted (i.e. with the same LO signal used in the transmitter) and digitized by a signal processor that applies a matched filter, resulting in the waveform of figure 5.3b(iii). The matched filter exhibits the optimum performance under noise and aids in the calculation of the delay τ_D. The cycle is then repeated with frequency f_{rep}.

If two pulses reflected by different targets overlap at the receiver input, then it is impossible to distinguish between them. In order to avoid this adverse situation, the pulse duration must be sufficiently small, i.e. $\tau < 2\Delta R/c$ and consequently:

$$\Delta R \geq \frac{c\tau}{2} \tag{5.8}$$

The Nyquist criterion for bandlimited pulses requires that at least a bandwidth of $1/\tau$ is necessary to represent a pulse. Substituting in the above equation:

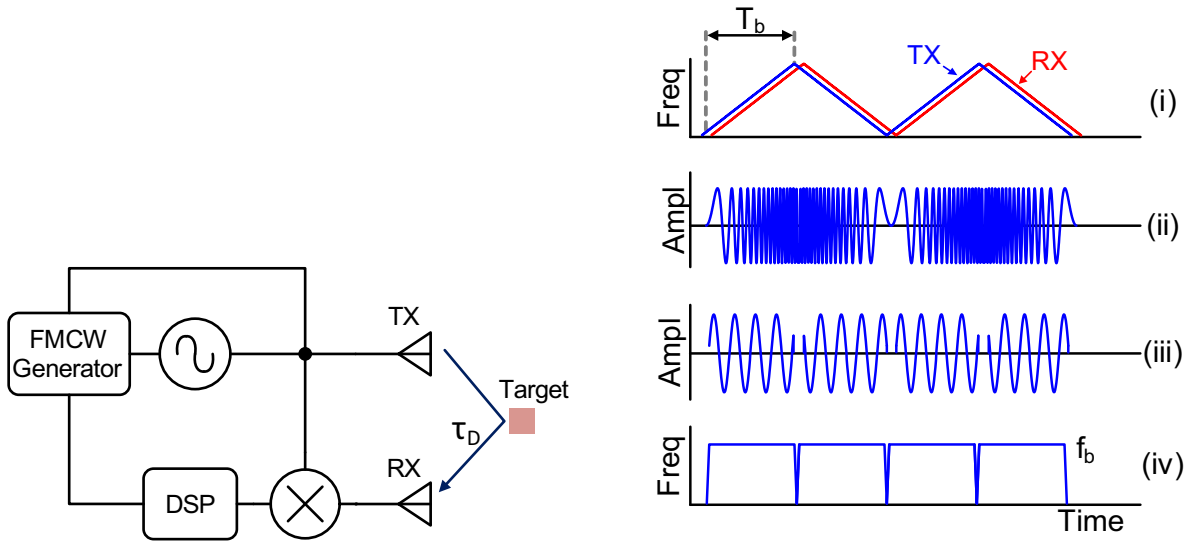
$$\Delta R \geq \frac{c}{2B} \tag{5.9}$$

which affirms the validity of equation (5.1) for the radar range resolution.

Another important consideration stems from the pulse repetition rate f_{rep}. If a target is far, the round trip delay of the pulse will be large enough to appear at the receiver during the next cycle, resulting in range ambiguity. Assuming that $\tau \ll 1/f_{rep}$, the following condition must be satisfied to avoid this condition:

$$R_{max} \leq \frac{c}{2f_{rep}} \tag{5.10}$$

The above equation leads to a trade-off. Due to the presence of noise in the received signal, several pulses from different cycles need to be averaged in order to improve the SNR and thus the range accuracy. However, since f_{rep} is generally small to avoid ambiguities, the averaging process will require more time, increasing the overall measurement time. This trade-off can



(a) Generic block diagram. (b) Waveforms.

Figure 5.4: FMCW Radar.

be resolved by increasing the transmitter power, and not the measurement time, in order to improve the SNR. Nevertheless high power transmitters are especially difficult to realize at high frequencies.

Another disadvantage of the pulsed radar is that it suffers from synchronization problems, which are common to all time division multiplexing systems. As can be seen in figure 5.3, any unwanted delay in the generation of the TX pulse will directly affect the range accuracy of the system.

Apart from its simplicity, the main reason why pulsed radar is the most popular architecture is the fact that the transmitter and receiver do not need to operate simultaneously (figure 5.3). This is a very important advantage since it automatically prevents leakage from the transmitter into the receiver, allowing for the two to share the same antenna, reducing the overall cost of the system.

5.3.2. Frequency Modulated - Continuous Wave Radar

The Frequency Modulated - Continuous Wave radar (FMCW) is the most common type of continuous wave radar and has been widely adopted in 77 GHz automotive radar systems [7, 8]. The main drive behind developing continuous wave (CW) systems is that, in contrast to the pulsed radar which transmits short bursts of energy, the CW radar generates constant output power over time. As a result, the peak power required by the transmitter circuits for the same amount of transmitted energy is significantly lower in the CW case, rendering its implementation much friendlier to silicon integrated circuits.

The block diagram of a generic FMCW system is illustrated in figure 5.4a. The transmitter generates a continuous wave signal whose frequency is modulated by alternating upwards - downwards ramps, also known as chirps, shown in figure 5.4b (i-ii). Assuming that each ramp linearly sweeps the transmit frequency from f_0 to $f_0+\Delta f$ over time T_b, the transmitted signal can be expressed as:

$$\mathrm{TX}(t) = A\cos\left(2\pi f_0 t + \frac{2\pi\Delta f}{T_b}\int_0^t \tau\,\mathrm{d}\tau\right) = A\cos\left(2\pi f_0 t + \pi\frac{\Delta f}{T_b}t^2\right)$$

The transmitted signal is reflected by the target, arrives at the input of the receiver with round-trip delay τ_D and gets multiplied by the transmitted signal:

$$\mathrm{RX}(t) = A_r\cos\left(2\pi f_0(t-\tau_D) + \pi\frac{\Delta f}{T_b}(t-\tau_D)^2\right)\cos\left(2\pi f_0 t + \pi\frac{\Delta f}{T_b}t^2\right)$$

After low-pass filtering:

$$\mathrm{RX}(t) = A_r\cos\left(2\pi\frac{\tau_D\Delta f}{T_b}t + 2\pi f_0\tau_D - \frac{\tau_D^2\Delta f}{T_b}\right) = A_r\cos\left(2\pi\frac{\tau_D\Delta f}{T_b}t + \phi\right) \tag{5.11}$$

Therefore, the receiver output contains a low frequency beat tone, depicted in figure 5.4b (iii-iv), at

$$f_b = \frac{\tau_D\Delta f}{T_b} = \frac{2\Delta f R}{cT_b} \tag{5.12}$$

The beat tone encodes the distance information in frequency. If many targets are present, equation (5.11) will result in the linear superposition of several sinusoids, each at a frequency related to the distance of a specific target. Therefore, the distances can be calculated from the IF signal using a simple Fast Fourier Transform (FFT) [160] or other, more advanced, spectral estimation techniques such as the Root-MUSIC algorithm [161]. If the target is moving, its speed can be easily calculated from the beat frequency difference between the upwards and downwards ramps [162, 163].

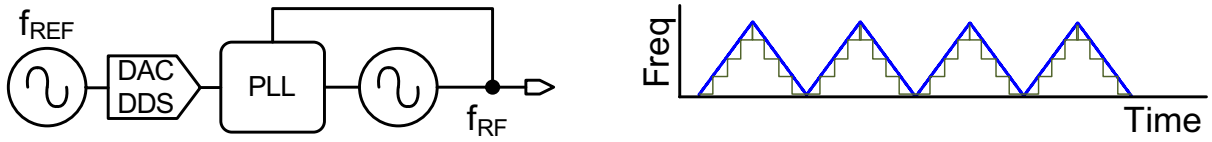
Similarly with the pulsed radar, the range resolution of the FMCW radar is [162]:

$$\Delta R \geq \frac{c}{2\Delta f} = \frac{c}{2B}$$

and the maximum range is related to the chirp period:

$$R_{max} \leq \frac{cT_b}{2}$$

where the assumption that $T_b \gg \tau_D$ was made.



(a) Block diagram. (b) Waveform.

Figure 5.5: FMCW ramp generation.

There are two important aspects that usually plague the performance of FMCW systems. First, the transmitter and receiver operate simultaneously leading to significant TX-to-RX leakage problems. Because of this, most FMCW radars have to employ separate antennas for the transmitter and receiver (bistatic radar). In case a single antenna is used in a monostatic FMCW radar, highly linear receiver circuits are required, resulting in higher receiver noise and higher power consumption.

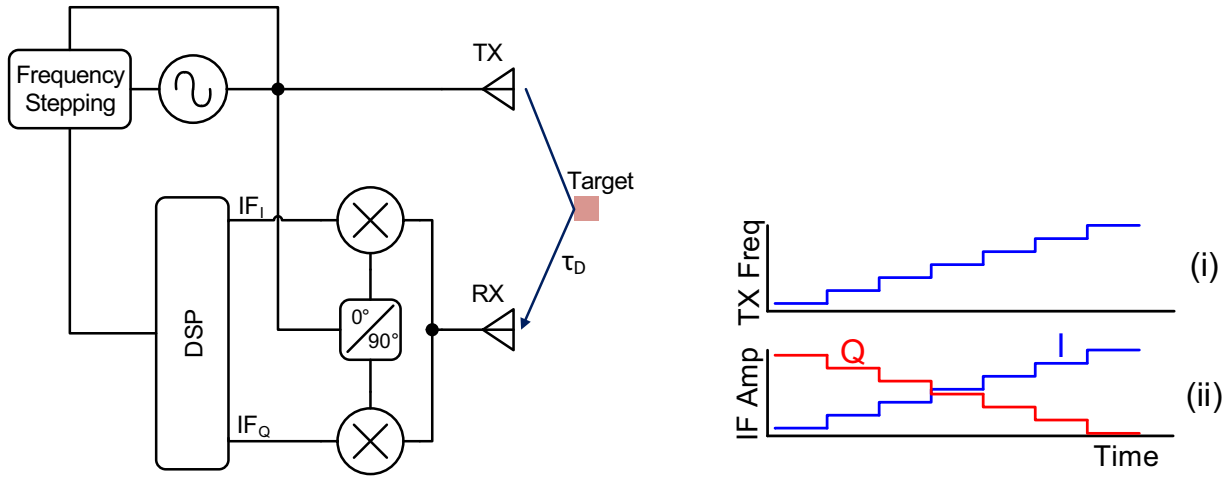
The second consideration has to do with the ramp generation process. As illustrated in figure 5.5a, a common method of generating a frequency ramp is to drive an adequately wideband Phase Locked Loop (PLL) with a Direct Digital Synthesizer (DDS), which in turn, is based on a multi-bit Digital-to-Analog Converter (DAC). However, due to the finite frequency resolution of the DDS, as well as its limited precision, the generated ramp will have a staircase-like shape as shown in figure 5.5b with generally unequal steps [164]. Therefore, the generated ramp will be nonlinear, which will have a deleterious impact on the beat frequency, rendering equation (5.12) invalid [165, 166] and resulting in systematic errors in the detected range.

5.3.3. Step Frequency Radar

The Step Frequency Radar (SFR) is a type of CW radar that has been proposed for detection of targets when a large SNR is required [154, 167, 168]. It shares the FMCW radar advantage of low peak output power and in addition, it is very flexible in terms of adjusting the measurement time for a variety of target scenarios.

The simplified block diagram of an SFR system is depicted in figure 5.6a. The physical principle is similar to the Time Domain Reflectometer (TDR) [169] and relies on the inverse Fourier transform. Suppose that a CW signal at frequency f, $A\cos(2\pi ft)$, is transmitted. The signal will be reflected by each of the N targets present and the superposition of all the reflections will appear at the input of the receiver:

$$\mathrm{RX} = \sum_{n=0}^{N} A_n \cos\left(2\pi f(t - \tau_{Dn})\right)$$



(a) Generic block diagram. (b) Waveforms.

Figure 5.6: Step frequency radar.

where A_n and τ_{Dn} are the amplitude of the reflected signal and round trip delay to the nth target respectively. The receiver performs quadrature downconversion with the same LO signal and low-pass filtering. This can be conveniently expressed using analytical signals:

$$\begin{aligned}
\text{IF} &= \text{IF}_I + \text{IF}_Q = \\
&= \sum_{n=0}^{N} A_n \cos\big(2\pi f(t-\tau_{Dn})\big)\cos(2\pi f t) - j\cdot\sum_{n=0}^{N} A_n \cos\big(2\pi f(t-\tau_{Dn})\big)\sin(2\pi f t) = \\
&= \sum_{n=0}^{N} \frac{A_n}{2}\cos(2\pi f\tau_{Dn}) - j\cdot\sum_{n=0}^{N}\frac{A_n}{2}\sin(2\pi f\tau_{D_n}) + \\
&\quad + \Big[\sum_{n=0}^{N}\frac{A_n}{2}\cos\big(4\pi f(t-\tau_{Dn}/2)\big) + j\cdot\sum_{n=0}^{N}\frac{A_n}{2}\cos\big(4\pi f(t-\tau_{Dn}/2)\big)\Big]
\end{aligned} \tag{5.13}$$

After low pass filtering and absorbing the factor-of-2 division into A_n,

$$\text{IF} = \sum_{n=0}^{N}\frac{A_n}{2}\big[\cos(2\pi f\tau_{Dn}) - j\sin(2\pi f\tau_{D_n})\big] = \sum_{n=0}^{N} A_n e^{-2\pi j f\tau_{Dn}} \tag{5.14}$$

The inverse Fourier transform of the above expression becomes:

$$\begin{aligned}
F^{-1}\{IF(f)\} &= \int_{-\infty}^{\infty}\left(\sum_{n=1}^{N} A_n e^{-2\pi j f\tau_{Dn}}\right)e^{2\pi j f t}df = \sum_{n=1}^{N} A_n\int_{-\infty}^{\infty} e^{2\pi j f(t-\tau_{Dn})}df = \\
&= \sum_{n=1}^{N} A_n\delta(t-\tau_{Dn}) = \sum_{n=1}^{N} A_n\delta\Big(t-\frac{2R_n}{c}\Big)
\end{aligned} \tag{5.15}$$

Consequently, the inverse Fourier transform of the receiver output signal is a superposition of N delta functions that "sit" at the target locations. The amplitude of the delta function for target n is equal to A_n.

Obviously, it is impossible for the radar to sweep through the entire frequency space as is implied in the integral calculation in equation (5.15). In practice, the transceiver sweeps over a finite set of frequencies $f_k = f_0 + k\Delta f$ (figure 5.6b) where Δf is the frequency step and $k = 1 \dots K$. The inverse Fourier transform of equation (5.15) is replaced by the Inverse Discrete Fourier Transform (IDFT), which introduces some important limitations.

First, the total sampled frequency bandwidth by the SFR is finite and equal to $B = K\Delta f$. According to the Shannon-Nyquist sampling theorem, the sampling period in time domain is:

$$\Delta t = \frac{1}{2K\Delta f} = \frac{1}{2B} \tag{5.16}$$

However, this period corresponds to a distance based on the formula:

$$\Delta t = \frac{\Delta R}{c} \tag{5.17}$$

Therefore, the range resolution is limited to:

$$\Delta R = \frac{c}{2K\Delta f} = \frac{c}{2B} \tag{5.18}$$

which is identical to that of the pulsed and FMCW radars.

Since there are K samples of distance ΔR, the maximum unambiguous range is:

$$R_{max} = K\,\Delta R = \frac{c}{2\Delta f} \tag{5.19}$$

There are several advantages in the SFR approach compared to the FMCW radar. The ramp linearity limitation is completely alleviated by using discrete frequency steps. Furthermore, perfect knowledge of the absolute frequency of operation $f_0 + k\Delta f$ is not strictly necessary, only of the frequency step Δf. Additionally, the number of frequency points K can be adjusted based on the number of targets N that are inside the field of view of the radar.

On the downside, if numerous targets are present, as in the automotive radar, the number of frequencies that need to be swept increases significantly. At each frequency the radar needs to spend some time to collect data and integrate to perform an accurate measurement, resulting in a high overall measurement time. The PLL will need extra time to hop from frequency to frequency, and settle, which would also add to the overall measurement time.

Consequently, the SFR is preferred for applications where the targets are static and the overall measurement time can be large, or when only few targets are present.

5.3.3.1. Step Frequency Radar with Two Steps

Assuming that only one target is present and two closely spaced frequencies f_1 and f_2 are scanned, substituting in equation (5.14) yields:

$$\begin{aligned} \mathrm{IF}(f_0) &= Ae^{-2\pi j f_0 \tau_D} \\ \mathrm{IF}(f_0+\Delta f) &= Ae^{-2\pi j (f_0+\Delta f)\tau_D} \end{aligned} \tag{5.20}$$

the phase of $\mathrm{IF}(f_0)/\mathrm{IF}(f_0+\Delta f)$ can be easily found to be:

$$\Delta\phi = \angle \frac{\mathrm{IF}(f_0)}{\mathrm{IF}(f_0+\Delta f)} = \angle e^{-2\pi j (f_0 - f_0 - \Delta f)\tau_D} = 2\pi \Delta f \tau_D \tag{5.21}$$

Simply dividing by $2\pi\Delta f$, the time of flight, τ_D, and therefore R is calculated:

$$R = \frac{\tau_D\, c}{2} = \frac{c}{4\pi\Delta f} \cdot \angle \frac{\mathrm{IF}(f_0)}{\mathrm{IF}(f_0+\Delta f)} = \frac{c\Delta\phi}{4\pi\Delta f} \tag{5.22}$$

Therefore, only two frequencies are necessary to calculate the distance to a single target.

This can also be intuitively explained by observing that if τ_D was estimated by only one frequency, say f_0, then $\tau_D = \angle\mathrm{IF}(f_0)/(2\pi f_0)$. However, if the phase of $\mathrm{IF}(f_0)$ becomes larger than 2π, it will wrap-around and make it impossible to resolve the distance unambiguously. In order to unwrap the phase, two frequencies are necessary with spacing that depends on the maximum unambiguous range, as predicted by equation (5.19).

If the phases of $\mathrm{IF}(f_0)$ and $\mathrm{IF}(f_0+\Delta f)$ are estimated with phase errors σ_{ϕ_1} and σ_{ϕ_2} respectively then, from equation (5.22), the distance error will be:

$$\sigma_{R,\Delta f} = \frac{c}{4\pi\Delta f} \cdot \sqrt{\sigma_{\phi_1}^2 + \sigma_{\phi_2}^2} \tag{5.23}$$

The error in the distance does not depend on the absolute frequency f_0 but only on the difference Δf. This implies that the phase error is divided by a small frequency Δf (e.g. $\Delta f = 75\,\mathrm{MHz}$ when $R_{max} = 2\,\mathrm{m}$) and does not improve if f_0 is increased. This limitation can be circumvented by the method proposed in [145], where the phase difference from equation (5.21) is used only as a coarse estimate. Subsequently, this estimate is used to calculate the number of multiples of 2π needed to unwrap the phase of $\mathrm{IF}(f_0)$ and calculate the distance directly from it, i.e.:

$$R = \frac{c}{4\pi f_0} \cdot \angle\mathrm{IF}(f_0) \tag{5.24}$$

The distance error then becomes:

$$\sigma_{R,f_0} = \frac{c}{4\pi f_0} \cdot \sigma_{\phi_1} \tag{5.25}$$

which can benefit significantly from increasing the frequency of operation f_0. Intuitively, this was expected since at higher frequencies, the wavelength becomes smaller resulting in larger phase shifts for smaller distances.

Nevertheless, in order for the phase unwrapping to be performed without errors, it is necessary that $(f_0/\Delta f)\Delta\phi < \pi/2$ [145] which translates to the following condition:

$$\sigma_{R,\Delta f} = \frac{c}{4\pi\Delta f} \cdot \sqrt{\sigma_{\phi_1}^2 + \sigma_{\phi_2}^2} < \frac{c}{8 f_0} \tag{5.26}$$

which introduces a trade-off in selecting Δf and f_0: larger Δf decreases the maximum range R_{max} but relaxes the above condition for phase unwrapping. Larger f_0 tightens the phase unwrapping condition but increases the accuracy. This trade-off can be partially alleviated by performing a progressive unwrapping with more than two frequency steps, as suggested in [154]. However, this would complicate the design of the frequency synthesizer and introduce new frequency errors.

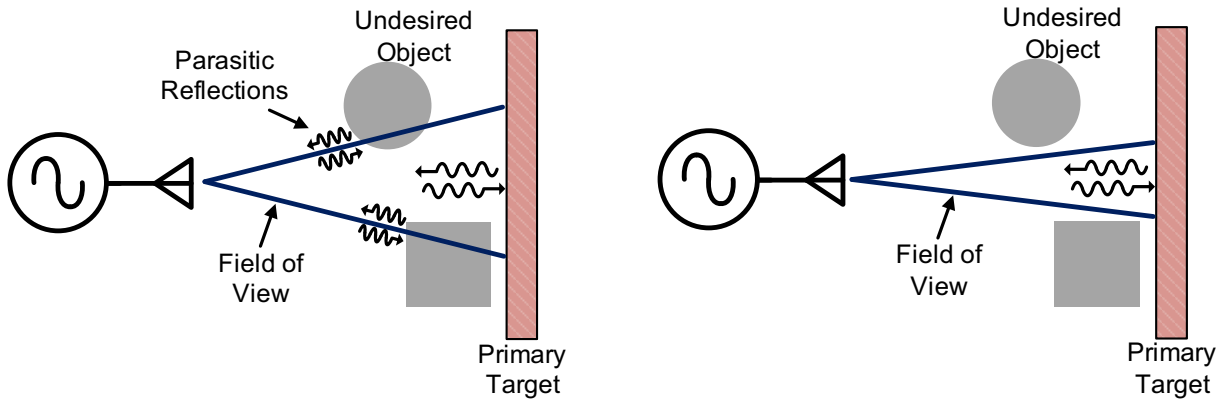
Another method to improve on the accuracy of the range estimation is through compressed sensing [170,171]. It is best suited for multiple targets and involves scanning through multiple frequencies, not necessarily equally spaced.

5.4. Selection of Frequency of Operation and Radar Waveform

5.4.1. Frequency of Operation

only a single target is present in the radar system under consideration. Consequently, the range resolution is not an important performance factor and the bandwidth requirement is considerably relaxed. Furthermore, unlike the case where multiple targets need to be detected with transceiver arrays, only one transceiver is necessary, leading to significantly lower system complexity and power dissipation.

However, if an undesired object (clutter) is present near the primary target, as in figure 5.7a, and a waveform with insufficient bandwidth to distinguish between them is employed, the detected distance will be the average of the two, degrading the accuracy. Consequently, to avoid increasing the bandwidth of the system, a narrow radar field-of-view is necessary, as shown in figure 5.7b. This can be achieved by employing a high gain - low beamwidth antenna. For example, a 30 dBi gain antenna would result in 2.3° beamwidth



(a) Presence of parasitic reflections.

(b) Restricted field of view.

Figure 5.7: The potential impact of the field of view on accuracy

which is higher than laser systems, but adequate for our applications. A 2.3° beamwidth parabolic antenna has a diameter of 180 cm at 5 GHz but only 7.5 cm at 120 GHz, indicating that there can be a significant size and beamwidth benefit at higher frequencies.

The unlicensed ISM band at 122 GHz - 123 GHz is ideal for the application under consideration for several reasons. First, the 1 GHz of available bandwidth is sufficient for the detection of a single large target. Second, the maximum allowed Effective Isotropically Radiated Power (EIRP), i.e. the product of the antenna gain times the output power $G_{TX}P_o$, is 20 dBm, which is adequate for operation up to 3 m, as will be shown below. Third, the possibility of an integrated antenna becomes viable, thus avoiding the cumbersome and high cost packaging solutions that would be required at a lower band like 60 GHz or 77 GHz. Lastly, as illustrated in figure 5.8, the atmospheric attenuation at 122 GHz is insignificant, even in high humidity, indicating that the environmental conditions will not impede with the correct operation of the system.

5.4.2. Radar Waveform

The range accuracy of the system needs to be at least 0.5 mm, or even better for applications like surface roughness monitoring. The accuracy of a radar is limited by both random and systematic errors. Random errors are the result of the receiver noise and will be analyzed in the following section. However, depending on the radar waveform employed, its can introduce systematic errors. For example, if ramp nonlinearity in FMCW modulation introduces a 1% systematic error in the measurement, as reported in [172], the accuracy goal of 0.5 mm would be impossible to meet at a distance of 1 m.

Therefore, frequency stepping becomes the best choice. In the simplest scenario, only two frequencies will be used to detect a single target. Stepping through more frequencies can

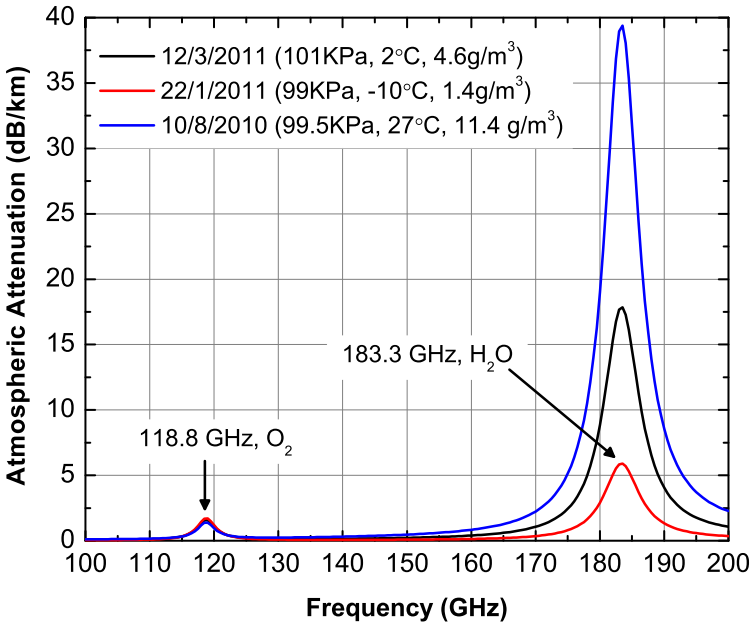


Figure 5.8: Atmospheric attenuation versus frequency in three days in Toronto, Canada based on the ITU-R P.676-8 model. The attenuation difference between the three days is due to the humidity.

Table 5.1: Radar system parameters

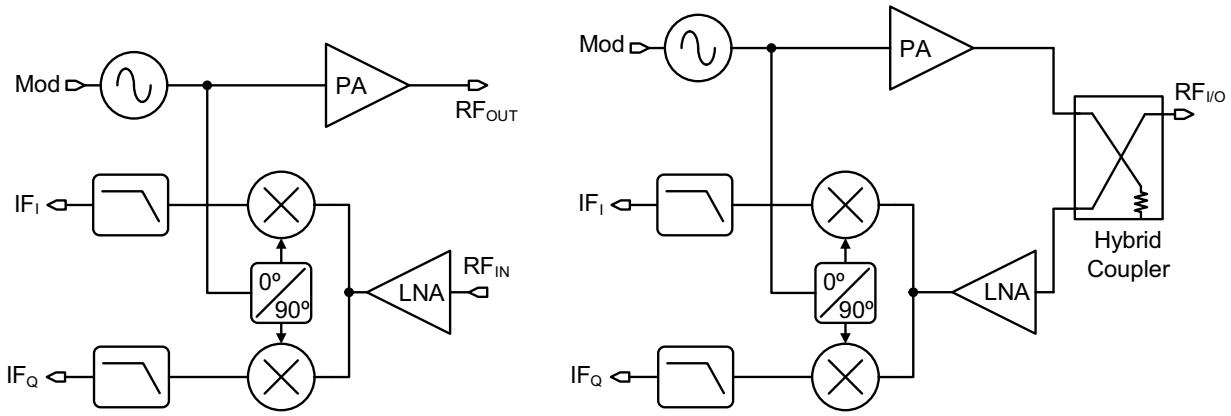
Maximum Range	3 m
Number of Targets	1
Field of View	$< 3°$
Accuracy	$<$0.5 mm
Frequency of Operation	122-123 GHz
Waveform	Frequency stepping

also be used to verify the presence of parasitic targets or for more advanced reconstruction using compressed sensing [170, 171].

Table 5.1 summarizes the radar system parameters.

5.5. Transceiver Architecture

In this section, the principal considerations for the radar system architecture will be analyzed with respect to their impact in the system accuracy. New solutions to circumvent the 1/f noise will be proposed and analytic simulations will be presented.



(a) Monostatic. (b) Bistatic.

Figure 5.9: Homodyne radar transceiver architectures.

5.5.1. Monostatic Versus Bistatic

Consider the monostatic transceiver architecture of figure 5.9a, where the transmitter output is separated from the receiver input, as opposed to the one of figure 5.9b, where the TX output and RX input are coupled to the same antenna through a directional coupler. The coupler is sensitive to the direction of the signals, routing the TX output to the antenna and the reflected signal to the RX, thus isolating the two. Both architectures have been employed extensively in radar systems [158] and there are certain design trade-offs in selecting one over the other.

In continuous wave radar, the transmitter and receiver operate simultaneously, resulting in TX to RX leakage problems. Therefore the RX must remain linear to avoid being blocked by the strong TX signal. The increased linearity comes with a noise figure cost. For example, in [173] a 77 GHz receiver with input 1-dB compression point of -22 dBm and associated double-sideband noise figure of 4.8 dB is reported, whereas in [7, 174] a receiver at the same frequency and in a similar SiGe technology exhibits an IP1dB of -0.3 dBm and DSB noise figure of 14 dB. For a transmit power of 7 dBm, typical of 77 GHz automotive radar, the former RX would require a TX-to-RX isolation of at least 30 dB while in the latter, an isolation of approximately 10 dB would suffice. Designing an on-chip coupler with 30 dB of isolation over different process corners is extremely challenging, necessitating the use of a bistatic architecture. In the second case, an integrated directional coupler can easily be employed but with a noise penalty in the receiver, on top of which one must add the loss of the coupler itself.

At 122 GHz, the guided half-wavelength is approximately $\lambda_g/2 = c/(\sqrt{\epsilon_r} f_0) = 600\mu$m and if the antenna is integrated on-chip, it will still require a small, but still considerable amount of silicon area, comparable to that of the transceiver itself. Furthermore, if two

antennas are on the same die, their mutual coupling is going to increase due the small distance between them, diminishing the advantage of the bistatic architecture. Consequently, if an on-chip antenna is to be employed, the monostatic architecture is the preferred approach. The increased receiver noise figure will be circumvented by a moderate increase in the measurement time to overcome the accuracy degradation.

5.5.2. Homodyne Transceiver

Consider the implementation of the step frequency radar with the homodyne transceiver of figure 5.9b. At each frequency step, the transmitter generates a CW signal of power P_{TX} and frequency f_0. The signal is attenuated by the coupling ratio a_{c} of the directional coupler before being transmitted by the antenna. The reflected signal will be collected back, after being attenuated by a_{radar} during its round trip to the target. This attenuation is governed by the radar equation [158]:

$$a_{\mathrm{radar}} = \frac{G_{ant}^2 c^2 \sigma}{(4\pi)^3 f_0^2 R^4} \tag{5.27}$$

where G_{ant} is the antenna gain of the monostatic system, R is the target distance and σ is its radar cross-section. The cross-section of a large surface will primarily depend on the relative angle between the surface and the antenna of the radar.

The antenna is connected through the direct arm of the coupler to the input of the receiver, resulting in further attenuation of the signal by $1 - a_{\mathrm{c}}^2$. Consequently, the signal power at the input of the receiver becomes:

$$P_{\mathrm{RX}} = P_{\mathrm{TX}} a_{\mathrm{radar}} a_{\mathrm{c}}^2 (1 - a_{\mathrm{c}}^2) = P_{\mathrm{TX}} a_{\mathrm{radar}} a_{\mathrm{coupler}} \tag{5.28}$$

where

$$a_{\mathrm{coupler}} = a_{\mathrm{c}}^2 (1 - a_{\mathrm{c}}^2) \tag{5.29}$$

is the total loss due to the directional coupler. Figure 5.10 illustrates a_{coupler} versus the coupling ratio in dB, $-20 \log a_{\mathrm{c}}$. The minimum loss is 6 dB when the coupling ratio is 3 dB and monotonically increases thereafter. The plot does not include the additional loss due to the finite conductivity of the metals used to realize the coupler. In general, the metal loss increases and the isolation degrades along with the coupling ratio [175, 176]. In order to relax this trade-off and improve the isolation, the coupling ratio is set to 6 dB which results in 7.25 dB loss.

The quadrature receiver performs complex multiplication of the received signal with the

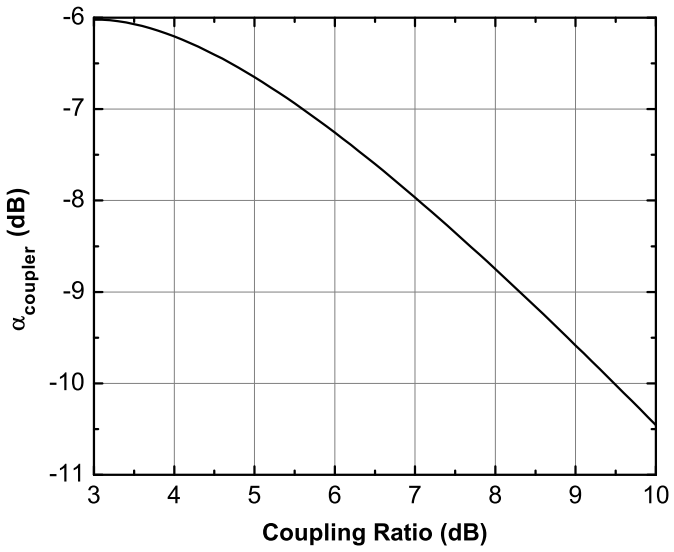


Figure 5.10: a_{coupler} versus coupling ratio a_{c}

transmitted one. The I and Q receiver outputs, IF_I and IF_Q, become:

$$\begin{aligned} \text{IF}_I &= \sqrt{G_{\text{RX}} P_{\text{RX}}} \cos\phi \\ \text{IF}_Q &= \sqrt{G_{\text{RX}} P_{\text{RX}}} \sin\phi \end{aligned} \tag{5.30}$$

where G_{RX} is the receiver power gain and $\phi = 2\pi f_0 \tau_D$ is the phase shift due to the round trip delay to the target τ_D. Ideally, in the absence of noise, ϕ can be calculated as:

$$\phi = \tan^{-1} \frac{\text{IF}_Q}{\text{IF}_I} \tag{5.31}$$

Nevertheless, due to the presence of receiver noise, an estimate of the angle, $\hat{\phi}$, is acquired:

$$\hat{\phi} = \tan^{-1} \left(\frac{\text{IF}_Q + \sigma_Q}{\text{IF}_I + \sigma_I} \right) \simeq \tan^{-1} \left(\frac{\text{IF}_Q}{\text{IF}_I} - \frac{\sigma_Q \text{IF}_I + \sigma_I \text{IF}_Q}{\text{IF}_I^2} \right) \tag{5.32}$$

where σ_I and σ_Q are the standard deviations of the noise present at the IF_I and IF_Q outputs respectively, assumed to be Gaussian but not necessarily white. The approximation $1/(a + x) \simeq 1/a - x/a^2$ for small x, has been used to derive the right hand side of the above equation. Since $\tan^{-1}(a + x) \simeq \tan^{-1} a + x/(1 + a^2)$ for small x, we can further approximate:

$$\hat{\phi} = \phi + \sigma_\phi \simeq \tan^{-1} \left(\frac{\text{IF}_Q}{\text{IF}_I} \right) - \frac{\sigma_Q \text{IF}_I + \sigma_I \text{IF}_Q}{\text{IF}_I^2 + \text{IF}_Q^2} \tag{5.33}$$

Consequently, the standard deviation of the estimated phase becomes:

$$\sigma_\phi = \frac{\sigma_Q \mathrm{IF}_I + \sigma_I \mathrm{IF}_Q}{\mathrm{IF}_I^2 + \mathrm{IF}_Q^2} = \frac{\sigma \sin(\pi/4 + \phi)}{\sqrt{2 G_{\mathrm{RX}} P_{\mathrm{RX}}}} \tag{5.34}$$

where the standard deviation of the noise at the two outputs is assumed equal $\sigma_I = \sigma_Q = \sigma$ and the fact that $\mathrm{IF}_I^2 + \mathrm{IF}_Q^2 = G_{\mathrm{RX}} P_{\mathrm{RX}}$ and $\mathrm{IF}_I + \mathrm{IF}_Q = \sqrt{2 G_{\mathrm{RX}} P_{\mathrm{RX}}} \sin(\pi/4 + \phi)$ has been used.

Since we are interested only in the worst case σ_ϕ, which occurs when $\sin(\pi/4 + \phi) = 1$, the formula for the estimation error of the phase can be derived:

$$\sigma_\phi = \frac{1}{\sqrt{2 G_{\mathrm{RX}} P_{\mathrm{RX}} / \sigma^2}} = \frac{1}{\sqrt{2 SNR_o}} \tag{5.35}$$

Therefore, the standard deviation of the estimated phase depends on the SNR of the system.

Accounting for flicker noise, the receiver output noise power spectral density (PSD), at each of the I and Q outputs, can be expressed as:

$$N_{\mathrm{RX}} = k_B T \left(1 + \frac{f_c}{f}\right) F G_{\mathrm{RX}} \tag{5.36}$$

where f_c is the output 1/f noise corner. If the IF low-pass filters in figure 5.9b are implemented by averaging several measurements over the measurement time τ_s, the SNR at the receiver output is calculated by employing equation (5.94) of appendix 5.A for DC signals under combined 1/f and white noise:

$$SNR_o = \frac{P_{\mathrm{RX}} G_{\mathrm{RX}}}{k_B T F G_{\mathrm{RX}} (1/\tau_s + f_c \ln 16)} = \frac{P_{\mathrm{RX}}}{k_B T F (1/\tau_s + f_c \ln 16)} \tag{5.37}$$

The estimation of the target distance will be performed by switching between two frequencies spaced by Δf. The accuracy of the radar, measured by the standard deviation of the distance, is given by equation (5.23):

$$\sigma_{R,\Delta f} = \frac{c}{4\pi \Delta f} \cdot \sqrt{\sigma_{\phi_1}^2 + \sigma_{\phi_2}^2} = \frac{\sqrt{2} c \, \sigma_\phi}{4\pi \Delta f} \tag{5.38}$$

where $\sigma_{\phi_1} = \sigma_{\phi_2} = \sigma_\phi$. In order to unwrap the phase and improve the accuracy, the following condition must be true at 122 GHz:

$$\sigma_{R,\Delta f} < \frac{c}{8 f_0} = 0.307 \,\mathrm{mm} \tag{5.39}$$

Table 5.2: Radar transceiver parameters

Symbol	Description	Value
R	Target distance	2-3 m
σ	Radar cross-section	0.01 m^2
f_0	Center frequency	122 GHz
T	System temperature	300 K
G_{ant}	Antenna gain	20 dBi
P_{TX}	Transmit power	0 dBm
a_c	Coupler coupling ratio	6 dB
F	Receiver noise figure	15 dB
f_c	Receiver flicker noise corner	10 kHz
τ_s	Measurement time	variable

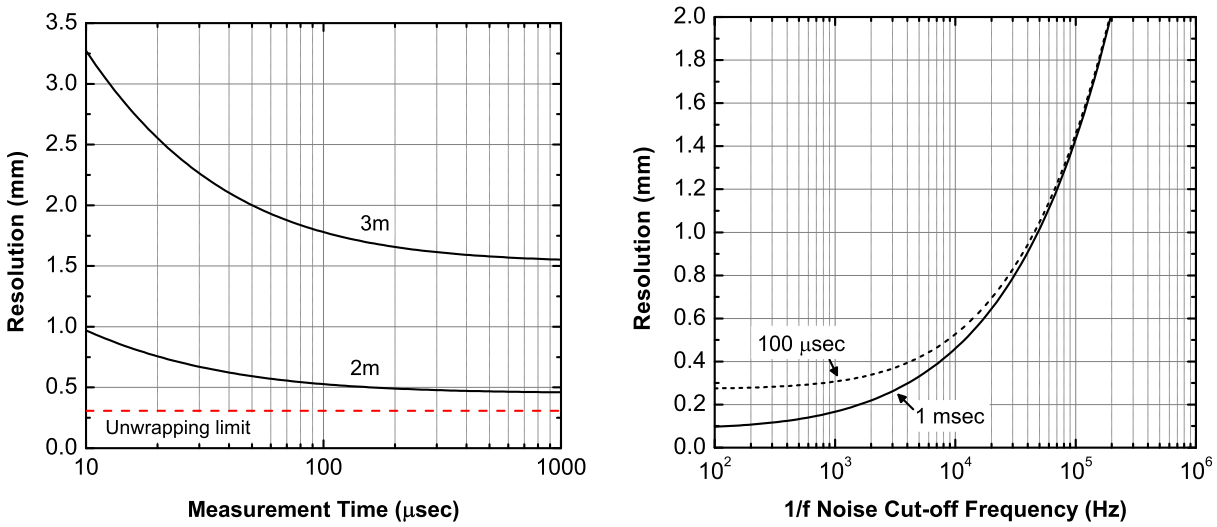
Equations (5.35), (5.37) and (5.38) can be combined to predict the accuracy of the two frequency steps radar, $\sigma_{R,\Delta f}$, versus the measurement time and the 1/f noise corner frequency for the system parameters summarized in table 5.2. The EIRP is set to 20 dBm, in accordance with the regulation for the 122 GHz ISM band while the receiver noise figure has been set to 15 dB, higher than that of state-of-art D-band SiGe HBT transceivers (e.g. [177–180]) in order to account for the degradation due to the coupler loss. The flicker noise corner for the receiver was set to 10 kHz, which is based on simulations of the receiver presented in chapter 6.

Figure 5.11a illustrates the accuracy $\sigma_{R,\Delta f}$ of the detected distance versus measurement time when the distance of the target is 2 and 3 m. Although the accuracy can be improved substantially with increasing τ_s, none of the two cases achieves the limit of 0.307 mm required to unwrap the phase. As indicated in equation (5.37), the 1/f noise component is immune to averaging, leading to a plateau, after which increasing the measurement time does not improve the accuracy. This plateau can be calculated from equation (5.37) by letting $\tau_s \to \infty$:

$$SNR_{o,max} = \frac{P_{\mathrm{RX}}}{k_B T F f_c \ln 16} \tag{5.40}$$

The impact of 1/f noise is also highlighted in figure 5.11b where the accuracy is plotted versus the 1/f noise corner frequency, for two measurement times. As the corner frequency increases above approximately 1 kHz, the accuracy rapidly degrades even when $\tau_s = 1$ msec.

In 77 GHz automotive radars found in the literature [7,8], the receivers are homodyne and are implemented in SiGe, which exhibits better 1/f noise corners than CMOS. For example in [8] a corner frequency of 60 kHz is reported, which would have been unacceptable for this application. However, the problem is less severe in the FMCW radar because the distance



(a) Accuracy versus measurement time.

(b) Accuracy versus 1/f noise cut-off frequency.

Figure 5.11: Accuracy ($\sigma_{R,\Delta f}$) of the homodyne radar transceiver.

information is encoded at a finite beat frequency, rather than DC. As reported in [8], the frequency ramps can be made sufficiently steep for the beat frequencies to be near or even above the 1/f noise corner frequency.

Therefore, it becomes evident that a different transceiver design is necessary to alleviate the 1/f noise problem.

5.5.3. Heterodyne Transceiver

Consider the proposed heterodyne transceiver (TXRX) of figure 5.12. The TX LO is tuned at frequency f_0 and the transmitter generates a CW signal:

$$\mathrm{TX}(t) = A_{\mathrm{TX}} \cos(2\pi f_0 t)$$

After being reflected by the target, it appears at the input of the receiver:

$$\mathrm{RX}(t) = A_{\mathrm{RX}} \cos(2\pi f_0 t - 2\pi f_0 \tau_D)$$

Unlike the homodyne transceiver, the receiver now has a separate VCO that is tuned to a slightly different frequency to that of the transmitter, $f_0 - f_{\mathrm{IF}}$. As a result, the receiver output is at a finite frequency f_{IF}:

$$\mathrm{IF}(t) = A_{\mathrm{RX}} \cos(2\pi f_{\mathrm{IF}} t - 2\pi f_0 \tau_D)$$

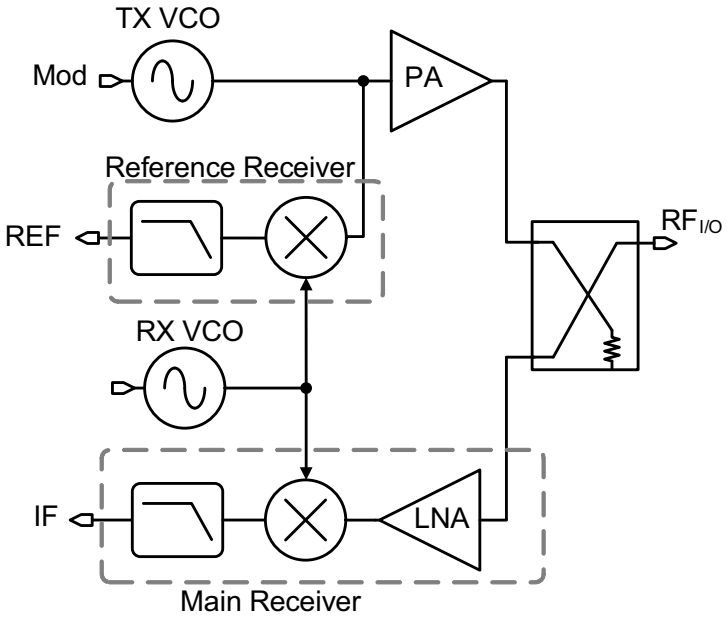


Figure 5.12: Heterodyne radar transceiver architecture.

The phase of the IF signal cannot be measured directly as it would depend on the start time of observation. To circumvent this problem, a separate reference receiver is introduced, which also generates a sinusoid at f_{IF} but whose phase is independent of the target distance:

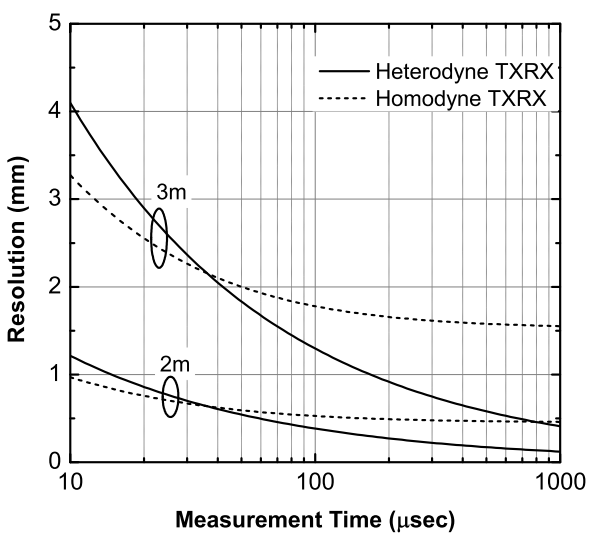
$$\mathrm{REF}(t) = A_{\mathrm{REF}} \cos(2\pi f_{\mathrm{IF}} t + \phi_{ref})$$

The phases of the IF and REF output sinusoids are compared, yielding $\phi_1 = 2\pi f_0 \tau_D - \phi_{ref}$. The process is repeated at the next frequency step $f_0 + \Delta f$ where the RX LO is tuned at $f_0 + \Delta f - f_{\mathrm{IF}}$, yielding $\phi_2 = 2\pi(f_0 + \Delta f)\tau_D - \phi_{ref}$. When ϕ_1 and ϕ_2 are subtracted, ϕ_{ref} vanishes and the distance can be calculated directly by equation (5.22).

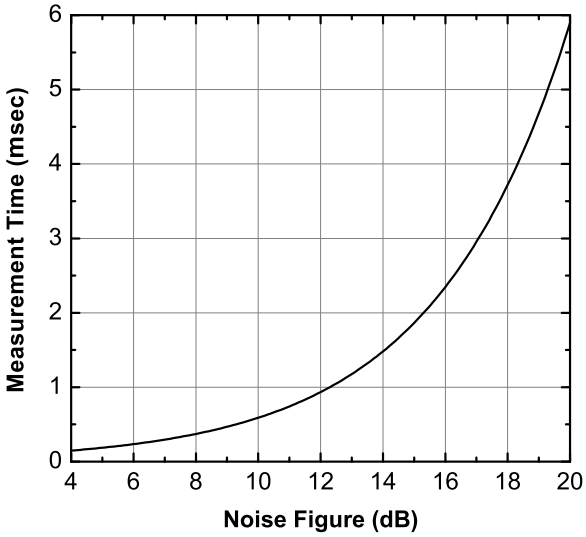
The advantage of the heterodyne transceiver becomes apparent when the IF frequency f_{IF} is set to a larger value than the 1/f noise corner frequency f_c. In that case, the receiver filter that will act on the IF and REF outputs and will cut-off the 1/f noise, drastically improving the SNR compared to that of the homodyne transceiver. A simple way to implement the filter and simultaneously calculate the phase difference between the IF and REF outputs is to directly digitize them, calculate their DFTs and subtract the phases of the Fourier coefficients that correspond to frequency f_{IF} [181, 182]. The phase error due to noise then becomes [181]:

$$\sigma_\phi = \frac{1}{\sqrt{SNR_o}} \tag{5.41}$$

where SNR_o is the SNR at the IF output of the receiver.



(a) Accuracy versus measurement time of the heterodyne TXRX compared with the homodyne.



(b) Measurement time required to achieve a resolution of 0.3 mm versus the receiver noise figure for a distance of 3 m.

Figure 5.13: Accuracy ($\sigma_{R,\Delta f}$) of the heterodyne radar transceiver.

Assuming that the frequency f_{IF} is larger than the 1/f noise corner, the SNR_o at the IF output, after the calculation of the DFT, is calculated by equation (5.100) of appendix 5.A:

$$SNR_o = \frac{S_i^2 \tau_s}{N_i} \tag{5.42}$$

where S_i, is the amplitude of the sinusoid and N_i is the white noise PSD. In this case, $S_i^2 = P_{\text{RX}} G_{\text{RX}}$ and $N_i = k_B T F G_{\text{RX}}$ resulting to:

$$SNR_o = \frac{G_{\text{RX}} P_{\text{RX}} \tau_s}{k_B T F G_{\text{RX}}} = \frac{P_{\text{RX}} \tau_s}{k_B T F} \tag{5.43}$$

As in the homodyne transceiver, the distance accuracy is governed by equation (5.38), which can be combined with (5.41) and (5.43) to predict the system accuracy of the heterodyne radar. Figure 5.13a shows the accuracy versus measurement time of the heterodyne TXRX, compared to that of the homodyne for the parameters of table 5.2. At short measurement times, where the noise in dominated by its white component, the accuracy of the heterodyne radar is $\sqrt{2}$ times worse than that the homodyne due to the absence of a quadrature receiver. As the measurement time increases, the white noise is filtered out, drastically improving the accuracy and surpassing that of the homodyne. At a distance of 2 m, the distance accuracy becomes better than 0.3 mm, after 150 μsec whereas the corresponding measurement time for a distance of 3 m is 2 msec. Due to the absence of 1/f noise, no

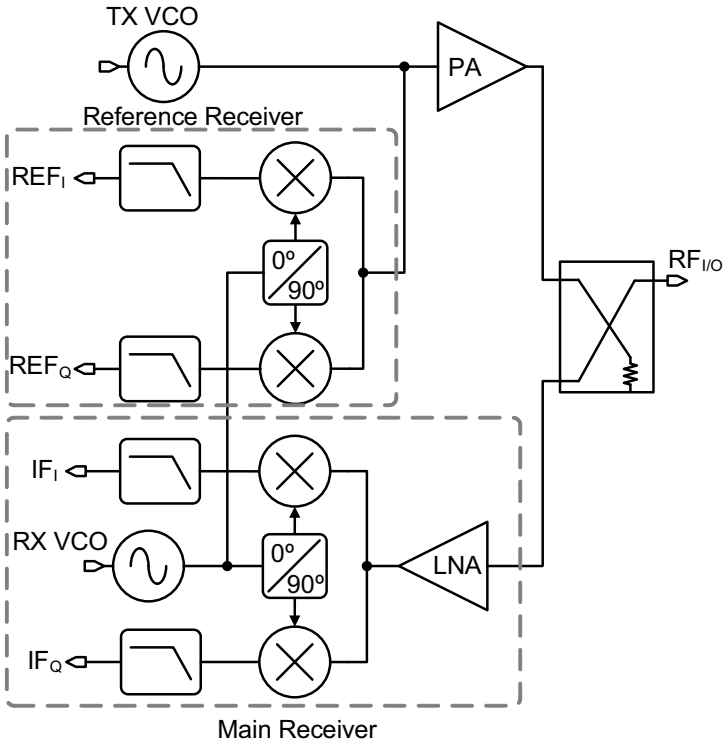


Figure 5.14: IQ heterodyne transceiver architecture.

plateau is formed in the accuracy curve, which keeps improving with τ_s.

Figure 5.13b depicts the measurement time required to achieve a resolution of 0.3 mm versus the receiver noise figure. A noise figure as high as 20 dB can be tolerated at the expense of increased integration time. If the resolution target of 0.3 mm is reached and phase unwrapping is performed, the accuracy of the system will be improved to less than 1 μm, indicating that it will not be dominated by noise, but rather by other systematic errors.

The σ_ϕ of the heterodyne TXRX can be improved by a factor of $\sqrt{2}$ with the IQ architecture of figure 5.14. The operation of this system is identical to the one of figure 5.12, apart from the fact that quadrature receivers are introduced at both the reference and receive channels. The output sinusoids are now complex and the estimation error becomes [181]:

$$\sigma_\phi = \frac{1}{\sqrt{2SNR_o}} \tag{5.44}$$

which is identical to that of the homodyne TXRX. Figure 5.15 illustrates the accuracy versus measurement time for the IQ heterodyne TXRX. As expected, the IQ heterodyne TXRX has superior performance since it avoids the 1/f noise without the $\sqrt{2}$ penalty in σ_ϕ.

There are disadvantages associated with the proposed heterodyne architectures. First,

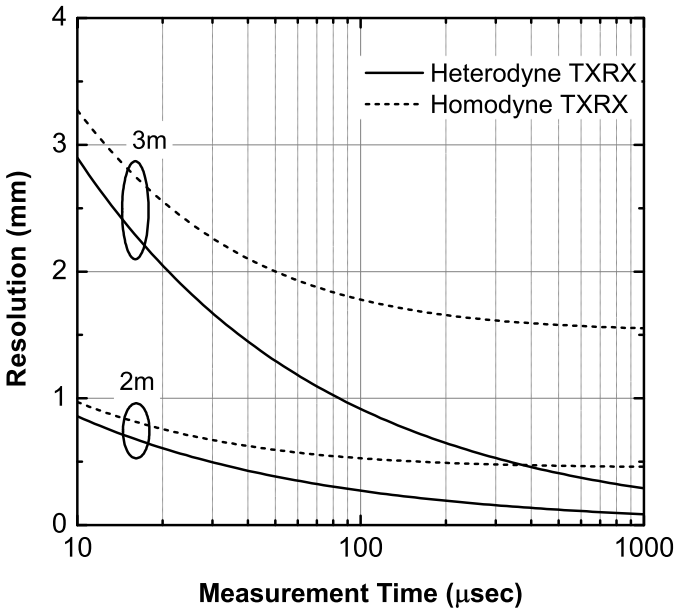


Figure 5.15: Accuracy versus measurement time of the IQ heterodyne TXRX compared with the homodyne.

the system complexity is increased substantially compared to the homodyne case, as two oscillators and two mixers are required. The increased complexity inevitably leads to increased design effort, especially at 122 GHz, as well as to additional power dissipation. Furthermore, having two oscillators on the same die operating at a small frequency offset poses a new design challenge because of the finite isolation that can be achieved on die. At small offsets, the two oscillators will tend to injection lock each other and force the system to operate at zero-IF, negating the heterodyne advantage. To avoid this adverse condition, f_{IF} needs to be sufficiently large to avoid injection locking. However, in that case, f_{IF} could be too high to be directly digitized by a simple ADC and further downconversion might be necessary.

5.5.4. Homodyne Transceiver with Chopping

An alternative implementation of the homodyne architecture that deals with the 1/f noise problem is also possible, through the use of chopping [183]. A straightforward way to implement chopping in the transceiver is to periodically change the transmitted signal, at a rate higher than the receiver 1/f noise corner, thus moving the IF signal from DC to a higher frequency. Such an implementation is shown in figure 5.16 where BPSK modulators (± 1 multipliers) are introduced in the transmit path.

Interestingly, this architecture is very similar to the pulsed radar. The TX phase switching can be used for a coarse distance estimation and perform the necessary phase unwrapping and the final distance estimation can be achieved by measuring phase delay at only one frequency.

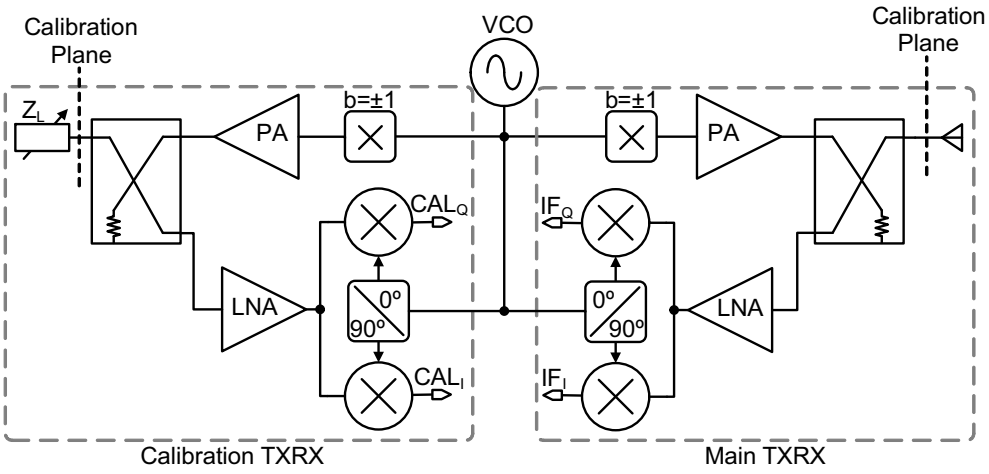


Figure 5.16: Radar transceiver with coarse/fine distance measurement and chopping.

However, this architecture suffers from the same synchronization challenges between the transmitter and the receiver, as in all pulsed radars. Furthermore, any IQ imbalance in the receivers is going to directly translate in distance measurement error, and I-Q calibration is necessary [184].

5.5.5. Impact of Phase Noise

The previous analysis assumed that the only source of noise in the system was the receiver output noise floor. It ignored the phase noise of oscillator signals which can be significant noise contributors. As will be shown below, short-range radars have an inherent advantage in dealing with phase noise compared to communication systems and long-range radars.

Consider the heterodyne TXRX of figure 5.12 with phase noise included in both oscillators. The signal generated by the transmitter is then given by:

$$\mathrm{TX}(t) = A_{\mathrm{TX}} \cos[2\pi f_0 t + \phi_{\mathrm{TX}}(t)]$$

where $\phi_{\mathrm{TX}}(t)$ is the phase noise signal generated by the TX LO. The REF output becomes:

$$\mathrm{REF}(t) = A_{\mathrm{REF}} \cos[2\pi f_0 t + \phi_{ref} + \phi_{\mathrm{TX}}(t) - \phi_{\mathrm{RX}}(t)]$$

where $\phi_{\mathrm{RX}}(t)$ is the corresponding phase noise of the RX LO. The signal at the input of the receiver after the round-trip delay τ_D is:

$$\mathrm{RX}(t) = A_{\mathrm{RX}} \cos[2\pi(f_0 + f_{\mathrm{IF}})(t - \tau_D) + \overline{\phi_{\mathrm{TX}}(t - \tau_D)}]$$

and after downconversion to IF:

$$\mathrm{IF}(t) = A_{IF}\cos[2\pi f_{\mathrm{IF}}t - 2\pi f_0\tau_D + \overline{\phi_{\mathrm{TX}}(t-\tau_D)} - \phi_{\mathrm{RX}}(t)]$$

The phases of IF(t) and REF(t) are then subtracted to calculate τ_D, yeilding:

$$\Delta\phi = 2\pi f_0\tau_D - \phi_{ref} - \overline{\Delta\phi_n(t)} \tag{5.45}$$

where

$$\Delta\phi_n(t) = \overline{\phi_{\mathrm{TX}}(t-\tau_D)} - \phi_{\mathrm{TX}}(t) \tag{5.46}$$

is the error due to the phase noise of the oscillator. If τ_D is small, the subtraction of a phase noise signal from its delayed replica leads to a noise cancelation effect, known as phase noise correlation or range correlation [185, 186]. If the single-sideband power spectral density of $\phi_{\mathrm{TX}}(t)$ is S_ϕ, the spectral density of $\Delta\phi_n(t)$ is given by [186]:

$$S_{\Delta\phi} = S_\phi \cdot 4\sin^2\left(\pi\tau_D f\right) = S_\phi\left[4\sin^2\left(2\pi\frac{Rf}{c}\right)\right] \tag{5.47}$$

where f is the frequency offset from the carrier. Therefore, range correlation acts on the noise close to the carrier and attenuates it.

The receiver output PSD, including phase noise, becomes:

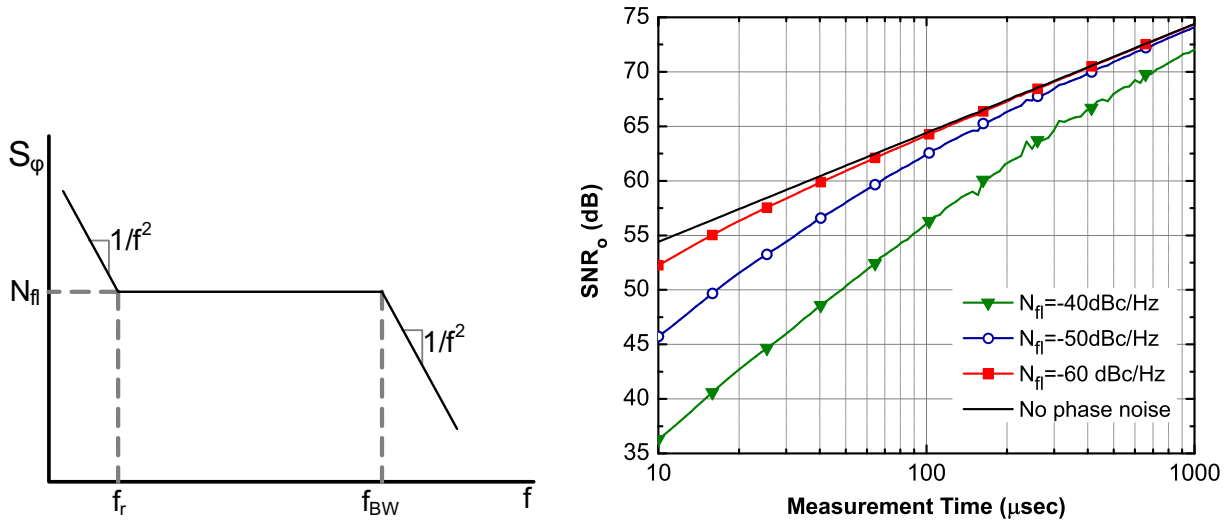
$$N_{o,RX} = 2G_{\mathrm{RX}}P_{\mathrm{RX}}S_{\Delta\phi} + k_BTFG_{\mathrm{RX}} \tag{5.48}$$

where the factor of 2 comes from the conversion of the single sideband phase noise spectrum of the VCO to double sideband After collecting N samples during time τ_s and calculating the DFT, the noise calculates to:

$$N_o = 2\int_0^{\frac{N}{2\tau_s}} (2G_{\mathrm{RX}}P_{\mathrm{RX}}S_{\Delta\phi} + k_BTFG_{\mathrm{RX}})\frac{\sin(\pi f\tau_s)^2}{N^2\sin(\pi f\tau_s/N)^2}\left[1-\cos(2\pi f\tau_s)\right]\mathrm{d}f \tag{5.49}$$

where the second term in the product accounts for the DFT filtering and the third term describes the fact that the measurements are paired (appendix 5.A). Assuming that N is large (>20), the above expression simplifies to:

$$\begin{aligned} N_o &\simeq 2\int_0^{\infty} (2G_{\mathrm{RX}}P_{\mathrm{RX}}S_{\Delta\phi} + k_BTFG_{\mathrm{RX}})\frac{\sin(\pi f\tau_s)^2}{\pi^2 f^2\tau_s^2}\left[1-\cos(2\pi f\tau_s)\right]\mathrm{d}f = \\ &= 2\int_0^{\infty} \left[8G_{\mathrm{RX}}P_{\mathrm{RX}}S_\phi\sin^2\left(2\pi\frac{d\,f}{c}\right) + k_BTFG_{\mathrm{RX}}\right]\frac{\sin(\pi f\tau_s)^2}{\pi^2 f^2\tau_s^2}\left[1-\cos(2\pi f\tau_s)\right]\mathrm{d}f \end{aligned} \tag{5.50}$$



(a) Transmitter oscillator phase noise profile. (b) SNR_o versus measurement time for different values for N_{fl}

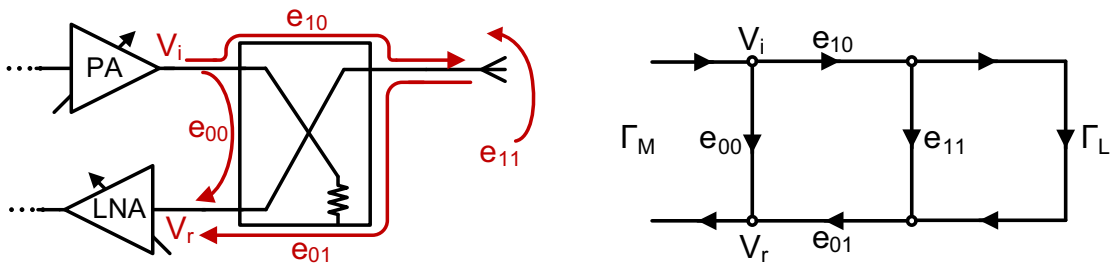
Figure 5.17: Impact of the oscillator phase noise.

The integral in (5.50) cannot be calculated analytically and numerical integration needs to be employed.

To simulate the SNR using the above formula, the phase noise profile, S_ϕ, of figure 5.17a is assumed for the TX oscillator. Typically, the LO is locked in a PLL with a low frequency crystal reference. The phase noise is then equal to that of the reference plus $20 \log N_d$ inside the PLL bandwidth, f_{BW}. N_d is the ratio of the TX frequency over the reference frequency. As illustrated in figure 5.17a, the reference phase noise drops with 20 dB/decade slope until a low frequency f_r, usually between 100 Hz to 1 kHz. After f_r, the phase noise is flat and equal to N_{fl} up until the PLL bandwidth f_{BW}, after which the PLL tracks the phase noise of the open loop oscillator, which typically exhibits a 20 dB/decade slope. Worst case f_r and N_{fl} values for a 50 MHz crystal reference are 1 kHz and -125 dBc/Hz+20log(2440) = -57 dBc/Hz respectively.

Figure 5.17b shows the SNR of the heterodyne radar system for a distance of 2 m and f_{BW}=50 kHz, for different values of N_{fl}. Even at the extreme case when N_{fl}=-40 dBc/Hz, the SNR is only 3 dB worse for 1 msec integration time. This degradation can easily be recovered by increasing the integration time to 2 msec.

The above simulations highlight the fact that the impact of phase noise on the system accuracy is minimal, as long as the maximum target distance R is small. This highly beneficial outcome is due to the extensive filtering of phase noise in the system. At low frequency offsets, the phase noise is filtered by the range correlation effect, whereas at higher offsets, it is filtered by the DFT averaging. Consequently the dominant source of noise in the system



(a) Sources of systematic error (b) Signal flow graph

Figure 5.18: Analysis of systematic errors

was properly assumed to be the receiver noise figure and not the TX LO phase noise.

5.6. Calibration

There are numerous sources of systematic errors in the transceivers presented in section 5.5 that need to be taken into consideration. The nature of there errors can be easily identified if the operation of the proposed transceivers is revisited from a different point of view. Essentially, they generate a CW signal and measure its relative change in amplitude and phase as this signal travels in free space and gets reflected back by a target. This is identical to the operation of the one-port reflectometer, or the one-port Vector Network Analyzer (VNA), for which the sources of errors are well understood and documented. Furthermore, well defined methods for VNA calibration exist and can be taken advantage of in the proposed radar system.

5.6.1. Sources of Error

The impairments that lead to systematic errors in the monostatic transceiver are summarized in figure 5.18a [187, 188]. The error coefficients e_{00}, e_{01}, e_{10} and e_{11} are generally complex numbers and capture the error sources in the one-port VNA. e_{00} is the directivity error and corresponds to the unavoidable direct leakage of signal from the transmitter to the receiver. The major source of this leakage is the imperfect isolation of the coupler, but also substrate and electromagnetic coupling between neighboring transmitter and receiver circuits. e_{01} and e_{10} are the reflection tracking errors and account for attenuation and finite delay that the signal experiences inside the transceiver. e_{11} is the source match error and corresponds to the unwanted reflections due to the imperfect matching of the antenna and coupler. The impact of any other phase shift and leakage that occurs inside the transceiver, can be lumped in the e_{00}, e_{01} and e_{10} coefficients.

5.6.2. External Calibration

The error coefficients of figure 5.18a are better illustrated and can be analyzed through the signal flow graph of figure 5.18b. If the reflection coefficient under measurement is Γ_{L}, then due to e_{00}, e_{01}, e_{10} and e_{11}, the system measures a different, erroneous reflection coefficient Γ_{M}. Simplifying the signal flow graph yields Γ_{M}:

$$\Gamma_{\mathrm{M}} = e_{00} + \frac{e_{10}e_{01}\Gamma_{\mathrm{L}}}{1 - e_{11}\Gamma_{L}} \tag{5.51}$$

Setting

$$\mathrm{a} = e_{10}e_{01} - e_{00}e_{11} \tag{5.52}$$

$$\mathrm{b} = e_{00} \tag{5.53}$$

$$\mathrm{c} = -e_{11} \tag{5.54}$$

transforms equation (5.51) to:

$$\Gamma_{\mathrm{M}} = \frac{\mathrm{a}\Gamma_{\mathrm{L}} + \mathrm{b}}{\mathrm{c}\Gamma_{\mathrm{L}} + 1} \tag{5.55}$$

The measured reflection coefficient Γ_{M} can be used to calculate the actual reflection coefficient Γ_{L}, provided that a, b and c are known.

The calibration procedure for characterizing a, b and c involves measuring three known, arbitrary reflection coefficients Γ_{L1}, Γ_{L2}, Γ_{L3} with corresponding measured system outputs Γ_{M1}, Γ_{M2}, Γ_{M3}. Substituting in equation (5.55) leads to a linear system of equations that can be solved for a, b and c:

$$\begin{bmatrix} \Gamma_{\mathrm{L1}} & 1 & -\Gamma_{\mathrm{L1}}\Gamma_{\mathrm{M1}} \\ \Gamma_{\mathrm{L2}} & 1 & -\Gamma_{\mathrm{L2}}\Gamma_{\mathrm{M2}} \\ \Gamma_{\mathrm{L3}} & 1 & -\Gamma_{\mathrm{L3}}\Gamma_{\mathrm{M3}} \end{bmatrix} \begin{bmatrix} \mathrm{a} \\ \mathrm{b} \\ \mathrm{c} \end{bmatrix} = \begin{bmatrix} \Gamma_{\mathrm{M1}} \\ \Gamma_{\mathrm{M2}} \\ \Gamma_{\mathrm{M3}} \end{bmatrix} \tag{5.56}$$

Once a, b and c have been calculated, the corrected Γ_{L} can simply be found from any measured Γ_{M} using the equation:

$$\Gamma_{\mathrm{L}} = \frac{\Gamma_{\mathrm{M}} - \mathrm{b}}{\mathrm{a} - \mathrm{c}\Gamma_{\mathrm{M}}} \tag{5.57}$$

In order to calibrate the radar transceivers of figures 5.9b, 5.12 or 5.15, a procedure that involves three external loads is necessary. The simplest way to implement this process in free space is to use an external reflector at predefined positions. For example, as illustrated in figure 5.19 [189], one measurement can be done with no reflector present, thus $\Gamma_{\mathrm{L1}} = 0$. Subsequently for the next two measurements, a reflector is brought at distances d_1 and

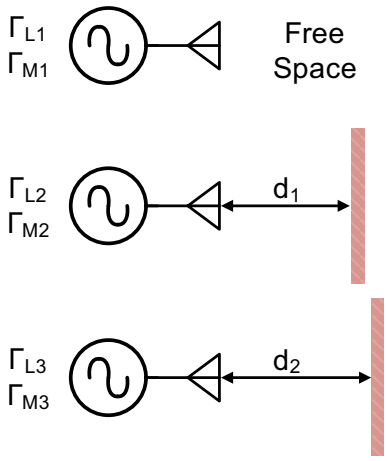


Figure 5.19: Three stage free space calibration.

d_2, that are not multiples of $\lambda/2$, corresponding to reflection coefficients[1] $\Gamma_{L2} = -1$ and $\Gamma_{L3} = -1/[4(d_2 - d_1)]^2 e^{j4\pi(d_2-d_1)/\lambda}$.

The free-space calibration method is straightforward, requires no additional hardware and inherently captures all the antenna imperfections, such as its delay and spurious reflections from nearby objects. However, it still requires external components and precise placement of a reflector, which is impossible in many situations. For example, in [145] the sensor waveguides and antenna are built in the chassis of a steam turbine, making the use of a reflector difficult. As a result, a built-in calibration method that would minimize or completely avoid the use of external hardware is desirable.

5.6.3. Built-in Calibration

The key idea for implementing a built-in calibration method in the heterodyne transceiver, without significant extra hardware, is to take advantage of the reference receiver. Consider the modified heterodyne transceiver of figure 5.20. The reference channel is now an exact duplicate of the main channel, but instead of an antenna, it is terminated on a tunable load. This load can be characterized externally, as a separate circuit, and the exact values of the reflection coefficients it can generate will be known. The reference channel in the transceiver can switch between three load values Γ_{L1}, Γ_{L2}, Γ_{L3} and perform an one-port calibration, according to equations (5.56) and (5.57).

Nonetheless, the question arises of which reference is appropriate to compare the output of the calibration receiver V_{CAL} with in order to measure the phase shifts between the different loads. To answer this, equations (5.56) and (5.57) need to be revisited and examine what

[1] With this calibration, the distances will be calculated relative to d_1. Distances will still be accurate if $\Gamma_{L3} = -e^{j4\pi(d_2-d_1)/\lambda}$

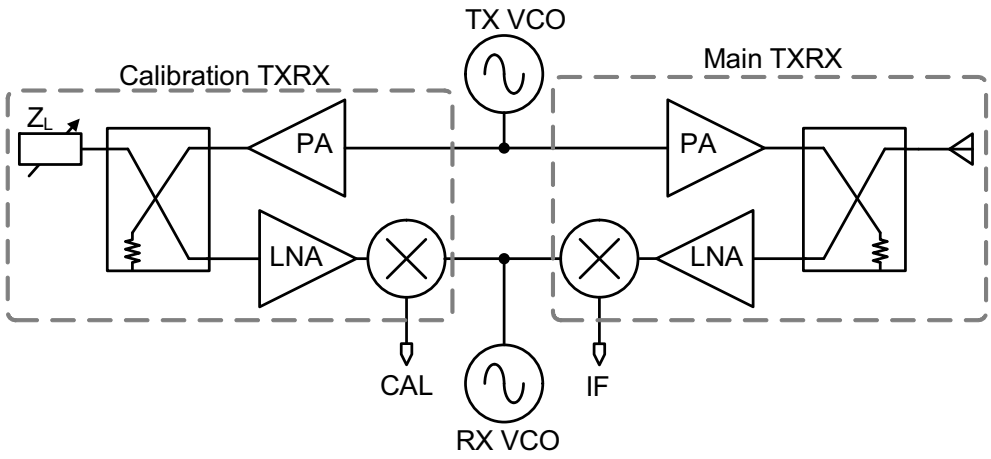


Figure 5.20: Heterodyne transceiver with built-in calibration.

happens if an arbitrary reference is used.

Assume that the reflection coefficients Γ_{M1}, Γ_{M2}, Γ_{M3} that correspond to the known Γ_{L1}, Γ_{L2}, Γ_{L3} are measured with an arbitrary reference V_{ref}. By definition, $\Gamma_{\mathrm{Mn}} = V_{r,\mathrm{Mn}}/V_{i,\mathrm{Mn}}$, where $V_{r,\mathrm{Mn}}$ and $V_{i,\mathrm{Mn}}$ are the reflected and incident voltages respectively and $n = 1, 2, 3$. Γ_{Mn} can be rewritten as:

$$\Gamma_{\mathrm{Mn}} = \frac{V_{r,\mathrm{Mn}}}{V_{\mathrm{ref}}} \cdot \frac{V_{\mathrm{ref}}}{V_{i,\mathrm{Mn}}} = \widetilde{\Gamma}_{\mathrm{Mn}} \cdot \frac{V_{\mathrm{ref}}}{V_{i,\mathrm{Mn}}} = \widetilde{\Gamma}_{\mathrm{Mn}} \cdot \frac{V_{\mathrm{ref}}}{V_i} \tag{5.58}$$

where $\widetilde{\Gamma}_{\mathrm{Mn}} = V_{r,\mathrm{Mn}}/V_{\mathrm{ref}}$ is the Γ_{Mn} referred to V_{ref}. Note that the assumption that $V_{i,\mathrm{Mn}} = V_i$, i.e. that all reflection coefficients are measured under the same transmitter conditions was made. Solving for $\widetilde{\Gamma}_{\mathrm{Mn}}$ yields:

$$\widetilde{\Gamma}_{\mathrm{Mn}} = \Gamma_{\mathrm{Mn}} \cdot \frac{V_i}{V_{\mathrm{ref}}} \tag{5.59}$$

Replacing Γ_{Mn} from (5.58) to (5.56):

$$\begin{bmatrix} \Gamma_{\mathrm{L1}} & 1 & -\Gamma_{\mathrm{L1}}\widetilde{\Gamma}_{\mathrm{M1}} \\ \Gamma_{\mathrm{L2}} & 1 & -\Gamma_{\mathrm{L2}}\widetilde{\Gamma}_{\mathrm{M2}} \\ \Gamma_{\mathrm{L3}} & 1 & -\Gamma_{\mathrm{L3}}\widetilde{\Gamma}_{\mathrm{M3}} \end{bmatrix} \begin{bmatrix} \mathrm{a}\frac{V_i}{V_{\mathrm{ref}}} \\ \mathrm{b}\frac{V_i}{V_{\mathrm{ref}}} \\ \mathrm{c} \end{bmatrix} = \begin{bmatrix} \widetilde{\Gamma}_{\mathrm{M1}} \\ \widetilde{\Gamma}_{\mathrm{M2}} \\ \widetilde{\Gamma}_{\mathrm{M3}} \end{bmatrix} \tag{5.60}$$

Defining $\tilde{\mathrm{a}} = \mathrm{a}V_i/V_{\mathrm{ref}}$ and $\tilde{\mathrm{b}} = \mathrm{b}V_i/V_{\mathrm{ref}}$, the above equation can be solved:

$$\begin{bmatrix} \tilde{\mathrm{a}} \\ \tilde{\mathrm{b}} \\ \mathrm{c} \end{bmatrix} = \begin{bmatrix} \Gamma_{\mathrm{L1}} & 1 & -\Gamma_{\mathrm{L1}}\widetilde{\Gamma}_{\mathrm{M1}} \\ \Gamma_{\mathrm{L2}} & 1 & -\Gamma_{\mathrm{L2}}\widetilde{\Gamma}_{\mathrm{M2}} \\ \Gamma_{\mathrm{L3}} & 1 & -\Gamma_{\mathrm{L3}}\widetilde{\Gamma}_{\mathrm{M3}} \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\Gamma}_{\mathrm{M1}} \\ \widetilde{\Gamma}_{\mathrm{M2}} \\ \widetilde{\Gamma}_{\mathrm{M3}} \end{bmatrix} \tag{5.61}$$

To examine the usefulness of the new calibration coefficients $\tilde{a}$ and $\tilde{b}$, equation (5.57) can be manipulated as follows:

$$\Gamma_L = \frac{\Gamma_M - b}{a - c\Gamma_M} = \frac{\Gamma_M \frac{V_i}{V_{ref}} - b\frac{V_i}{V_{ref}}}{a\frac{V_i}{V_{ref}} - c\Gamma_M \frac{V_i}{V_{ref}}} = \frac{\widetilde{\Gamma}_M - \tilde{b}}{\tilde{a} - c\widetilde{\Gamma}_M} \tag{5.62}$$

Therefore, absolute knowledge of a and b is not strictly necessary. It suffices if all the measurements are referenced to an arbitrary V_{ref}, since the absolute Γ_L can still be calculated.

An obvious choice for V_{ref} in the system of figure 5.20 is the output of the main receiver V_{IF}. Suppose that there is no reflector present during the measurement of $\widetilde{\Gamma}_{M1}$, $\widetilde{\Gamma}_{M2}$ and $\widetilde{\Gamma}_{M3}$, i.e., the antenna is pointing to the free space. Then, at a first step, the modified calibration coefficients are calculated from equation (5.61). At a second step, observing that

$$\widetilde{\Gamma}_M = \Gamma_M \cdot \frac{V_i}{V_{ref}} = \frac{V_{IF}}{V_i} \cdot \frac{V_i}{V_{IF}} = 1 \tag{5.63}$$

and from equation (5.62), Γ_L simply becomes:

$$\Gamma_L = \frac{1 - \tilde{b}}{\tilde{a} - c} = \Gamma_{ant} \tag{5.64}$$

Therefore, not only the modified calibration coefficients $\tilde{a}$, $\tilde{b}$ and c can be found, but also the antenna reflection coefficient Γ_{ant}.

For regular measurements, i.e. when a target is present, the tunable load is set to the calculated Γ_{ant} and the signal from the calibration channel becomes the reference signal for the main channel. The output V_{IF} is compared to V_{CAL} to measure $\widetilde{\Gamma}_M$ and the calibrated Γ_L is calculated from equation (5.62). The calculated Γ_L needs to be further corrected for the imperfect antenna reflection coefficient Γ_{ant}:

$$\Gamma'_L = \frac{\Gamma_L}{1 + \Gamma_L \Gamma_{ant}} \tag{5.65}$$

The proposed calibration procedure has several advantages. The impairments of the transceiver can be calibrated out without using any external component. The only requirement is for the antenna to point to free space during calibration. Furthermore, the antenna reflection coefficient Γ_{ant} is calculated during the calibration procedure, which can be used to correct Γ_L.

It should be emphasized that this method relies on the perfect matching between the calibration and measurement channels. Any mismatch due to process variation or layout imperfections between the two will lead to additional errors that cannot be calibrated out.

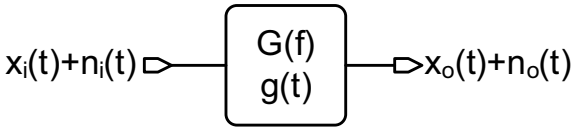


Figure 5.21: Filtering system

Layout asymmetries can be factored out with careful layout but process variation can still present a challenge. Moreover, the tuner in the calibration load needs to be characterized independently as a separate breakout and any variation between the pre-characterized tuner and the one in the TXRX will also lead to errors.

5.7. Summary

The different approaches for implementing a precision distance sensor have been presented and analyzed. First, frequency stepping was selected as the waveform of choice due to its simplicity and potential for high accuracy due to its inherent immunity to errors in the generation of the modulation waveform. Second, the different transceiver architectures that are suitable to implement the radar sensor were analyzed with respect to their resulting signal to noise ratio. The heterodyne architecture was proposed as a solution to the 1/f noise problem and was shown to achieve the highest SNRs. Third, the impairments in the RF front-end that lead to systematic errors in the frequency stepping method were analyzed and a new method for built-in calibration was proposed.

Implementation examples of the IQ heterodyne architecture and of the heterodyne with built-in calibration will be presented in the following chapter.

5.A. Appendix: Estimation of Signals Under Noise

5.A.1. Single Versus Paired Measurements

Consider the simple system of figure 5.21, where the input to a Linear Time Invariant (LTI) system with impulse response $g(t)$ and transfer function $G(f)$, is composed of a useful signal $x_i(t)$ corrupted by a random noise signal $n_i(t)$. The system outputs are assumed to be $x_o(t)$ and $n_o(t)$.

Generally, $x_i(t)$ is the output of a noisy sensor that needs to be measured and the aim of the LTI system $g(t)$ is to provide adequate filtering that will result in improving the signal-to-noise ratio (SNR) at its output compared to its input. This filtering action requires a certain amount of time τ_s which is referred to as measurement time or integration time.

There are two ways that the measurement can be performed. In the case of a singe measurement, the signal along with the noise pass through the filter $g(t)$ for time τ_s and the output $x_o(\tau_s) + n_o(\tau_s)$ is recorded. A second method, known as paired measurements, is to first record the output of the filter $n_o(t_0)$ when no useful input signal is present and after the filter has been acting for a sufficiently long time t_0. Following that, the useful signal x_i is introduced at t_0 and the output $x_o(t_0 + \tau_s) + n_o(t_0 + \tau_s)$ is recorded after time τ_s. The final result is formed by subtracting $x_o(t_0 + \tau_s) + n_o(t_0 + \tau_s) - n_o(t_0)$.

In all real-world systems, some form of paired measurement is necessary in order to subtract the effect of DC offsets and other non-idealities that would otherwise corrupt the input signal. The purpose of this section is to derive analytical formulas for the output SNR when paired measurements are considered for three popular filter transfer functions.

Most of the derivations in this appendix can be found in [161, 190, 191] and are repeated for completeness.

5.A.2. SNR Calculation

In the case of paired measurements, the useful input signal arrives at time t_0, and the output Δx_o is calculated after measurement time τ_s:

$$\Delta x_o = \big[x_o(t_0 + \tau_s) + n_o(t_0 + \tau_s)\big] - n_o(t_0) = x_o(t_0 + \tau_s) + \big[n_o(t_0 + \tau_s) - n_o(t_0)\big] \tag{5.66}$$

The signal-to-noise ratio is defined as:

$$\frac{S}{N} = \frac{x_o(t_0 + \tau_s)^2}{\sigma^2_{\Delta x_o}} \tag{5.67}$$

where $\sigma^2_{\Delta x_o}$ is the noise variance:

$$\sigma^2_{\Delta x_o} = \overline{\big[n_o(t_0 + \tau_s) - n_o(t_0)\big]^2} = \overline{n_o(t_0 + \tau_s)^2} + \overline{n_o(t_0)^2} - \overline{2n(t_0 + \tau_s)n_o(t_0)} \tag{5.68}$$

Assuming that $n(t)$ is *stationary* (which is true for both white and 1/f noise), $\overline{n(t_0 + \tau_s)^2} = \overline{n(t_0)^2} = \sigma^2_{no}$. Therefore,

$$\sigma^2_{\Delta x_o} = 2\sigma^2_{no} - \overline{2n_o(t_0 + \tau_s)n_o(t_0)} = 2\big[\psi(0) - \psi(\tau_s)\big] \tag{5.69}$$

where $\psi(\tau_s) = \overline{n_o(t_0 + \tau_s)n_o(t_0)}$ and as a result, $\psi(0) = \overline{n_o(t_0)n_o(t_0)} = \overline{n_o(t_0)^2} = \sigma^2_{no}$

Since $\psi(\tau_s)$ is a noise autocorrelation function, it can be expressed through the Wiener-Khintchine theorem:

$$\psi(\tau_s) = \int_0^\infty S_{no}(f)\cos(2\pi f\tau_s)\,\mathrm{d}f \tag{5.70}$$

where $S_{no}(f) = S_{ni}(f)|G(f)^2|$ is the noise power spectral density (PSD) at the output of the filter and $S_n(f)$ is the PSD at its input. Consequently,

$$\psi(\tau_s) = \int_0^\infty S_{ni}(f)|G(f)|^2 \cos(2\pi f \tau_s)\,\mathrm{d}f \tag{5.71}$$

and

$$\sigma^2_{\Delta x_o} = 2\int_0^\infty S_{ni}(f)|G(f)|^2\big[1-\cos(2\pi f\tau_s)\big]\,\mathrm{d}f \tag{5.72}$$

Substituting in (5.67) yields the final expression for the SNR [190]:

$$\frac{S}{N} = \frac{x_o(\tau_s)^2}{2\int_0^\infty S_{ni}(f)|G(f)|^2\big[1-\cos(2\pi f\tau_s)\big]\,\mathrm{d}f} \tag{5.73}$$

The factor $2\big[1-\cos(2\pi f\tau_s)\big]$ in the above equation arises because of the paired readings and is equal to unity for a single measurement.

5.A.3. DC Signal in the Presence of White Noise

The SNR at the output of the filter will be calculated when the input signal is DC with amplitude S_i and is corrupted by white noise with PSD $S_{ni}(f) = N_i$. Since the DC signal arrives at the input of the filter abruptly at time t_0, the output $x_o(\tau_s)$ after time τ_s corresponds to the step response.

5.A.3.1. Low-pass filtering

Suppose $G(f)$ is a low-pass filter with cut-off frequency $f_c = 1/\tau_c$:

$$G(f) = \frac{1}{1+j2\pi\tau_c f} \tag{5.74}$$

while

$$|G(f)|^2 = \frac{1}{1+(2\pi\tau_c f)^2} \tag{5.75}$$

The step response of the filter after time τ_s is:

$$x_o(\tau_s) = S_i(1-e^{-\tau_s/\tau_c}) \tag{5.76}$$

Substituting (5.75) and (5.76) in (5.73) yields [190]:

$$\frac{S}{N} = \frac{S_i^2(1-e^{-\tau_s/\tau_c})^2}{2\int_0^\infty \dfrac{N_i}{1+(2\pi\tau_c f)^2}\big[1-\cos(2\pi f\tau_s)\big]\,\mathrm{d}f} = \frac{2S_i^2(1-e^{-\tau_s/\tau_c})\tau_c}{N_i} \tag{5.77}$$

In order for the filter to settle to within 1% of its final value, $\tau_s = 4.6\tau_c$. Then the $1 - e^{-\tau_s/\tau_c}$ term in (5.77) vanishes:

$$\frac{S}{N} = \frac{2S_i^2\tau_c}{N_i} = \frac{0.43S_i^2\tau_s}{N_i} \tag{5.78}$$

Consequently, the SNR improves linearly as the filter cut-off frequency f_c is reduced and the measurement time τ_s increases

5.A.3.2. Integration

Suppose G is an integrator with integration time τ_i:

$$G(f) = \frac{1 - e^{-j2\pi f\tau_i}}{j2\pi f} \tag{5.79}$$

and

$$|G(f)|^2 = \frac{1 - \cos(2\pi f\tau_i)}{2\pi^2 f^2} \tag{5.80}$$

The step response of the integrator is simply

$$x_o(\tau_s) = S_i\tau_i \tag{5.81}$$

Substituting (5.80) and (5.81) in (5.73) and setting $\tau_s = \tau_i$ (i.e. the measurement time is the integration time) yields [190]:

$$\frac{S}{N} = \frac{S_i^2\tau_i^2}{2\int_0^\infty N_i \frac{\left[1-\cos(2\pi f\tau_i)\right]^2}{2\pi^2 f^2}\,\mathrm{d}f} = \frac{S_i^2\tau_i}{N_i} = \frac{S_i^2\tau_s}{N_i} \tag{5.82}$$

Comparing (5.82) with (5.78), it can be seen that for fixed measurement time τ_s the integrator output will have an approximately 2.3 times better SNR than the low-pass filter.

5.A.3.3. Averaging

A more realistic scenario for G would be to apply an analog anti-aliasing filter and collect N samples of the signal with rate $f_s = 1/T_s$. Subsequently, the DC signal can be estimated by averaging the samples, which is equivalent to calculating the DC component of the Discrete Fourier Transform. The transfer function of the averaging filter can then be expressed as:

$$G(f) = \frac{\sin(N\pi fT_s)}{N\sin(\pi fT_s)}\exp[-j(N-1)\pi fT_s]\,, \quad f < \frac{1}{2T_s} \tag{5.83}$$

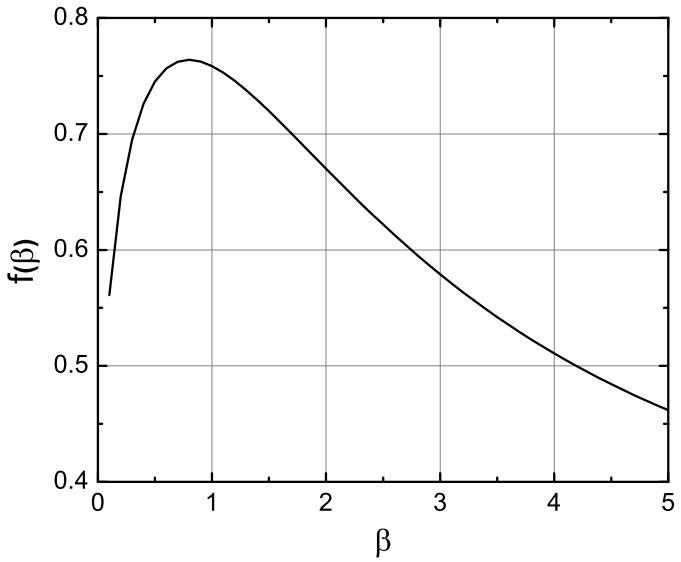


Figure 5.22: Plot of $f(\beta)$ versus β

and $G(f) = 0$ when $f > \frac{1}{2T_s}$ due to the anti-aliasing filter. However, the measurement time τ_s is related to the number of samples collected and the sampling period: $\tau_s = N\,T_s$ and as a result:

$$G(f) = \frac{\sin(\pi f \tau_s)}{N \sin(\pi f \tau_s / N)} \exp[-j(N-1)\pi f \tau_s / N]\,, \quad f < \frac{N}{2\tau_s} \tag{5.84}$$

and $|G(f)|^2$ becomes:

$$|G(f)|^2 = \frac{\sin(\pi f \tau_s)^2}{N^2 \sin(\pi f \tau_s / N)^2} \tag{5.85}$$

Substituting in (5.73) yields:

$$\frac{S}{N} = \frac{S_i^2}{2\displaystyle\int_0^{\frac{N}{2\tau_s}} N_i \frac{\sin(\pi f \tau_s)^2}{N^2 \sin(\pi f \tau_s / N)^2} \big[1 - \cos(2\pi f \tau_s)\big]\,\mathrm{d}f} = \frac{S_i^2 \tau_s}{N_i} \tag{5.86}$$

Consequently, the DFT averaging has the same SNR performance as the integrator.

5.A.4. DC Signal in the Presence of 1/f Noise

In this section, the SNR of the filter will be calculated when the input DC signal S_i is corrupted by 1/f noise with PSD $S_{ni}(f) = K/f$.

5.A.4.1. Low-pass filtering

For the 1/f noise case, equation (5.77) becomes [190]:

$$\frac{S}{N} = \frac{S_i^2(1-e^{-\tau_s/\tau_c})^2}{2\int_0^\infty \frac{K}{f}\frac{1}{1+(2\pi\tau_c f)^2}\big[1-\cos(2\pi f\tau_s)\big]\,\mathrm{d}f} = \frac{S_i^2}{2K}\cdot f(\beta) \tag{5.87}$$

where

$$f(\beta) = \frac{(1-e^{-\beta})^2}{\int_0^\infty \frac{1-\cos z}{z(1+z^2/\beta^2)}\mathrm{d}z} \tag{5.88}$$

while $\beta = \tau_s/\tau_c$ and $z = 2\pi f\tau_s$. $f(\beta)$ can only be calculated numerically and is depicted in figure 5.22. It attains a singe maximum value of approximately 0.76 when $\beta \simeq 0.8$, i.e. when $\tau_s = 0.8\tau_c$. Most importantly, increasing the measurement value τ_s beyond its maximum value will *decrease* the output SNR.

At the maximum of $f(\beta)$, the expression (5.87) for the SNR becomes:

$$\frac{S}{N} = \frac{0.38S_i^2}{K} \tag{5.89}$$

5.A.4.2. Integration

Similarly, the expression for 1/f noise in the integrator can be calculated [190]:

$$\frac{S}{N} = \frac{S_i^2\tau_i^2}{2\int_0^\infty K\frac{\big[1-\cos(2\pi f\tau_i)\big]^2}{2\pi^2 f^3}\,\mathrm{d}f} = \frac{S_i^2\tau_i^2}{4K\tau_i^2\ln 2} = \frac{S_i^2}{K\ln 16} = \frac{0.36S_i^2}{K} \tag{5.90}$$

The output SNR is independent of the measurement (integration) time. Furthermore, the SNR is essentially identical to the low-pass filter case.

5.A.4.3. Averaging

The DFT averaging under 1/f noise results in the following SNR:

$$\begin{aligned}\frac{S}{N} &= \frac{S_i^2}{2\int_0^{\frac{N}{2\tau_s}} \frac{K}{f}\frac{\sin(\pi f\tau_s)^2}{N^2\sin(\pi f\tau_s/N)^2}\big[1-\cos(2\pi f\tau_s)\big]\,\mathrm{d}f} \\ &\simeq \frac{S_i^2}{2\int_0^\infty \frac{K\sin(\pi f\tau_s)^2}{\pi^2 f^3\tau_s^2}\big[1-\cos(2\pi f\tau_s)\big]\,\mathrm{d}f} = \frac{S_i^2}{K\ln 16}\end{aligned} \tag{5.91}$$

where the assumption that N is large has been made, which is valid to within at least two digits when $N > 20$. Therefore, as with the white noise case, averaging results in the same SNR as the integrator.

5.A.5. DC Signal in the Presence of Combined Noise

When both white and 1/f noise are present, it is convenient to express the 1/f component using its noise corner f_{cn}, which is defined as the frequency where the 1/f noise is equal with the white noise component. Therefore, $K = N_i f_{cn}$ and the input noise PSD can be expressed as:

$$S_{ni} = N_i\Big(1 + \frac{f_{cn}}{f}\Big) \tag{5.92}$$

5.A.5.1. Integration

Since the integrator performs better than the low-pass filter in the case of white noise, only the integrator SNR will be considered. Equation (5.73) reduces to:

$$\frac{S}{N} = \frac{S_i^2\tau_i^2}{2\displaystyle\int_0^\infty N_i\Big(1 + \tfrac{f_{cn}}{f}\Big)\tfrac{\left[1-\cos(2\pi f\tau_i)\right]^2}{2\pi^2 f^2}\,\mathrm{d}f} = \frac{S_i^2\tau_i}{N_i + N_i f_{cn}\tau_i \ln 16} = \frac{S_i^2}{N_i/\tau_s + N_i f_{cn}\ln 16} \tag{5.93}$$

It is evident from equation (5.93) that in the limit of $\tau_i = \tau_s \to \infty$, the SNR is dominated only by the 1/f noise whereas when $f_{cn} \to 0$, white noise dominates.

5.A.5.2. Averaging

Similarly, the expression for DFT averaging calculates to:

$$\begin{aligned}\frac{S}{N} &\simeq \frac{S_i^2\tau_i^2}{2\displaystyle\int_0^\infty N_i\Big(1 + \tfrac{f_{cn}}{f}\Big)\tfrac{\sin(\pi f\tau_s)^2}{\pi^2 f^2\tau_s^2}\left[1-\cos(2\pi f\tau_s)\right]\mathrm{d}f} = \frac{S_i^2\tau_i}{N_i + N_i f_{cn}\tau_i\ln 16} = \\ &= \frac{S_i^2}{N_i/\tau_s + N_i f_{cn}\ln 16}\end{aligned} \tag{5.94}$$

5.A.6. 1/f Noise Singularity

If only a single measurement is performed in the presence of noise and signal, as opposed to the paired measurements of the previous sections, equation (5.73) simplifies to:

$$\frac{S}{N} = \frac{x_o(\tau_s)^2}{\displaystyle\int_0^\infty S_{ni}(f)|G(f)|^2\,\mathrm{d}f} \tag{5.95}$$

In the case of 1/f input noise and low-pass filtering the above equation becomes:

$$\frac{S}{N} = \frac{S_i^2(1-e^{-\tau_s/\tau_c})^2}{2\int_0^\infty \frac{K}{f}\frac{1}{1+(2\pi\tau_c f)^2}\,\mathrm{d}f} \tag{5.96}$$

Nevertheless, the integral $\int_0^\infty \frac{K}{f}\frac{1}{1+(2\pi\tau_c f)^2}\,\mathrm{d}f$ is singular and grows to infinity as $f \to 0$, preventing any safe conclusion to be reached about the SNR (the singularity persists in the case of the integrator as well). In order for the equation (5.96) to be evaluated, the assumption that the 1/f noise *flattens* close to zero frequencies is necessary. This can be expressed as:

$$\frac{S}{N} = \frac{S_i^2(1-e^{-\tau_s/\tau_c})^2}{2\int_0^\infty \frac{K}{\gamma+f}\frac{1}{1+(2\pi\tau_c f)^2}\,\mathrm{d}f} \tag{5.97}$$

where γ is a small, non-zero number. Ignoring the $1-e^{-\tau_s/\tau_c}$ factor ($\tau_s = 4.6\tau_c$) the SNR evaluates to:

$$\frac{S}{N} = \frac{S_i^2}{N_i}\frac{2+8\gamma^2\pi^2\tau_c^2}{2K\gamma\pi^2\tau_c + K\ln\left(\frac{1}{4\gamma^2\pi^2\tau_c^2}\right)} \tag{5.98}$$

Interestingly, the SNR is now sensitive to the measurement time and increases unboundedly as τ_s and τ_c increase. However, flattening of the 1/f noise spectrum has never been observed experimentally [192, 193] and consequently only the expressions developed in the previous sections will be considered valid.

5.A.7. Sinusoidal Signals

In the case the input signal is a sinusoid with frequency f_0 and amplitude S_i. The easiest method to improve the SNR would be collect N samples with rate $f_s = 1/T_s = 2f_0$, calculate the Discrete Fourier Transform (DFT) and estimate the sinusoid from the coefficient $k = \mathrm{round}(Nf_0/f_s)$. This essentially equivalent to the averaging that was employed for DC signals and has a system transfer function:

$$G(f) = \frac{\sin[N\pi(f-f_0)T_s]}{N\sin[\pi(f-f_0)T_s]}\exp[-j(N-1)\pi(f-f_0)T_s]\,, \quad f < \frac{1}{2T_s} \tag{5.99}$$

Since the PSD of white noise is flat and independent of frequency, the value of f_0 is irrelevant and the SNR is the same with the DC case:

$$\frac{S}{N} = \frac{S_i^2\tau_s}{N_i} \tag{5.100}$$

For combined white and 1/f noise, the analysis becomes more involved and the result will

depend on the corner frequency f_{cn}. However, in all cases we are interested in, the frequency of the sinusoid will be higher than the noise corner and consequently, the above formula will suffice.

6 Implementation of 122 GHz and 145 GHz Distance Sensors

This chapter presents two distance sensors, at 122 and 144 GHz respectively. The 122 GHz sensor is based on the proposed IQ heterodyne architecture whereas the 144 GHz follows the heterodyne with built-in calibration. In both cases, the transceivers are based on *fundamental* frequency oscillators and represent the first such transceivers reported in the literature operating above 100 GHz.

In both cases, the proposed circuits attempt to minimize the power consumption, in order to render the sensors feasible for use in portable systems. This is achieved by eliminating all the circuit topologies operating from 2.5 V or higher, commonly found in bipolar designs.

The overall system cost, which is an important consideration for mm-wave systems, is usually overwhelmed by the complicated package and the test equipment. Solutions to these problems will be proposed by using a superstrate antenna along with an on-chip feed at 122 GHz and by employing several self-test functionalities in addition to the built-in calibration in the 145 GHz transceiver.

The IC technology of choice for circuit implementation is the 130 nm SiGe BiCMOS process from STMicroelectronics, whose details have been presented in sections 2.1.3, 2.2 and 2.3. The technology features f_{max} of 280 GHz, adequate for circuit operation up to approximately 160 GHz [13], along with high Q passives.

The chapter is structured as follows. The 122 GHz system architecture is covered in Section 6.1. A detailed description of the transistor-level design considerations and the simulated performance of each block are provided in Section 6.2. The experimental verification of the circuit breakouts and of the packaged chip, including radar experiments over a distance of up to 2.1 m are described in Section 6.4. Section 6.5 presents the system and circuit details of the 144 GHz sensor and section 6.6 the associated measurement results.

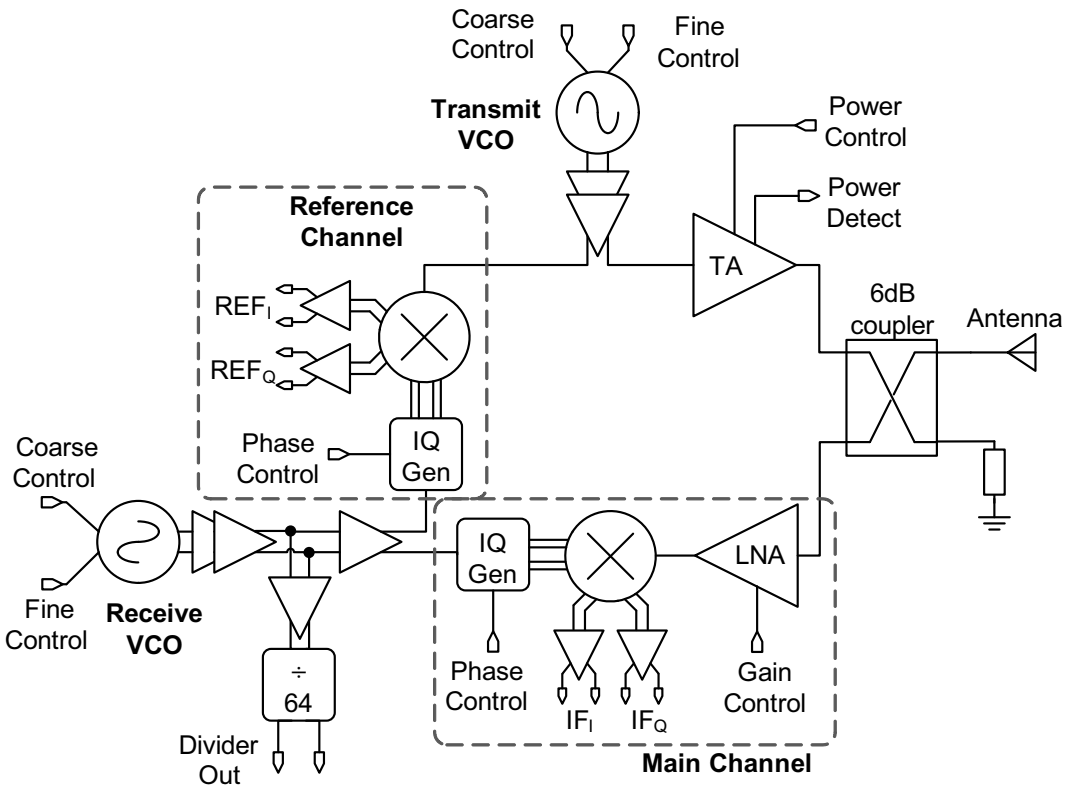


Figure 6.1: Block diagram of the transceiver.

6.1. 122 GHz Sensor Architecture

The block diagram of the proposed monostatic 122-GHz transceiver is illustrated in figure 6.1. Two fundamental frequency Voltage Controlled Oscillators (VCOs), operating at a frequency difference of up to 5 GHz, are employed to generate the transmit and receive LO signals, allowing for an adjustable receiver output frequency. The signal from the receive VCO is distributed to the two IQ mixers and to a divide-by-64 chain whose output, at approximately 1.9 GHz, is provided to an off-chip PLL that locks the VCO to a stable frequency reference.

The local transmit and receive VCO outputs are also downconverted by a second quadrature receiver. The reference path IQ downconverter is identical to the main receiver, apart from the low noise amplifier, which has been omitted since the amplitude of the two multiplied signals are large enough for the mixer noise to be irrelevant. The reference receiver serves a dual purpose. First, it can be used to lock the transmit VCO in an external PLL. Second, it provides the phase reference for the main receiver (section 5.5.3).

The primary consideration for selecting the IQ heterodyne architecture over the simple heterodyne, despite its increased complexity, has been the minimum realizable IF frequency of the system. Specifically, when the two VCOs operate at a sufficiently small frequency

offset, they will injection lock each other, due to the finite isolation between them, forcing themselves to oscillate at precisely the same frequency. In the system of figure 6.1, this would result in the transceiver turning into a homodyne radar, with adverse impact on the SNR.

Because of the lack of previous experience on mm-wave systems with two VCOs, it was not known in advance at which offset frequency the injection locking will happen. If it occurred at even large frequency offsets, the minimum realizable IF frequency might have been too high for the chip package to handle. Even worse, if the VCOs injection lock when the TX VCO is operating at the higher end of its frequency tuning curve, while the RX VCO is at its lower end (or vice versa), it would be impossible for the system to operate at any finite IF frequency in general. Consequently, to account for the case of having to operate as a homodyne system, IQ mixers were implemented in both receivers. As discussed in section 5.5.3 having IQ receivers in heterodyne operation, will also result in improving the SNR of the system by a factor of two and as a result, if injection locking is avoided, having IQ receivers will not plague the system performance. However, this SNR advantage by itself would hardly justify the increased complexity and power dissipation of IQ downconversion in the RF path.

Due to imperfect antenna, LNA input, and Transmit Amplifier (TA) output matching, as well as because of imperfections in the antenna coupler, a portion of the transmitted signal will leak into the receiver, potentially desensitizing it. After conducting an analysis of different microstrip structures for the directional coupler, in which realistic values were assumed for the reflection coefficients of the circuit blocks connected to the coupler, and the process variation of the termination resistor was accounted for, it was concluded that the isolation between the output of the transmit amplifier and the input of the receiver will not be greater than 20-25 dB. As a result, it was decided to limit the transmitter output power to 0 dBm. Therefore, to ensure that the receiver performance is not impaired by the leakage of the transmitter, the receiver input compression point has to be better than -20 dBm. Both the transmitter output power and the receiver input compression point are relatively easy to satisfy using silicon technology at 122 GHz. To provide adequate margin in case of even poorer isolation, output power control in the transmit amplifier and gain control in the receiver LNA were introduced.

6.2. 122 GHz Circuit Design

Unlike all other D-Band SiGe HBT receivers, transmitters and transceivers reported to date, which operate from 3.3 V or higher supplies and consume 1.5 W or more, one of the most important goals of this design was to reduce power consumption below 1 W, as needed

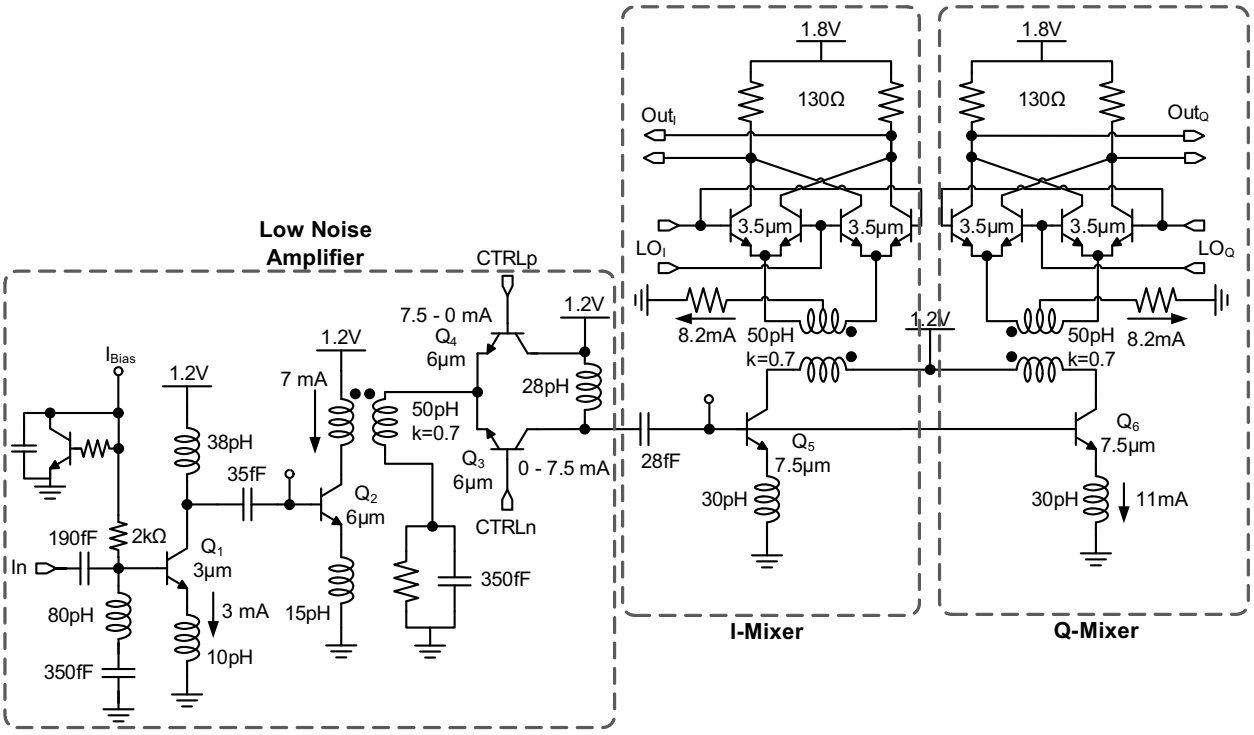


Figure 6.2: Schematic of the quadrature receiver.

for a portable system. As a result, a decision was made from the outset to use only circuit topologies that operate from 1.2 V or 1.8 V supplies. Telescopic cascodes were intentionally avoided and replaced with capacitively or transformer-coupled cascodes or common-emitter topologies when necessary.

6.2.1. IQ Receiver

The schematic of the IQ receiver is illustrated in figure 6.2. The input signal is first amplified by a three-stage LNA followed by common-emitter transistors Q_5 and Q_6 that double up as transconductors for the I and Q mixers and as an active power splitter. Transformer-coupling is employed in the Gilbert-cell mixers that downconvert the 120-GHz RF signal to a low IF frequency. Highly linear, unity-gain, 50-Ω IF buffers (not shown in the figure) are used as an interface to the external, 50-Ω environment.

In the system architecture under consideration, the receiver input needs to be well-matched to 50 Ω in order for the 6-dB coupler to be terminated symmetrically. As demonstrated in section 2.2, the penalty in noise figure when matching directly for maximum power gain $G_{\max}$ is relatively low, approximately 0.7 dB. As a result, matching for power ($G_s + jB_s = G_{og} + jB_{og}$) was preferred to using inductive feedback to achieve simultaneous

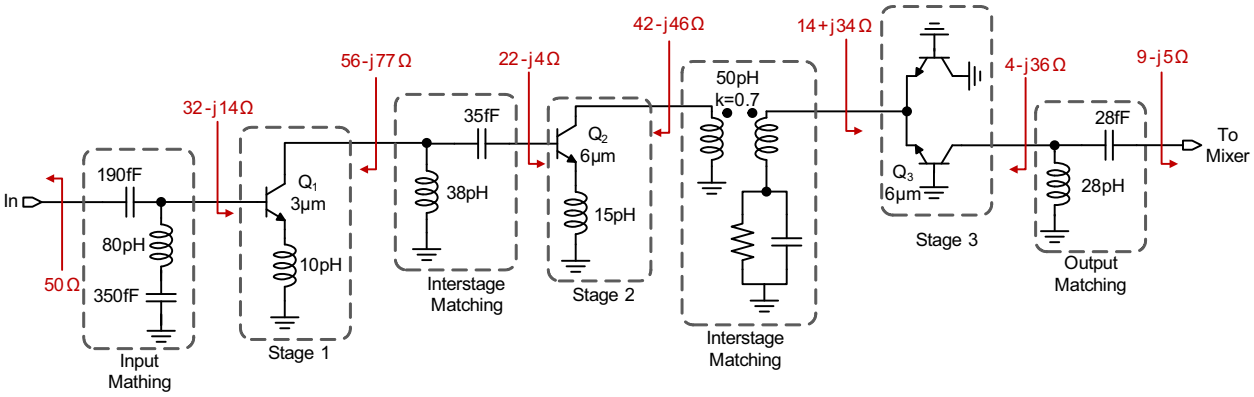


Figure 6.3: Small signal equivalent circuit of the LNA.

matching for noise and power. The first stage of the LNA is biased for minimum noise measure while the second and third stages are biased progressively at slightly higher current densities in order to increase the amplifier gain with negligible impact on the noise figure.

Gain control is implemented in the third, common-base, stage of the LNA by steering the bias current between transistor Q_3, whose output drives the tuned L-C load, and Q_4, whose collector is connected directly to the power supply. The input impedance of Q_3 in parallel with Q_4 remains constant as the current is steered, ensuring that the S_{11} of the LNA does not vary significantly with gain control.

Figure 6.3 depicts a detailed small-signal equivalent circuit of the LNA. The amplifier stages and associated matching networks are denoted in the figure along with the impedances for simultaneous conjugate matching at 122 GHz. All the impedance transformations, apart from those between the second, common-emitter, stage and the third, common-base, stage, were performed using simple L-matching networks [41]. A 1:1 transformer between stages two and three was preferred due to the fact that its secondary provides an immediate DC current path to ground for the emitter of Q_3.

Transistors Q_5 and Q_6, forming the power splitter, are of identical size and loading, guaranteeing that the signal applied at their bases is split equally and in-phase among the I and Q mixers and avoiding unwanted I-Q imbalance. Both the LNA and the transconductor cells operate with a 1.2 V supply.

Because the linearity of the receiver is limited by the mixer, special attention was paid to maximizing its input compression point. To achieve this, Q_5 and Q_6 are biased at the peak f_T current density of 1.5 mA/μm and their emitter length is set to 7.5 μm, more than two times larger than the size of Q_3. Similarly, the mixing quad transistors have an emitter length of 3.5 μm, forming a one-to-one folded cascode with Q_5 and Q_6 respectively. The input compression point of the mixer is further improved by inserting 30-pH degeneration

inductors at the emitters of Q_5 and Q_6 and by employing a 1.8 V supply for the mixing quad and IF buffers to maximize the linear output voltage swing.

The simulated conversion gain, input 1-dB compression point and double-sideband noise figure of each of the I and Q mixers (including signal splitting at Q_5 and Q_6) are 5 dB, -10 dBm, and 18.5 dB respectively, including the unity gain 50-Ω IF buffers.

Although the linearity of the mixer could, in theory, be improved by further increasing the size of Q_5 and Q_6 and the supply voltage of the mixing quads, the required power consumption becomes prohibitively high. As a result, the number of stages in the LNA was limited to three so that the receiver achieves the goal of -20 dBm input 1-dB compression point at maximum gain. Nevertheless, limiting the number of LNA stages does degrade the overall noise figure resulting in a power-noise-linearity trade-off where noise is traded for linearity.

The total power consumption of the receiver is 130 mW, including the 50-Ω IF buffers. The simulated downconversion gain, DSB noise figure and input compression point from the LNA input to either the I or the Q IF outputs are 13.5 dB, 12.5 dB and -20.5 dBm respectively.

6.2.2. I-Q Generation and Calibration

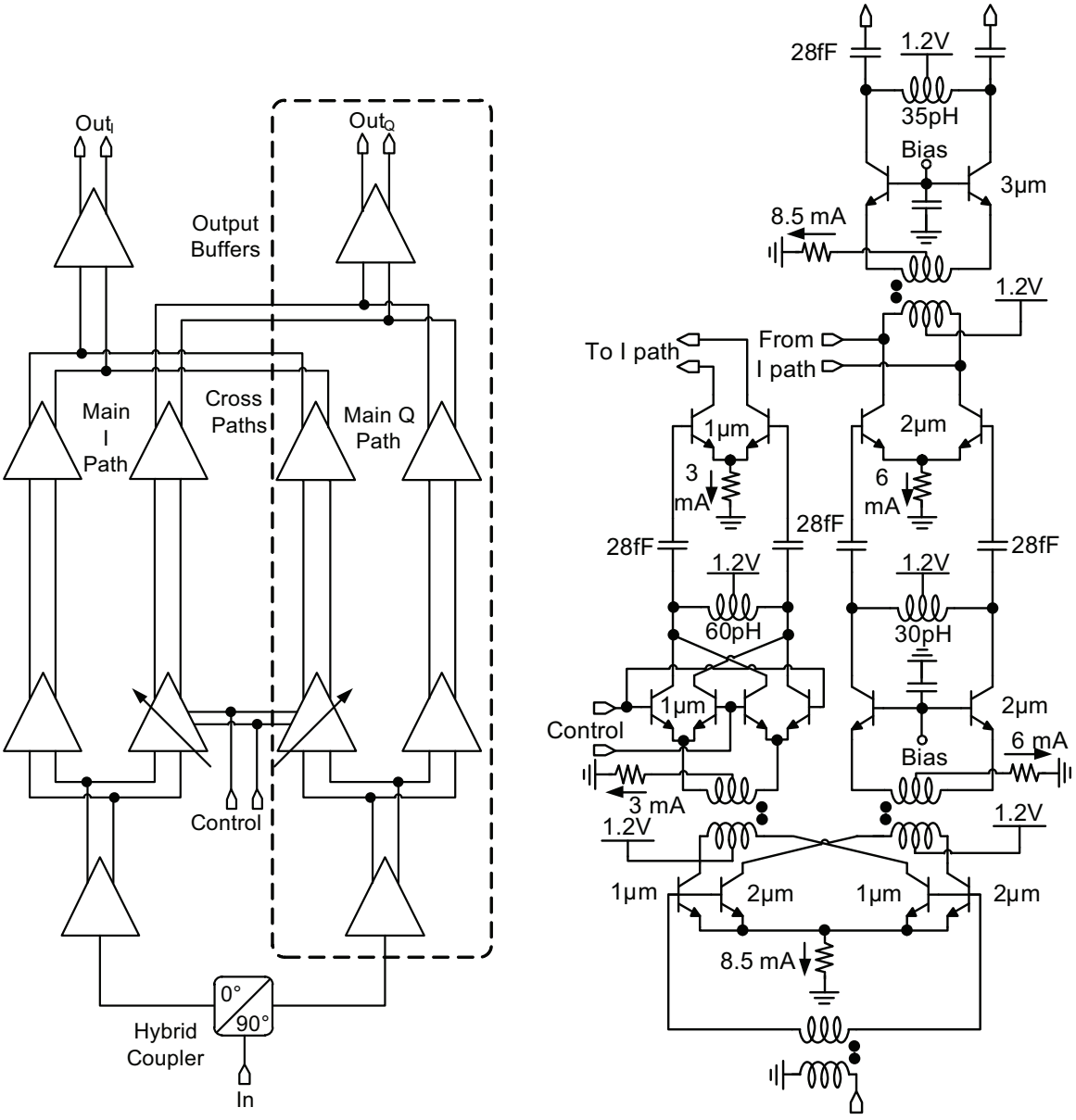
Due to possible modeling inaccuracies and layout asymmetries in the I-Q LO distribution circuits, phase calibration is necessary in order to guarantee that the receiver IF outputs are in quadrature.

Figure 6.4a illustrates the block diagram of the IQ phase correction circuit that splits the LO signal into quadrature outputs and allows for their phases to be calibrated. Calibration is performed by adding weighted I and Q cross paths to the main LO distribution path of opposite phase, i.e. the I cross path is added to the Q main path and vice versa. The operation of the phase correction circuits can be described in the phasor domain as:

$$V_{\text{out,I}} = V_I + kV_Q = A + kAe^{-j\pi/2} = A - jkA \tag{6.1}$$

$$V_{\text{out,Q}} = V_Q + kV_I = Ae^{-j\pi/2} + kA = kA - jA \tag{6.2}$$

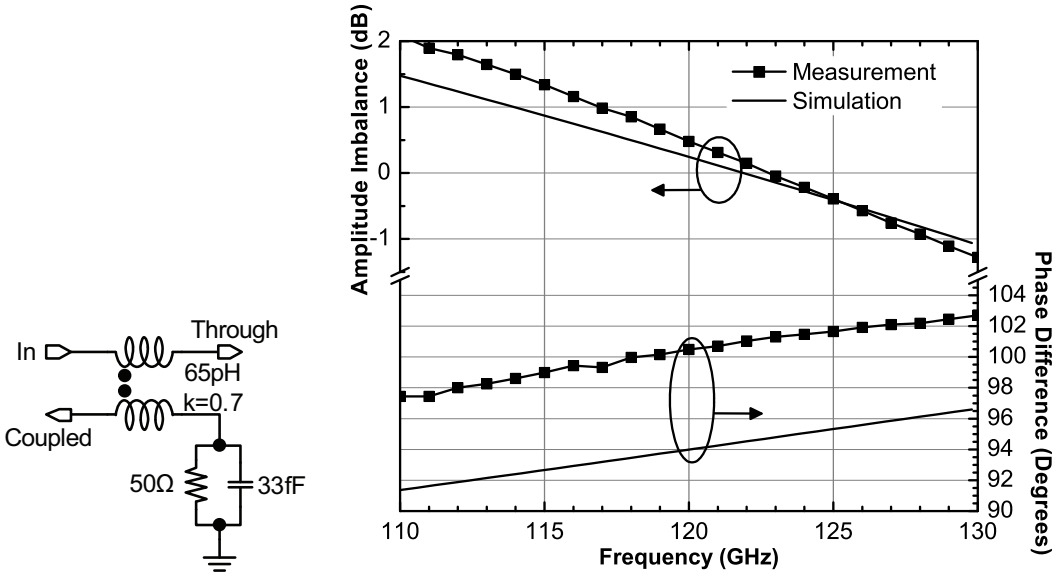
where A is the amplitude of the I and Q outputs before calibration (assumed equal) and k is the variable gain of the cross path. The amplitude imbalance and phase difference of the



(a) Block diagram.

(b) Schematic of one of the two parallel paths highlighted in the block diagram.

Figure 6.4: IQ generation and calibration circuits.



(a) Schematic. (b) Measured and simulated results.

Figure 6.5: Quadrature hybrid.

I and Q outputs are given by:

$$\left|\frac{V_{\text{out,I}}}{V_{\text{out,Q}}}\right| = \frac{\sqrt{A^2 + k^2A^2}}{\sqrt{k^2A^2 + A^2}} = 1 \tag{6.3}$$

$$\angle V_{out,I} - \angle V_{out,Q} = \tan^{-1}\frac{1}{k} - \tan^{-1} k \tag{6.4}$$

As a result, under ideal conditions, the amplitude ratio of the I and Q outputs remains constant and the phase difference between $V_{\text{out,I}}$ and $V_{\text{out,Q}}$ can be varied around 90°, according to the sign of k.

Figure 6.4b reproduces the transistor-level schematic of the Q path, highlighted in the block diagram of figure 6.4a. The input signal is first converted from single-ended to differential mode and is split into the main and cross paths using two parallel, differential, common-emitter amplifiers in a 2:1 size ratio. The output signal from each differential amplifier is coupled via 1:1 symmetrical baluns to a buffer amplifier on the main path, and to a Gilbert-cell Variable Gain Amplifier (VGA) on the cross path. The advantage of using a Gilbert-cell based VGA is that both variable gain and sign selection are possible. The output of the VGA is then buffered by a differential common-emitter stage before being added to the main I path. This buffering stage is absolutely necessary because the output impedance of the VGA varies with the gain-control setting and would otherwise lead to uncontrolled loading of the main I path, and ultimately to parasitic amplitude modulation. After adding the currents of the main and cross paths at the primary of yet another balun, the resulting

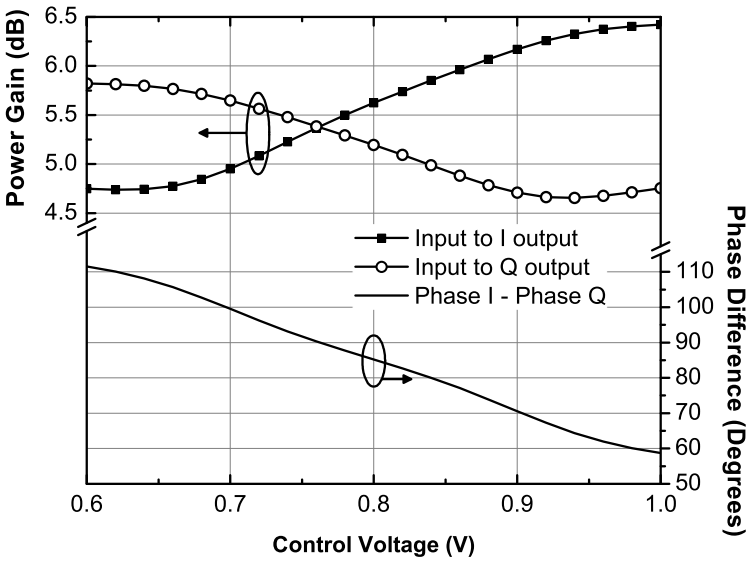
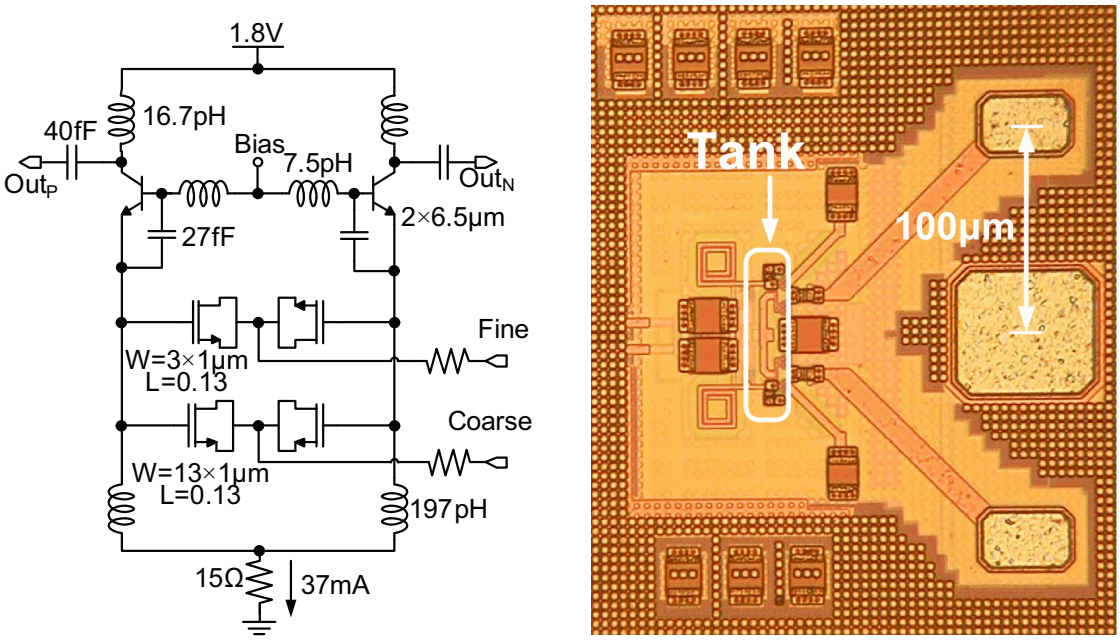


Figure 6.6: Simulated performance of the I-Q generation and calibration circuit at 122 GHz.

output signal is buffered by a 3 μm differential common-base stage to ensure that the output power is sufficiently large to fully switch the HBTs in the mixing quads.

As shown in figure 6.4a, the input signal to the phase correction circuit is first split into in-phase and quadrature paths by a lumped 90°-hybrid, illustrated in figure 6.5a, which is designed according to the methodology presented in [194]. Figure 6.5b reproduces the measured and simulated amplitude imbalance and phase difference between the I and Q outputs of the hybrid. To perform the S-parameters measurement, two separate test structures with on-chip 50-Ω terminations were employed. The 6° disagreement between measurement and simulation of the phase difference is attributed to limitations in modeling the lumped components, as well as to possible deviation of the on-chip terminations from their ideal 50-Ω value, indicating the necessity of calibration.

The simulated small signal power gain and phase difference at 122 GHz from the input to the I and Q outputs is illustrated in figure 6.6. A phase adjustment range of 60° to 110° is predicted. The absolute amplitude imbalance between the I and Q outputs is less than 1.8 dB for the entire phase adjustment range. This amplitude imbalance is not predicted by equation (6.3) and is due to non-idealities such as unequal delays between the main and cross paths, as well as due to variation of the VGA transmission phase as a function of its gain setting. The power consumption of the I-Q generation and calibration circuit is 92 mW from a single 1.2 V power supply.



(a) Schematic (b) Die microphotograph.

Figure 6.7: 122 GHz Colpitts VCO.

6.2.3. Voltage Controlled Oscillator

The schematic of the 120-GHz fundamental frequency VCO is shown in figure 6.7a and follows a differential Colpitts topology [195], a common choice for low-phase noise mm-wave VCOs [196–198]. In order to achieve low phase noise, a single transistor topology is employed. It has the added benefit of operation from a low-voltage supply, avoiding the use of a stacked common-base buffer [197, 199].

Leeson's phase noise model predicts that the phase noise at offset $\Delta\omega$ from the carrier frequency ω_0 can be calculated by:

$$L(\Delta\omega) = \frac{2Fk_BT}{P_s}\left[1 + \frac{\omega_0}{2Q\Delta\omega}\right] \tag{6.5}$$

where P_s is the signal power provided by the oscillator core to the tank, Q is the quality factor of the tank and F is a factor that depends on the noise added by the transistors. Inspecting the above equation reveals that to minimize the phase noise, the quality factor and the signal power, i.e. the voltage swing on the tank need to be maximized.

In order to optimize the tank quality factor, multifinger, double-side contacted, 130-nm accumulation-mode FET varactors with 1-μm wide gate fingers were employed. The measured quality factor of these varactors is 4-16 in the D-band [1], and is typically higher than that of pn-junction varactors [1, 198], while their measured $C_{\max}/C_{\min}$ ratio is 2:1 [1]

with voltage tuning range that is CMOS compatible (0 - 1.2 V).

For the case of the Colpitts VCO, the voltage swing on the tank can be calculated by [200]:

$$V_{\text{tank}} \simeq 2I_{\text{bias}}\left(\frac{C_1}{C_1 + C_2}\right)R_p \tag{6.6}$$

where I_{bias} is the oscillator bias current and R_p is the equivalent parallel resistance of the tank. As a result, to maximize the voltage swing and thus minimize phase noise, the HBT emitter length is set to the largest possible value that still permits a reasonable tank inductance value, 7.5 pH in this case. The HBTs are biased at their peak-f_T current density, resulting in the largest possible I_{bias}. Biasing at peak-f_T current density is associated with increased F in Leeson's formula (6.5). Lower current densities, closer to the minimum-F_{min}, are generally preferred at lower frequencies [196]. However, in this case, biasing at peak-f_T was found to yield the highest output power and the highest robustness in terms of satisfying the oscillation conditions, over supply and temperature variations, and was therefore selected.

To simplify the design of the external PLL and to minimize its noise contribution and spurs, coarse and fine control of the VCO frequency is provided by two groups of varactors. The grouping for coarse control has a total varactor size of $13 \times 1\,\mu\text{m} \times 0.13\,\mu\text{m}$ and can be adjusted once, at power up, to bring the VCO frequency close to the intended value. A $3 \times 1\,\mu\text{m} \times 0.13\,\mu\text{m}$ varactor is used for fine tuning as part of the PLL. This arrangement minimizes the VCO gain in the PLL. Single-ended controls were preferred in order to simplify the chip interface with the external loop filter, but are known to slightly degrade the VCO phase noise.

Optimization of the symmetrical VCO layout is crucial to achieve the intended oscillation frequency and tuning range at D-Band. The 7.5-pH tank inductors were formed as a single, 5-μm wide, inductive line, with the top two copper layers shunted together for a total thickness of $6\,\mu\text{m}$ in order to be able to precisely control the inductance. Figure 6.7b reproduces the die microphotograph of a VCO breakout.

The total power consumption of each VCO is 76 mW from 1.8 V power supply.

6.2.4. LO Distribution

Distributing the 120-GHz LO signal to a relatively large number of circuits without degrading its amplitude is a challenging task. Passive power splitting has been employed at lower frequencies (e.g. [6, 122, 201, 202]). However, such an approach would suffer from very high insertion loss which, ultimately, requires very powerful buffers to be placed at the output of the VCO. In order to avoid designing such buffers, an active power distribution solution was preferred, as illustrated in Figs. 6.8a and 6.8b for the cases of the receive VCO

and the transmit VCO, respectively.

The first set of buffers, placed immediately after the VCOs, employ the smallest transistor size in order to avoid overloading the VCO, which would endanger the oscillation condition. Furthermore, in order for the two VCOs to be loaded similarly and to oscillate in the same frequency range, these buffers are identical for both the TX and RX LO trees. The second-level buffers are scaled by a factor of two (4 μm) and operate close to their 1-dB output compression point, ensuring that their output power is sufficiently large to drive the subsequent stages in compression. The third set of buffers in the receiver LO tree have their inputs connected in parallel and the transistor size is selected such that the divider receives adequate input power to guarantee proper frequency division.

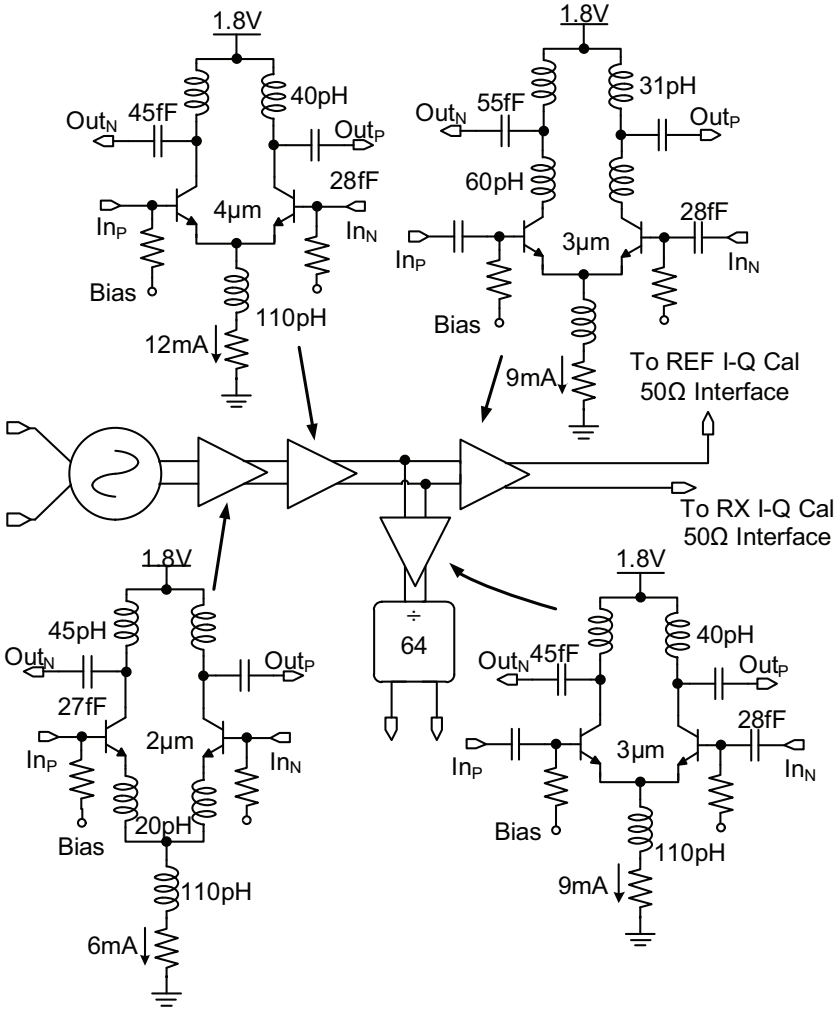
Conjugate power matching with L-sections was employed between buffer stages, whereas Π-matching networks were used at the 50-Ω interfaces (fig. 6.8a and 6.8b) since the corresponding L-sections would require very small capacitors and would result in very narrow bandwidth. Simple common-emitter amplifiers were preffered for the buffers over cascodes for two reasons. First, to minimize the power consumption and second, to avoid the detrimental effects of the increased sensitivity that cascodes exhibit to output matching. Common-mode inductors and resistors are introduced to increase the common-mode rejection ratio and improve the bias stability. All transistors in all buffers are biased at the peak-f_T current density.

The LO distribution network in the receiver delivers +4 dBm to the divider and +1 dBm to each of the I-Q generation and calibration circuits at 120 GHz, while consuming 90 mW from a single 1.8 V power supply. The transmitter LO distribution tree delivers +1 dBm to the transmit amplifier and to the reference downconverter while consuming 40 mW from 1.8 V power supply.

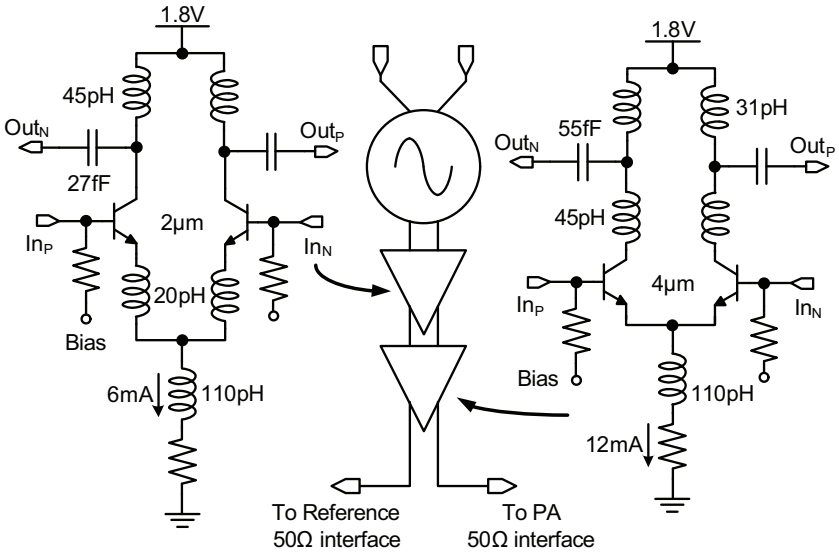
6.2.5. Transmit Amplifier

The schematic of the transmit amplifier is illustrated in figure 6.9a and is identical to that of the LNA, except for the power supply voltage of the last stage which was increased to 1.8 V in order to ensure that the output power is at least 0 dBm. The output network was modified to facilitate 50-Ω matching.

The power detector function is realized with a common-collector HBT biased at a low current density using a nFET current source [203]. The detector is AC-coupled through a 200-fF capacitor to the output of the amplifier. The smallest HBT size was selected in order to minimize the loading of the amplifier. The simulated power detector output voltage versus the output power of the transmit amplifier is reproduced in figure 6.9b.

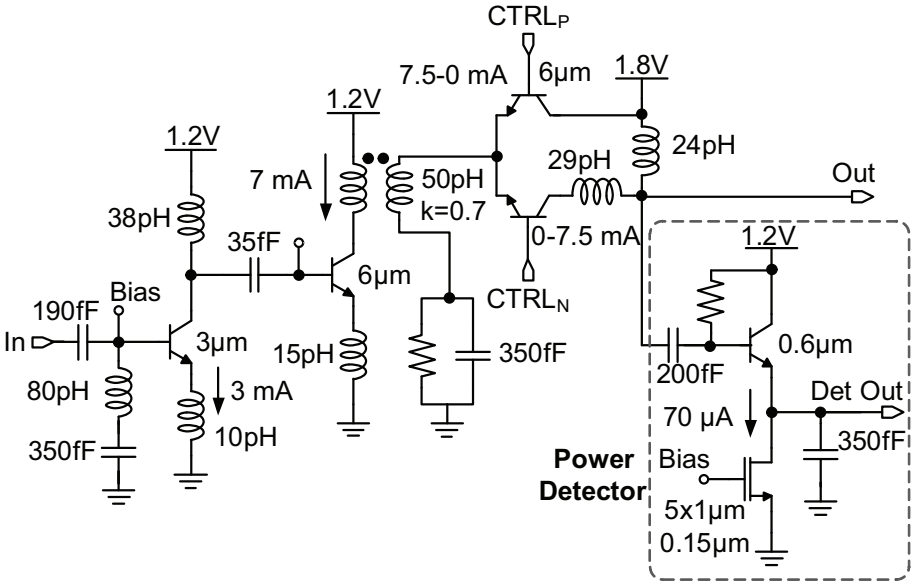


(a) Receive VCO signal distribution.

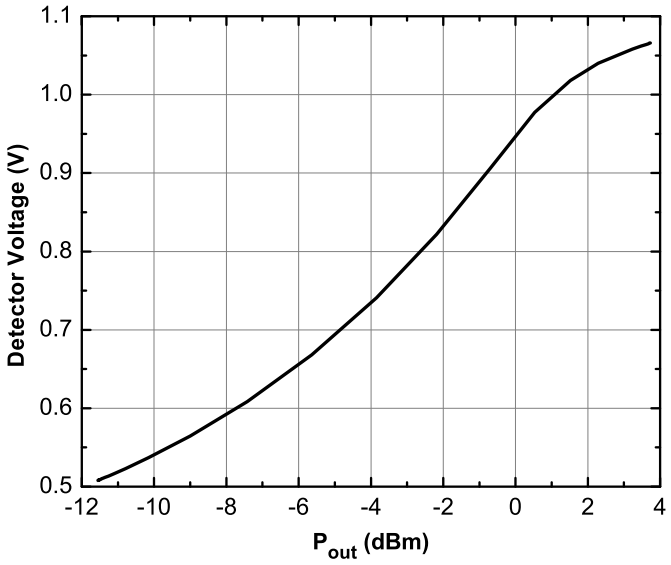


(b) Transmit VCO signal distribution.

Figure 6.8: LO distribution networks.

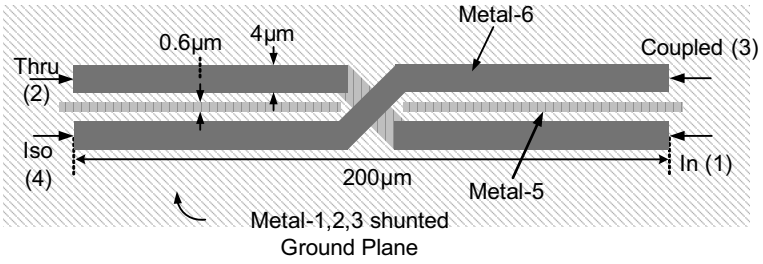


(a) Schematic of the variable gain transmit amplifier.

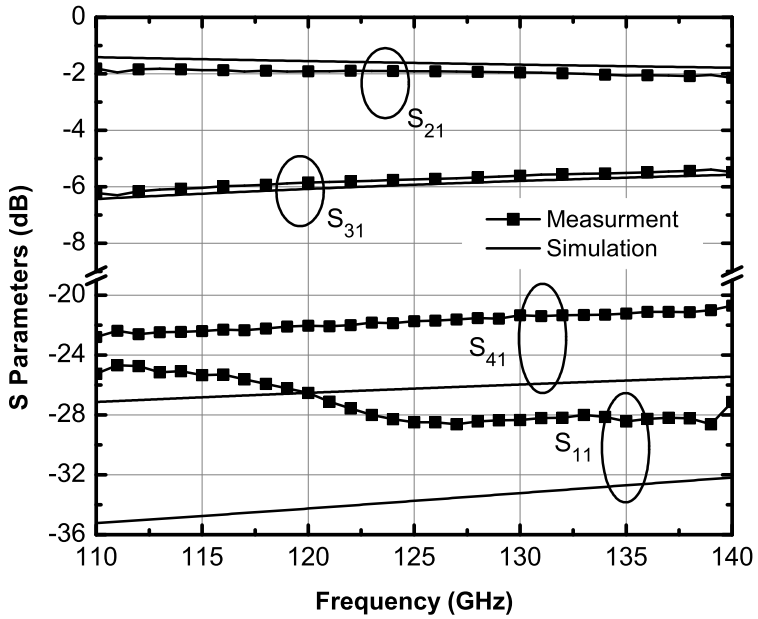


(b) Simulated power detector output voltage versus amplifier output power at 122 GHz.

Figure 6.9: Transmit amplifier schematic and simulation results.



(a) Sketch.



(b) Measured and simulated S-parameters.

Figure 6.10: 6-dB Coupler.

6.2.6. 6-dB Coupler

The sketch of the coupled-line, 6-dB antenna coupler is illustrated in figure 6.10a. The two coupled lines are realized in the top, 3-μm thick, metal layer while the ground plane is realized with the bottom three metal layers shunted together (figure 2.16a of section 2.3). In order to meet the required metal density rules, a floating bar in M5, located 1.5 μm below the top metal layer, is inserted between the two arms of the coupler. The width of the floating bar is chosen to improve the overall isolation of the coupler by optimizing the matching between the even and odd-mode velocities [175, 176].

A comparison between the measured and simulated S-parameters of the 6-dB coupler is reproduced in figure 6.10b. The 6-dB coupler was also characterized using two separate test structures with on-chip terminations. The loss of the "thru" arm remains below 2 dB at 120 GHz while the isolation and input reflection are better than 20 dB. Very good match-

Table 6.1: Power consumption breakdown of the 122 GHz sensor

Block	Count	Power (mW)
VCO	2	2×76
Divider chain	1	115
I-Q generation and calibration	2	2×92
Receiver	1	130
LO distribution	2	90 (RX), 40 (TX)
Reference generation	1	100
Transmit amplifier	1	35
Total (including bias)		**900**

ing between the simulated and measured "thru" and "coupled" transmission coefficients is achieved. The measured reflection coefficient and isolation slightly deviate from their simulated values. However, it is difficult to know if this is due to a design inaccuracy or because of deviations and mismatches of the on-chip terminations.

6.2.7. Power Dissipation

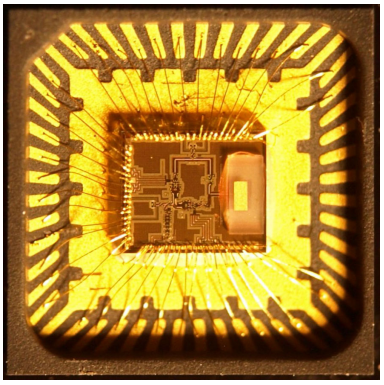
A breakdown of the 900 mW power consumption of the transceiver chip is presented in Table 6.1. Most of the power was consumed for generating, distributing and calibrating the LO signals which was necessary in order to ensure proper operation of the transceiver.

6.3. 122 GHz Antenna Design

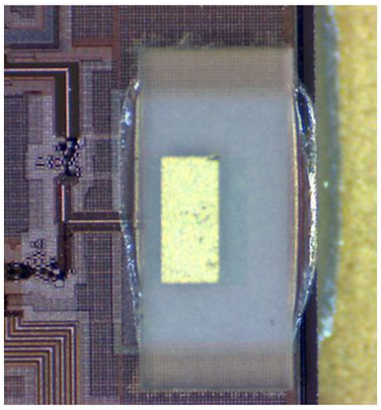
The antenna was designed by Jurgen Hasch from Robert Bosch GmbH and is a frequency scaled version of the antenna presented in [204]. It consists of a shorted patch on-die feed radiating from one side, electromagnetically coupled to an external $\lambda/2$ patch resonator, manufactured on a low-loss quartz superstrate. The electromagnetic coupling between the feed and the antenna averts the use of cumbersome mm-wave on-chip to off-chip transitions that are common in 60 and 77 GHz designs (e.g. in [205]). The chip is then mounted in a simple, open-lid QFN package with the external quartz antenna glued on top of the die, as illustrated in figure 6.11a. Figure 6.11b shows a close-up view of the resonator glued on the chip.

The antenna was simulated along with the package and the bondwires in a 3D EM-simulator. The gain and radiation pattern are reproduced in figure 6.12a. The simulated antenna gain and efficiency are 6 dBi and 50% respectively, including the mismatch loss. The simulated antenna 1-dB bandwidth of 9 GHz is shown in figure 6.12b.

Similar antenna solutions with comparable performance have also been reported at 94 GHz in [95, 206].

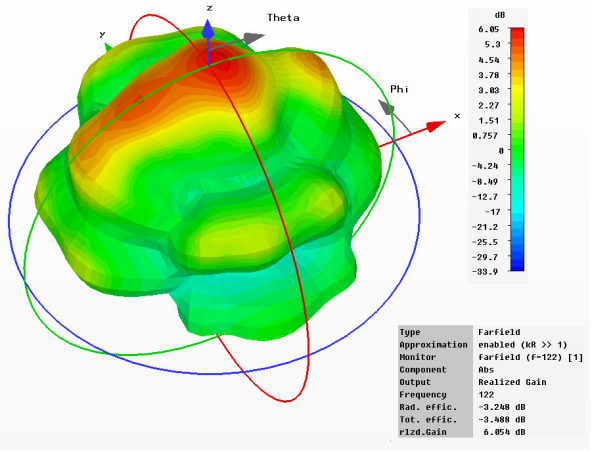


(a) Transceiver mounted in a QFN package.

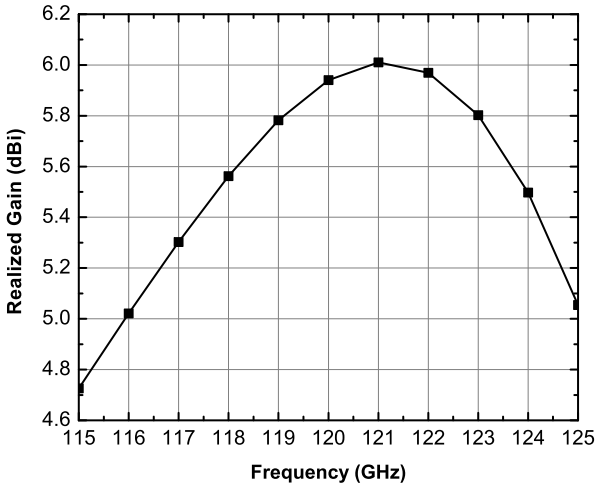


(b) Close up view of the quartz resonator.

Figure 6.11: Packaged transceiver.



(a) 3D radiation pattern.



(b) Simulated gain versus frequency, including mismatch losses.

Figure 6.12: Simulated performance of the antenna (courtesy of Jurgen Hasch).

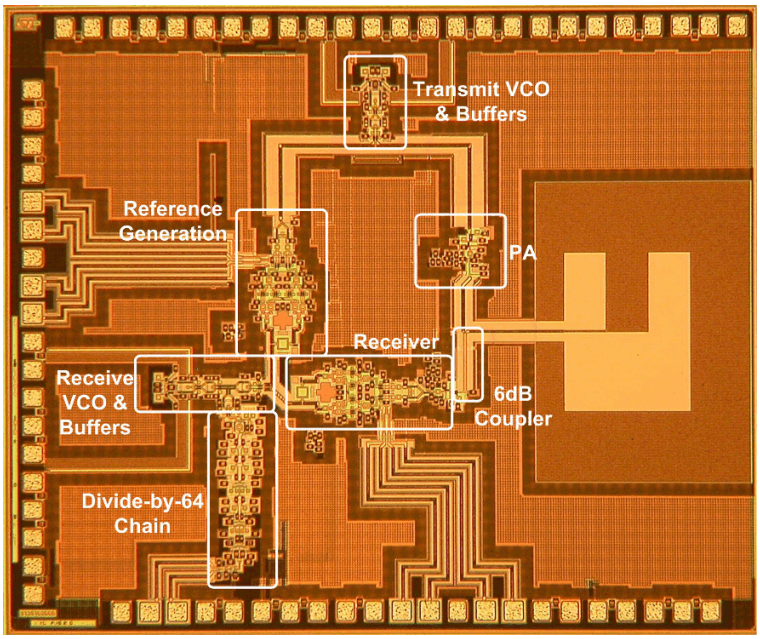


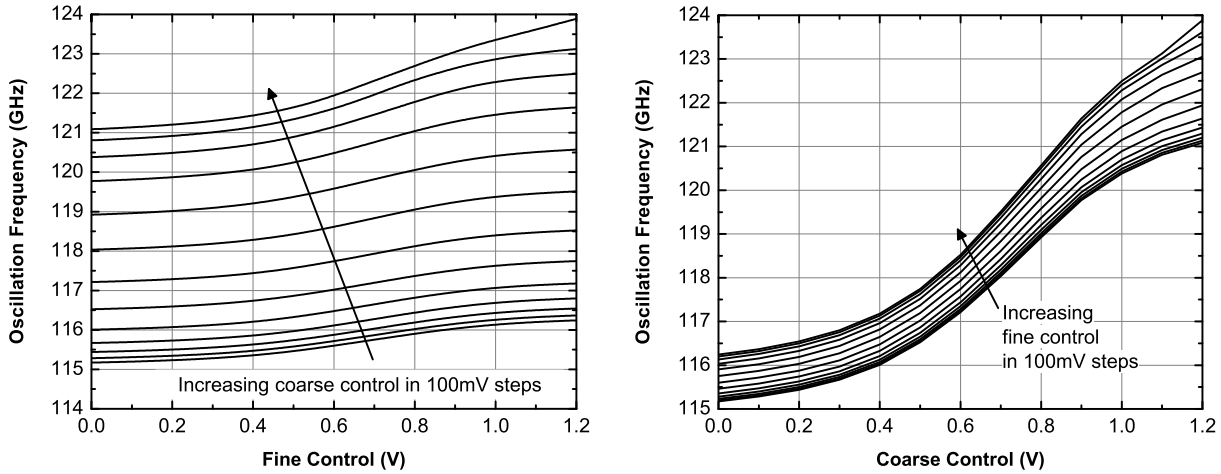
Figure 6.13: Die microphotograph. The dimensions are 2.2 mm×2.6 mm.

6.4. Characterization of the 122 GHz Sensor

Figure 6.13 shows a die microphotograph of the transceiver which occupies $2.2\,mm \times 2.6\,mm$. The physical layout methods of sections 2.4 and 2.5 have been extensively employed. In addition, the receiver, VCOs, and divider have been surrounded with ground shields to avoid noise coupling and their power supplies are provided from separate pads and employ isolated supply planes.

The physical distance from the transmit VCO and amplifier to the rest of the circuits has been maximized in an effort to minimize the leakage from the transmitter to the receiver and to reduce the frequency range over which the TX and RX VCOs are injection locking.

The antenna coupler, the on-chip power detector, and the phase generation and calibration circuits allow for the verification of the functionality and for some of the quantitative performance parameters of the packaged transceiver to be evaluated without mm-wave equipment and measurements. At the same time, due to the difficulty of accurately quantifying all performance parameters of the integrated transceiver with antenna through free-space measurements, and in order to evaluate the performance of the individual blocks, several breakouts of the transceiver circuits were fabricated and characterized separately on wafer.



(a) Tuning range versus fine control.

(b) Tuning range versus coarse control.

Figure 6.14: Measured VCO tuning range.

6.4.1. Breakouts

6.4.1.1. VCO

The VCO was characterized (i) as a standalone breakout, (ii) in a breakout that includes the VCO and the divider chain, (iii) in the receiver breakout and (iv) in the packaged transceiver at the divider output. Figure 6.14 reproduces the measured tuning range of the VCO at the IF output of the receiver, when the fine tuning control is swept from 0 to 1.2 V for different values of the coarse tuning control (fig. 6.14a), and when only the fine control is swept (fig. 6.14b). The tuning range spans 8.7 GHz from 115.2 GHz to 123.9 GHz.

Figure 6.15 illustrates the measured -100 dBc/Hz at 1 MHz phase noise of the VCO at 118 GHz. This is on par or better than that of other state of the art SiGe HBT VCOs in this frequency range [179, 207].

6.4.1.2. LNA

The S-parameters of the low noise amplifier were measured in a separate breakout using a D-band VNA. The S_{21} and S_{11} for the first few gain settings are depicted in figure 6.16. The dashed lines correspond to simulations in the highest gain state. The S_{11} variation with gain has been minimized by using the gain control technique described in section 6.2.1. Good agreement between simulation and measurements has been achieved due to careful modeling of all interconnects and passive components. Similar results, but with slightly higher S_{21}, were obtained for the transmit amplifier.

The utilized S-parameter measurement setup in the D-band (110-170 GHz) frequency range is illustrated in figure 6.17. It involves the use of a lower frequency VNA along with D-band OML frequency extenders, whose input/ouput port is in WR-6 rectangular

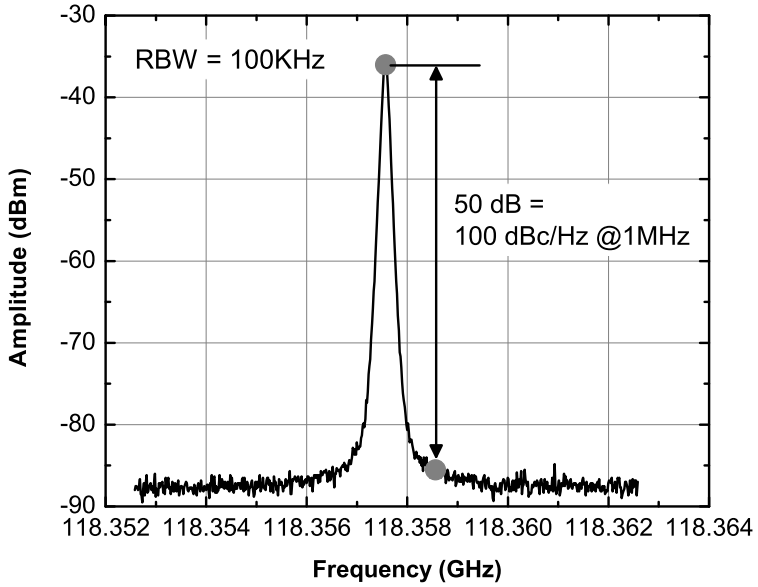


Figure 6.15: Measured phase noise of the VCO at 118 GHz.

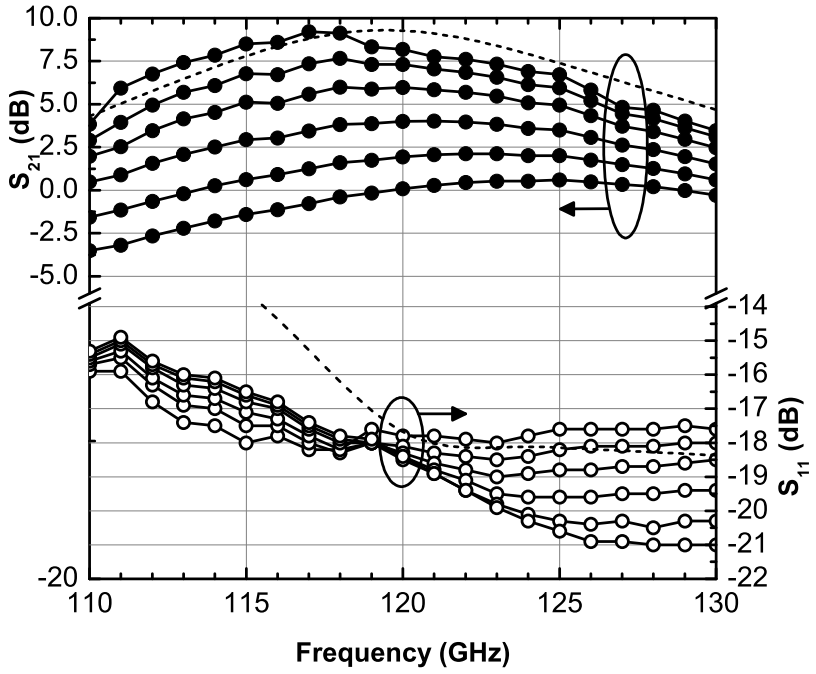
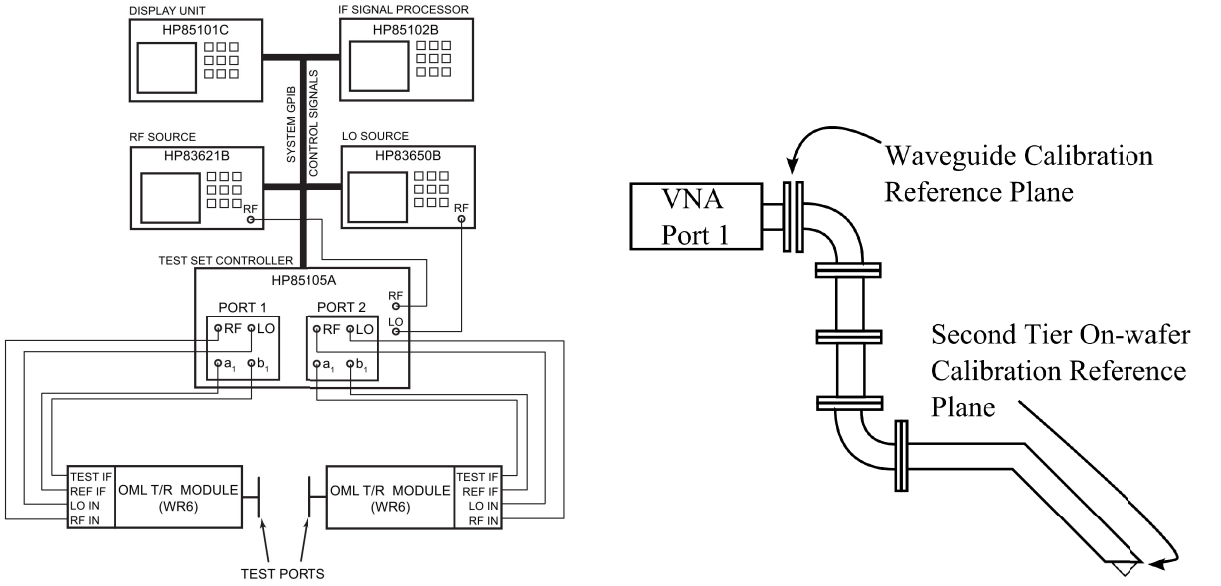


Figure 6.16: Measured (lines and symbols) S-parameters of the LNA breakout for 6 gain settings. The dashed lines correspond to simulations at maximum gain.



(a) D-band Vector Network Analyzer (VNA).

(b) D-band waveguide probe and calibration reference plane.

Figure 6.17: D-band S-parameter measurement setup and deembedding (after [2]).

waveguide. To interface with the circuit, a D-band waveguide circuit microprobe whose sketch is shown in figure 6.17b is employed. The non-idealities of the measurement setup are deembedded using a two-tier calibration process, involving a waveguide calibration step followed by an on-wafer calibration second step. The advantage of this procedure is that it can independently calculate the error boxes of the VNA and of the waveguide probes. The latter can the be used to calibrate the noise figure and power measurements presented in the following sections. A more detailed explanation of the measurement setup, and of the two-tier deembeding process can be found in [2].

6.4.1.3. Receiver

A separate receiver breakout was characterized with an ELVA-1 D-band noise source and the Y-factor method [41], as illustrated in figure 6.18. The measured single-ended conversion gain and DSB noise figure (from the RF input to the I IF output), at a 500-MHz IF frequency, are shown in figure 6.19. Since the 50-Ω IF buffers have no gain, the entire downconversion gain is achieved by the RF front-end alone. Although in this application the gain of the LNA had to be limited in order to preserve the overall linearity of the receiver, the 10-11.5 dB receiver noise figure is comparable to those of other D-band transceivers fabricated in technologies with similar performance [177–179].

The measured noise figure is 2 dB better than simulation, consistent with many other results obtained with this transistor model [1, 208] and is attributed to the inadequate cap-

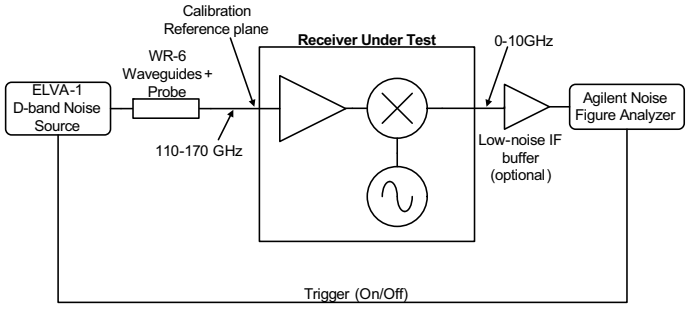


Figure 6.18: Noise figure measurement setup.

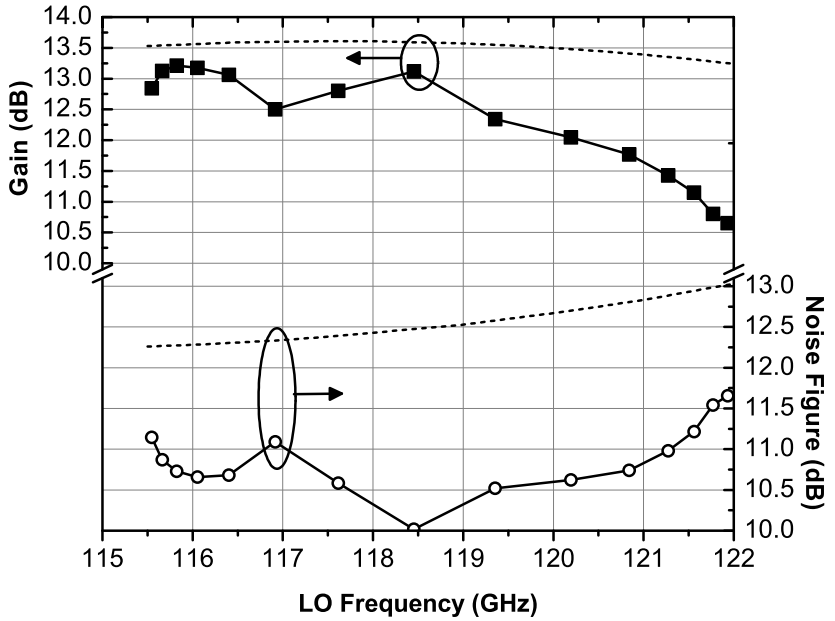


Figure 6.19: Measured (lines and symbols) and simulated (dashed lines) DSB noise figure and downconversion gain of the receiver breakout.

turing of the correlation between the collector and base noise currents [1, 2].

6.4.2. Transceiver

6.4.2.1. Receiver

Although the transceiver with the on-die antenna feed could not be wafer-probed, certain measurements could be carried out on the packaged chip, mounted on the PCB, by taking advantage of the leakage from the transmitter to the receiver and of the detector placed at the output of the transmitter.

The IF outputs of the receiver and divider chain were monitored with a spectrum analyzer and an oscilloscope, and the injection-locking properties of the two VCOs were analyzed. As the frequency difference between the transmit and receive VCOs, and thus the IF frequency,

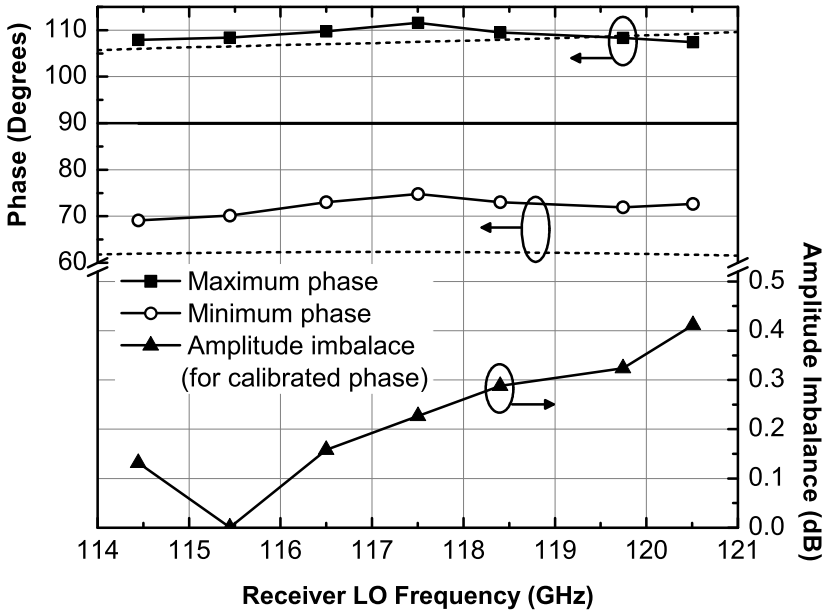


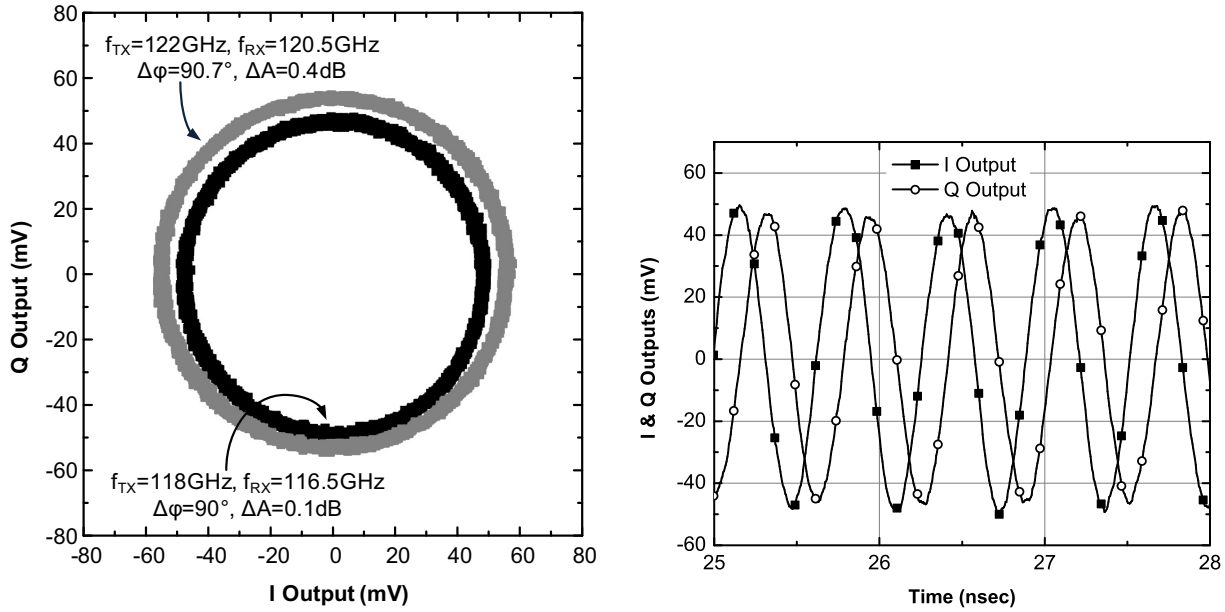
Figure 6.20: Measured (lines and symbols) and simulated (dashed lines) phase adjustment range of the I-Q receiver and amplitude imbalance when the phase is calibrated.

was decreased to less than approximately 1 GHz, the harmonics of the IF signal started to emerge and modulating tones appeared in the divided-down receive VCO spectrum. In order to avoid this injection-locked mode of operation, the rest of the transceiver measurements were performed with 1.5 GHz frequency offset between the transmitter and receiver VCOs. The upper limit of the IF frequency is 5 GHz due to receiver bandwidth and package limitations.

The phase difference and phase adjustment range of the I and Q IF outputs were measured by taking advantage of the leakage from the transmitter to the receiver. Figure 6.20 reproduces the measured phase adjustment range of the I and Q outputs of the receiver as a function of the LO frequency. The phase difference can be adjusted from 70° to 110°, which agrees reasonably well with the simulated values. The discrepancy between measurements and simulations in the lower range is attributed to the mismatch between the measured and simulated performance of the quadrature hybrid, as well as to other layout imbalances in the mixer that were intended to be calibrated out with this circuit.

Also shown in figure 6.20 is the measured amplitude imbalance between of the I and Q receiver outputs when the phase is calibrated to be exactly 90° for each receiver LO frequency. The amplitude imbalance remains less than 0.4 dB (better than simulation due to the saturated operation of the mixer in the LO port) for the entire range, and could be easily compensated by including variable gain in the baseband amplifiers.

Figure 6.21a reproduces an X-Y plot of the calibrated I and Q receiver outputs for two



(a) XY plot of the receiver IQ outputs at 2 frequencies.

(b) Time domain signals when $f_{TX} = 122\,\text{GHz}$.

Figure 6.21: Quadrature receiver outputs.

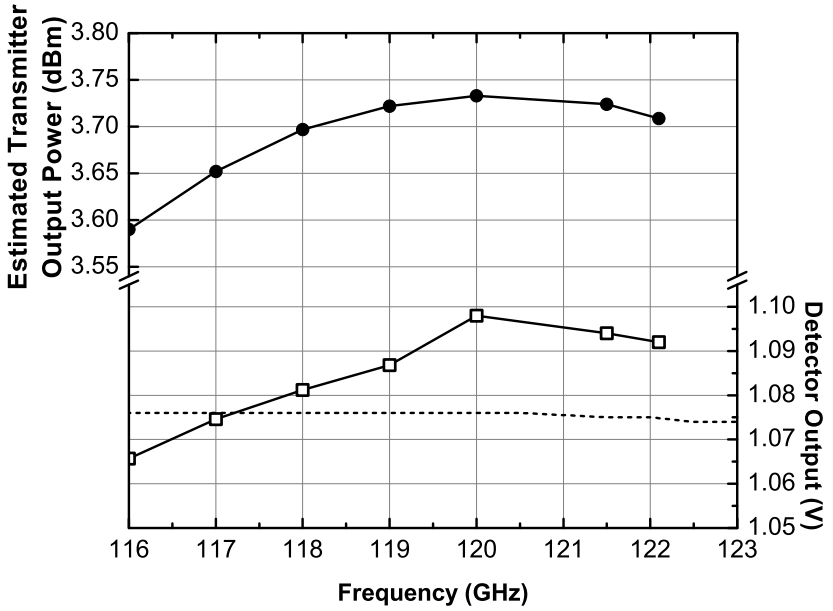
LO frequencies while the corresponding time domain signals are shown in figure 6.21b.

6.4.2.2. Transmitter

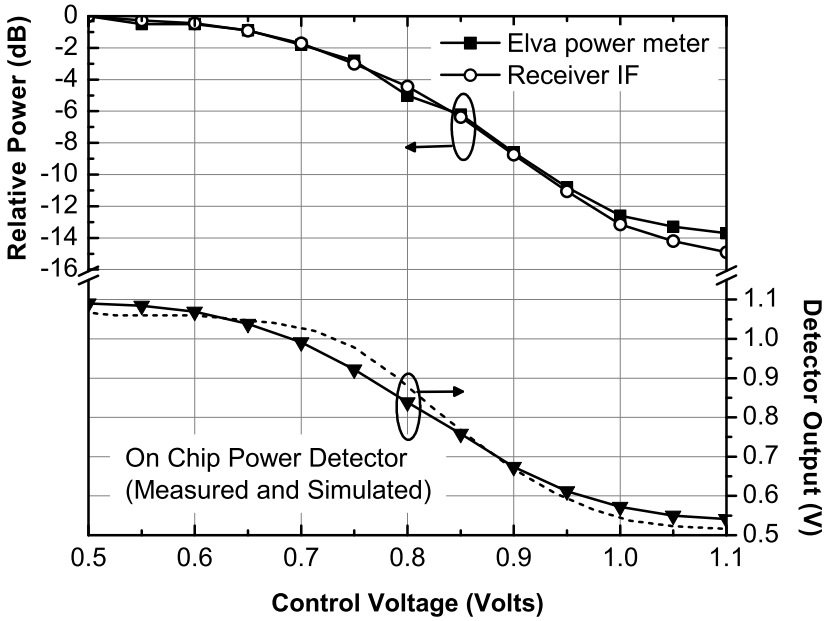
In order to characterize the output power of the transmitter, the DC voltage of the on-chip power detector was measured over frequency, as illustrated in figure 6.22a. Based on the simulated performance of that detector (fig. 6.9b), the signal power at the transmit amplifier output was estimated to be approximately 3.6-3.7 dBm across the entire TX VCO frequency range. Furthermore, the less than 20 mV mismatch between the measured and simulated detector output voltages indicate that the error in the estimated output power is expected to be less than 1 dB.

The output power control function was verified, as illustrated in figure 6.22b, by varying the transmit amplifier output power control and monitoring the power of the receiver IF output and the DC output voltage of the on-chip power detector. Furthermore, an ELVA-1 D-band power sensor with a horn antenna was brought close to the chip and its power reading was monitored [1]. As seen in figure 6.22b, the three power measurements track each other, indicating that the output power can be controlled over a range of at least 15 dB. Furthermore, this experiment also validates the linear operation of the receiver since the amplitude of the IF output signal due to transmitter leakage responds linearly to changes in

[1]This measurement was performed in the near field as the power sensor does not have enough sensitivity to measure the power in the far field. Therefore, deembeding the free space loss was not attempted.



(a) Measured (lines and symbols) and simulated (dashed line) on-chip power detector output voltage versus frequency and transmitter output power.



(b) Transmitter output power versus control voltage measured with a power meter, at the IF receiver outputs and at output of the on-chip power detector at 122 GHz.

Figure 6.22: Output power measurements.

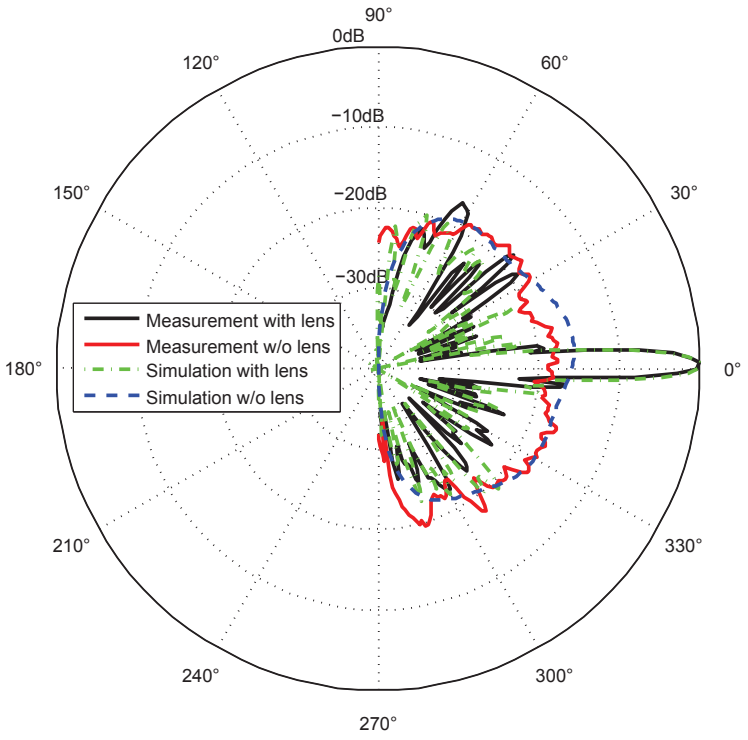


Figure 6.23: Measured and simulated H-plane normalized antenna radiation patterns with and without the focusing lens (courtesy of Jurgen Hasch).

the transmitter output power.

6.4.2.3. Antenna

The relative radiation pattern of the antenna was measured at the output of the receiver by placing the PCB with the packaged chip on a rotating table and illuminating it using a 122-GHz signal source. As illustrated in figure 6.23, two cases were compared: (i) the board with the EM-coupled antenna alone, and (ii) the board with EM-coupled antenna and a dielectric meniscus (convex-concave) lens with d=35 mm diameter and 20 mm focal length which yields a maximum theoretical directivity of

$$D = \frac{\pi^2 d^2}{\lambda^2} = 33\,\text{dBi} \tag{6.7}$$

Based on the measured patterns of figure 6.23, the antenna gain is improved by 20 dB by the lens. As a result, the estimated total antenna gain is 26 dBi.

6.4.2.4. Measurement of Rotation Speed

To verify the intended operation of the transceiver through the air, the speed of a rotating reflector placed at various distances was measured using the evaluation board with the

packaged chip and the dielectric lens.

In this simple test case, a portion of the transmitted signal at frequency f_{TX} leaks into the receiver input and another portion is transmitted and reflected back by the reflector. The signal at the I IF output of the receiver can be expressed as:

$$\mathrm{IF}(t) = A_{\mathrm{leak}} \cos\left(2\pi f_{\mathrm{IF}} t + \phi_{\mathrm{leak}}\right) + a_R(t) A_{\mathrm{refl}} \cos\left(2\pi f_{\mathrm{IF}} t + \phi_{\mathrm{refl}}\right) \tag{6.8}$$

where A_{leak} and A_{refl} are the amplitudes of the signal leaking from the transmitter to the receiver and the reflected signals respectively, ϕ_{leak} and ϕ_{refl} are the corresponding phases while $f_{\mathrm{IF}} = f_{\mathrm{TX}} - f_{\mathrm{RX}}$ is the IF frequency. $a_R(t)$ is a function that depends on the reflector type, the antenna radiation pattern, and the rotation speed. In its simplest form, $a_R(t)$ is unity when the reflector is in parallel with the evaluation board and zero elsewhere. The amplitude of the reflected signal can be calculated by using the radar equation:

$$A_{\mathrm{refl}} = \sqrt{2 Z_0 P_{\mathrm{RX}}} = \frac{(2 Z_0 G_{\mathrm{RX}} P_{\mathrm{TX}} a_{\mathrm{coupler}} \sigma)^{1/2} G_{\mathrm{ant}} c}{(4\pi)^{3/2} f_0 R^2} \tag{6.9}$$

where $Z_0 = 50\,\Omega$ is the termination impedance at the IF output.

Similarly, the I reference output is

$$\mathrm{REF}(t) = A_{\mathrm{ref}} \cos\left(2\pi f_{\mathrm{IF}} t + \phi_{\mathrm{ref}}\right) \tag{6.10}$$

By multiplying the reference and IF output signals and low-pass filtering, the baseband signal becomes:

$$\mathrm{BB}(t) = A_{\mathrm{leak}} A_{\mathrm{ref}} \cos\left(\phi_{\mathrm{ref}} - \phi_{\mathrm{leak}}\right) + a_R(t) A_{\mathrm{refl}} A_{\mathrm{ref}} \cos\left(\phi_{\mathrm{ref}} - \phi_{\mathrm{refl}}\right) \tag{6.11}$$

Equation (6.11) indicates that the baseband signal includes a DC component that depends on the leakage and a time varying part that depends on the rotation speed of the reflector. Based on the simple model for $a_R(t)$, the strongest harmonic in the spectrum of the baseband signal $\mathrm{BB}(t)$ will correspond to the rotation speed.

In practice, all four outputs of the chip could be further downconverted to a lower IF frequency of a few MHz, digitized and processed by a digital signal processor [209]. However, for a simple demonstration, the in-phase receiver IF output is amplified and multiplied with the in-phase reference output using an external passive mixer, as illustrated in figure 6.24. Figure 6.25 reproduces the $\mathrm{BB}(t)$ output in time and frequency domains for a $13 \times 13\,\mathrm{cm}^2$ reflector, located 30 cm away from the PCB (including the dielectric lens). Figure 6.26 shows $\mathrm{BB}(t)$ when the reflector is positioned 2.1 m away from the PCB. The lower frequency tones

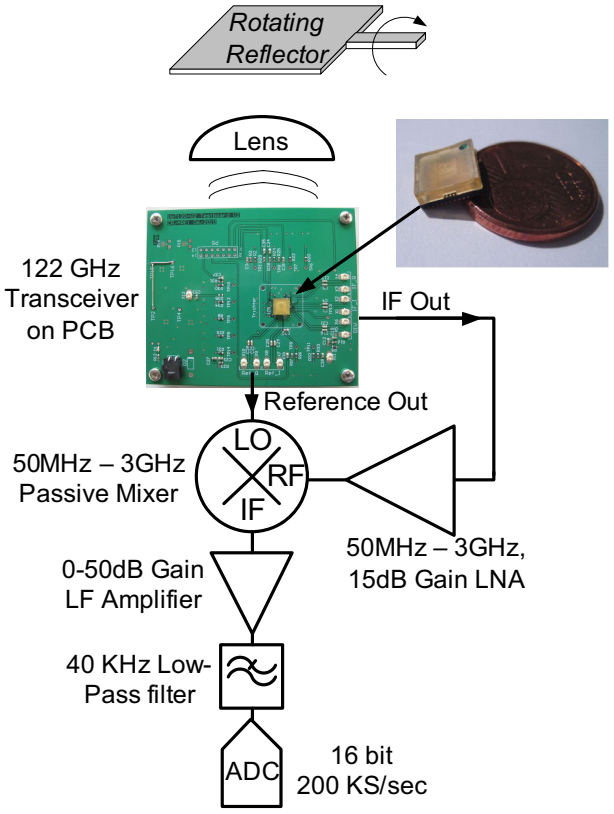
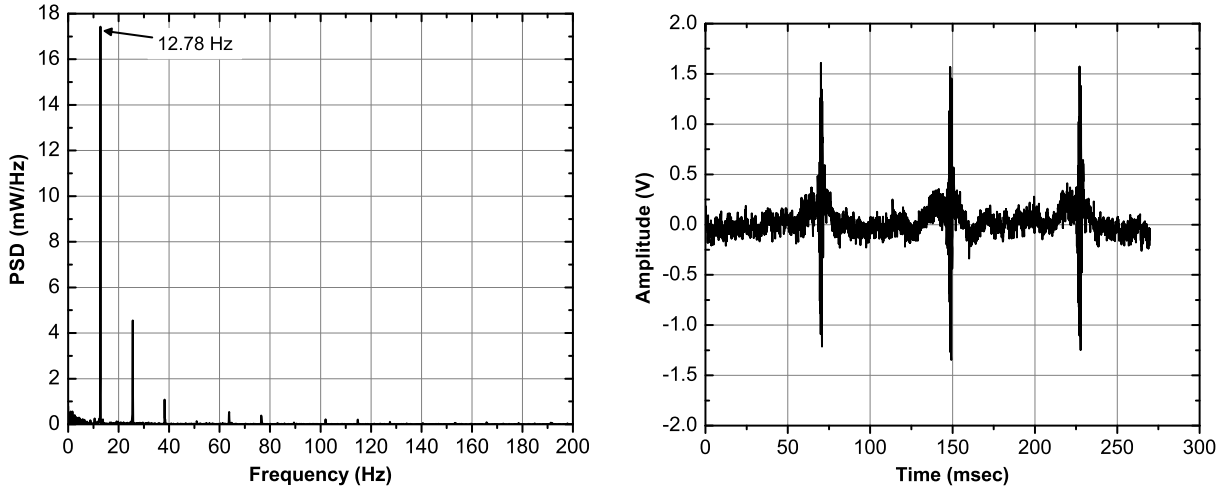
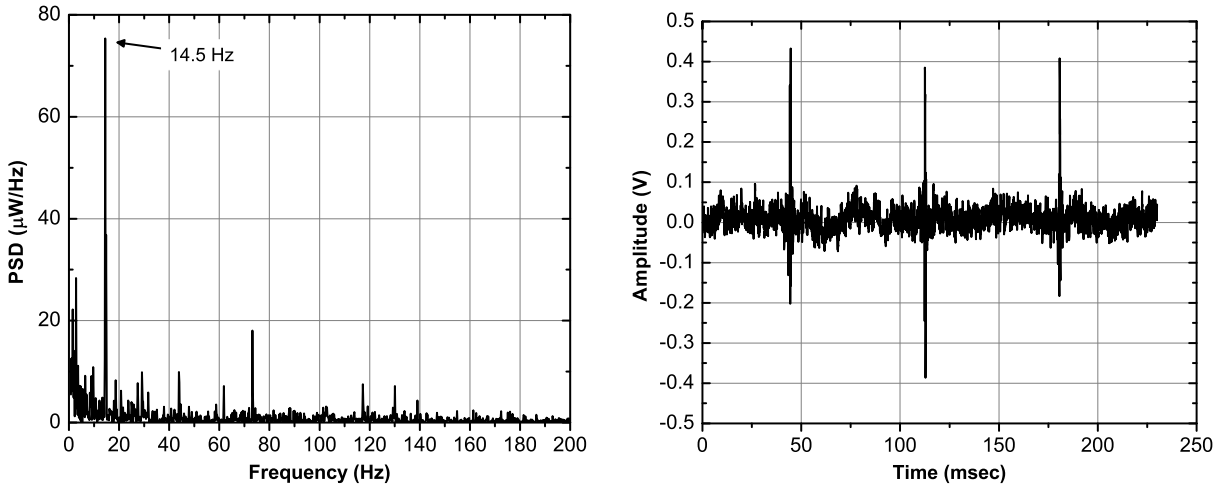


Figure 6.24: Rotation speed test setup.



(a) Spectrum. (b) Time domain signal.

Figure 6.25: Measured speed of a rotating reflector placed at 30 cm above the board.



(a) Spectrum. (b) Time domain signal.

Figure 6.26: Measured speed of a rotating reflector placed at 2.1 mabove the board.

in figure 6.26a correspond to movement spurs caused by the person who held the rotor in place. The difference in rotation speed between the two positions is also caused by this person's movement. In both experiments the transmitter frequency was 122 GHz while the receiver frequency was 120 GHz.

6.5. 145 GHz Sensor Architecture and Circuit Design

Circuits that operate in the mm-wave frequency range typically have little design margin and are sensitive to process variation and model inaccuracies, thus exacerbating the need to test every part before delivering it to the customer. This requirement is in sharp contrast with the fact that mm-wave test equipment, especially in the W and D-bands, is cumbersome, has limited capabilities and very high cost. Therefore, implementing self-test features on-chip will go a long way to solving these problems and to reduce (or eliminate) the use of mm-wave test equipment [210].

In previous work, self-test was demonstrated in 77 GHz phased arrays [211] and mixers [212] but not in integrated transceivers. In the 140-GHz band in particular, which has emerged as a potential candidate for next generation radar systems [213,214], neither self-test nor calibration functionality has been implemented.

In this section, the first fundamental frequency, fully integrated radar transceiver operating at 143-152 GHz is presented. The transceiver implements the proposed heterodyne architecture with built-in calibration. In addition, it features several built-in self-test (BIST) capabilities that allow for the testing of most of its functionality without using external mm-wave equipment.

6.5.1. Architecture

The block diagram of the proposed transceiver is shown in figure 6.27. The system architecture is similar to that of the 122 GHz IQ heterodyne sensor along with some improvements based on the experience gained from it. First, the receivers in the 145 GHz system are double sideband and not IQ. Measurements in the 122 GHz system indicated that low-IF frequencies, lower than 1 GHz, can be achieved without injection locking the TX and RX VCOs if adequate buffering is provided to the VCOs. Therefore, the IQ receivers were dropped in order to reduce the system complexity and associated power consumption, but also to reduce the number of circuits in the LO path and thus increase the LO signal power at the mixers. This will decrease the mixer contribution to the system noise and increase their gain. Second, the main and reference channels were made perfectly symmetrical, in contrast to the system of figure 6.1. This symmetry is necessary for the proper calculation of the error coefficients by the calibration method.

Several BIST features have been included in order to facilitate simple low-frequency, low-cost testing and on-line monitoring of the circuit performance. First, divide-by-64 chains have been introduced for both the TX and RX VCOs. Apart from allowing the VCOs to be locked by external PLLs, the dividers provide a low frequency signal that can be used to independently verify the tuning range and phase noise of the VCOs.

Bidirectional power detectors, capable of measuring the forward and reflected waves, have been inserted between the RX LO distribution tree and the mixers, as well as between the TX LO distribution and the transmit amplifiers. Their role is to monitor the signal power provided by the LO distribution trees and to isolate potential problems either in the mixers, transmit amplifiers, or in the VCOs. Similar detectors were also placed at the outputs of the primary and of the reference channel. Apart from measuring the transmit power, the detector reading of the reflected power can be employed to calculate the reflection coefficient at the TX output and thus determine whether the antenna is well matched.

A digital impedance tuner has been added at the output of the reference channel in order to implement the calibration procedure proposed in section 5.6.3. In addition to calibration, the same process also performs a full functional test on the capability of the reference channel to measure the amplitude and phase shifts generated by the tuner. Any unwanted degradation of the receiver gain, additional delays or increased TX to RX leakage, that would indicate a malfunction, will be reflected in the error coefficients and can be easily identified.

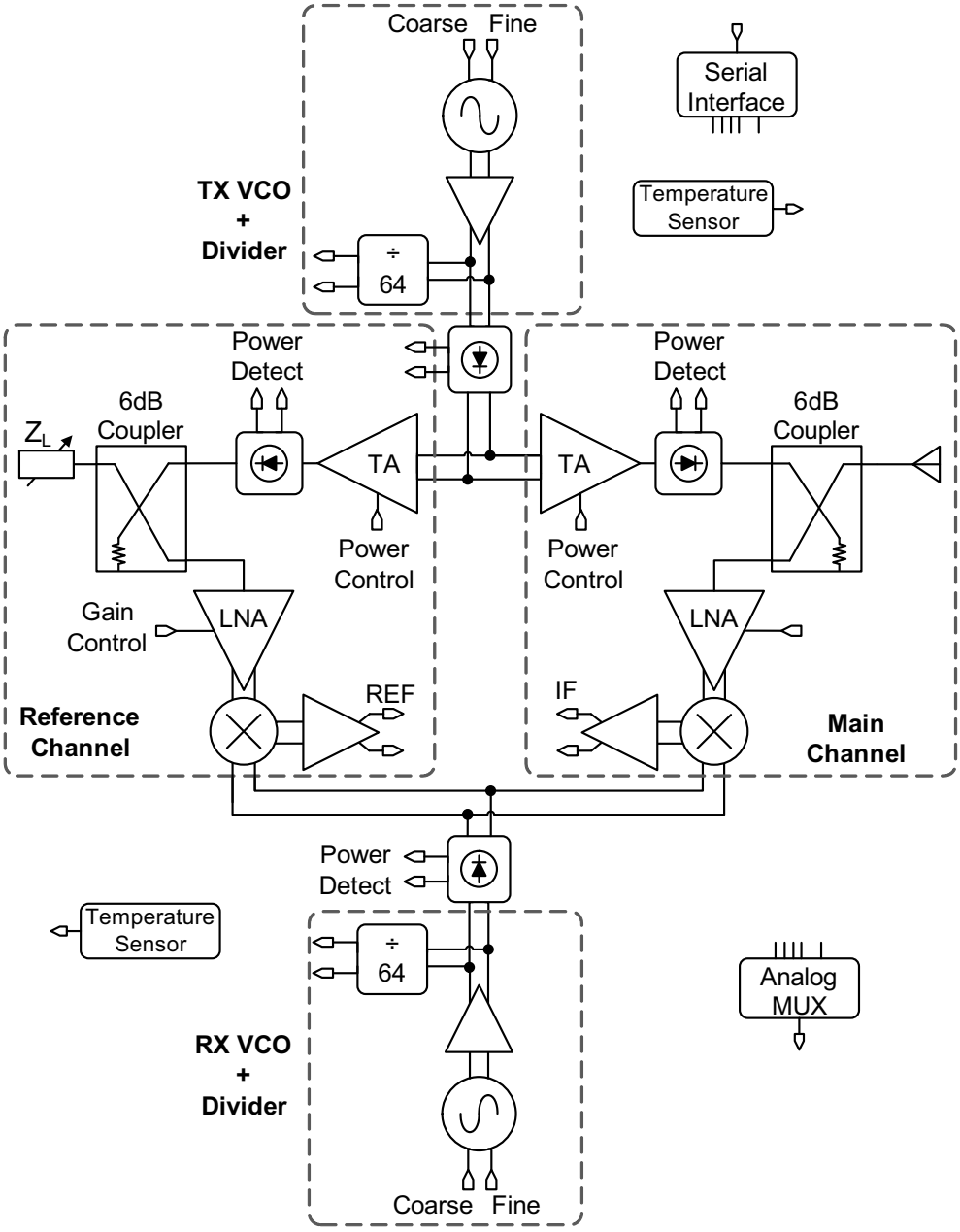


Figure 6.27: Block diagram of the radar transceiver.

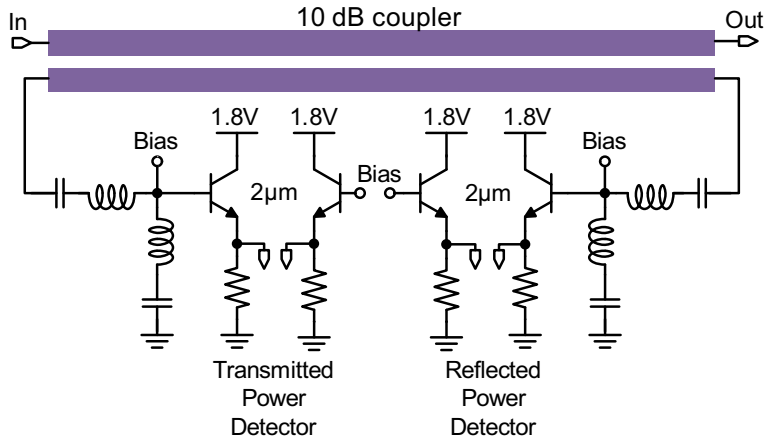


Figure 6.28: Schematic of the bidirectional power detector.

6.5.2. Circuit Design

The schematic of the bidirectional power detector is illustrated in figure 6.28. It is based on a 10-dB coupled-line directional coupler whose coupled outputs are terminated on bipolar detectors. The inputs of the bipolar detectors need to be matched to 50 Ω, i.e. to the coupler characteristic impedance. A coupling ratio of 10 dB was selected in order to minimize the impact of the detector on the signal path. Since the portion of the signal that will be routed to the detectors is decided by the coupling ratio, the transistor size in the bipolar detectors can be freely selected, in contrast to the detector employed in section 6.2.5, where the smallest transistor size had to be used. This allows for optimizing the transistor size in order to easily match the bipolar detectors to 50 Ω and to maximize their responsivity.

The schematic of the proposed, digitally controlled impedance tuner is shown in figure 6.29. The nFET transistors act as two-state (high/low) resistors, and their parasitic capacitance is tuned out by the shunt inductors. The 45° transmission line rotates Γ_C by a total of 90°, rendering it purely imaginary, as opposed to Γ_T which ideally assumes only real values. As a result, the reflection coefficient at the input of the coupler becomes:

$$\Gamma_{in} = \Gamma_T + j\Gamma_C \tag{6.12}$$

which ideally maps to a square around $0 + j0$ when Γ_T and Γ_C vary from 0 to 1. In this design, four bits were selected to control Γ_T and four bits to control Γ_C.

The LNAs, VCOs, TAs and downconvert mixers employ the same 1.2/1.8 V HBT-only topologies as those in the 122 GHz sensor. Digital gain control has been implemented in all amplifiers and the corresponding control bits, along with those needed to program the impedance tuner, are provided by a serial interface. The digital gain control has been

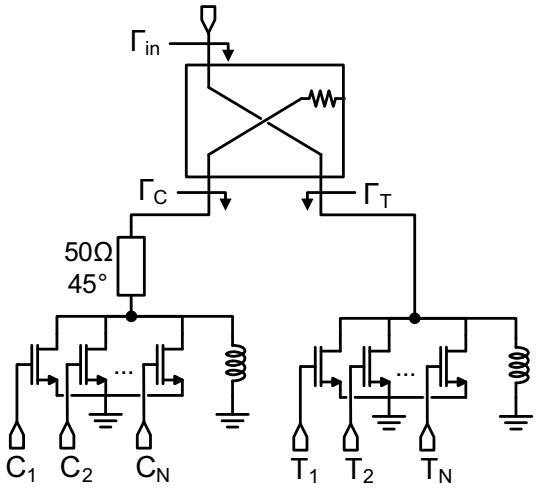
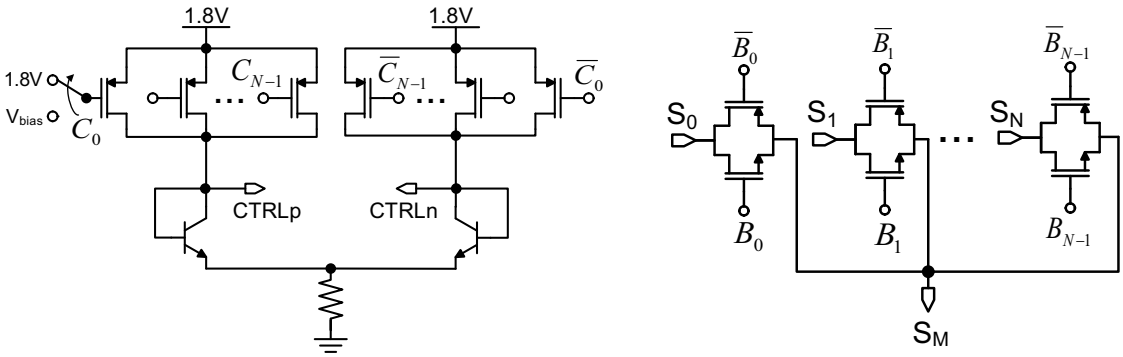


Figure 6.29: Schematic of the variable impedance tuner.



(a) Digital gain control.

(b) Analog multiplexer.

Figure 6.30: Digital control circuit schematics.

achieved by employing the circuit of figure 6.30a, where the current through the diode connected HBT current mirror is controlled through pFET switched current sources. The gate of each pFET finger can be switched between V_{DD}, when transistor is off, and V_{bias} which generates 100 μA current flow. Selecting the appropriate number of fingers that are set to V_{bias} will thus adjust the current through the HBT mirrors and the resulting gain of the circuit at which their current is mirrored (e.g. to transistors Q_3 and Q_4 of figure 6.2).

The outputs of the power detectors and temperature sensors can be scanned through the on-chip analog multiplexer (MUX) of figure 6.30b, which is also controlled by the serial interface. The MUX is composed of CMOS transmission gates whose outputs are shunted together and their inputs are connected to the analog signals to be read. By enabling only one transmission gate and disabling all the others, the corresponding analog signal is read from the common output.

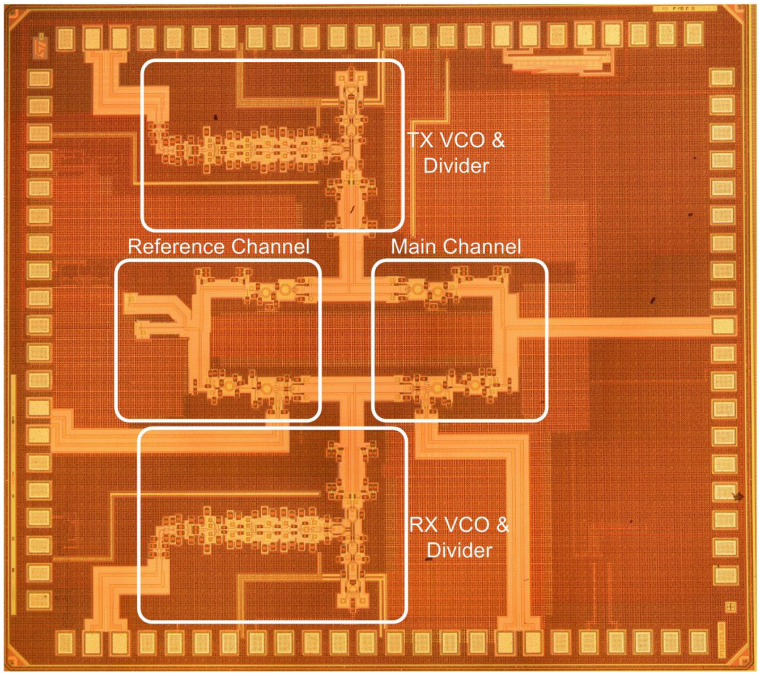


Figure 6.31: Die microphotograph.

6.6. Measurement Results for the 145 GHz Sensor

A die photograph is shown in figure 6.31. It occupies an area of 2.6 mm × 2.3 mm. The power consumption is 800 mW from 1.2 and 1.8 V power supplies. The measured tuning characteristics of the VCO are plotted in figure 6.32. The frequency range spans from approximately 143 to 152 GHz, with the range covered by the VCO fine control being approximately 800 MHz.

Figure 6.33 depicts the receiver downconversion gain and double-side band noise figure, as measured on a stand-alone receiver breakout. The NF remains below 10.5 dB across the whole band. Interestingly, the NF is slightly lower that the 10-11.5 dB of the 122 GHz receiver (fig. 6.19), although the frequency of operation is 20-30 GHz higher. This is attributed to two facts. First the IQ calibration block was removed and the mixers are double sideband, requiring less splitting of the LO signal and thus receiving higher LO power. Second, a differential layout asymmetry in the VCO layout has been corrected in the 145 GHz version. This fix allowed the VCO to provide higher output power, which also contributed to the increase of the LO power at the mixers.

Figure 6.34a reproduces the measured S-parameters of the bidirectional power detector breakout. The detector is well matched and its loss remains below 0.8 dB across the whole D-band. This demonstrates that the power detector introduces a very small loss burden

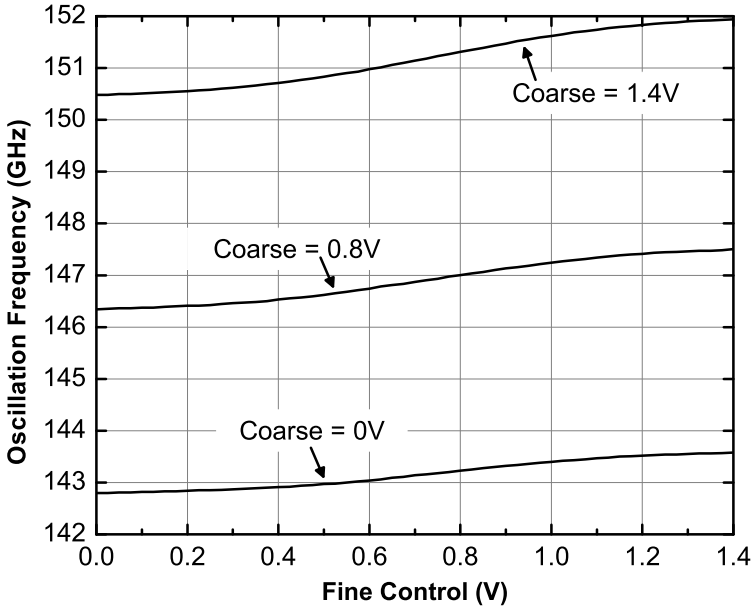


Figure 6.32: Measured VCO tuning curve.

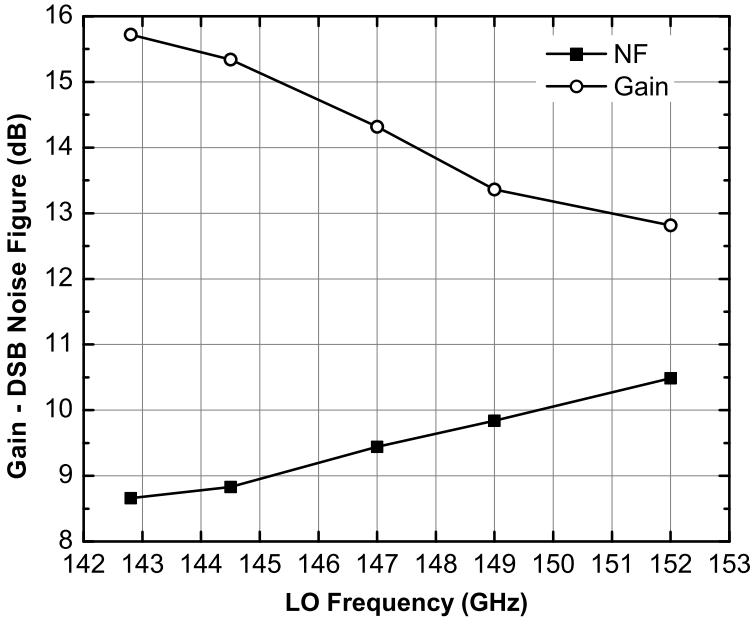
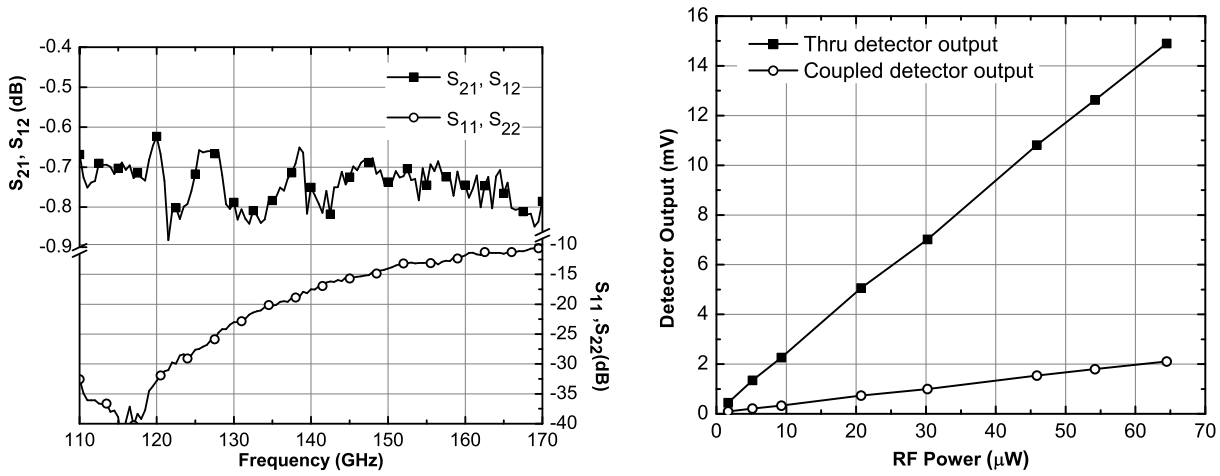


Figure 6.33: Measured receiver gain and DSB noise figure.



(a) S-parameters.

(b) $P_{in} - V_{DC}$ at 146 GHz.

Figure 6.34: Bidirectional power detector measured S-parameters and $P_{in} - V_{DC}$ characteristic.

when employed to monitor internal system nodes. Figure 6.34b depicts the measured V_{DC} versus P_{in} characteristic of the detector breakout. Because of the 10-dB coupler, only a fraction of the power reaches the bipolar detectors, which typically exhibit a linear $P_{in} - V_{DC}$ characteristic at low power levels. According to simulation, the linear behavior extends to power levels as high as +1 dBm.

Figure 6.35 shows the measured Γ_{in} states of the impedance tuner breakout at 146 GHz. Due to the non-idealities of the hybrid coupler used in the tuner and imperfect tuning of the output capacitance of the nFET resistors, the locus is not centered exactly around $0 + j0$. However, the impedance states are adequately separated and the tuner can provide a wide range of phase delays to test and calibrate the transceiver.

For testing purposes, the chip was wire-bonded in open-lid QFN package, similar to the one of figure 6.11a, which was then soldered on a simple FR-4 PCB. The board was mounted on a probe station in order to probe the antenna port of the primary channel of the transceiver.

Figure 6.36 illustrates the output power of the transceiver chip measured with the on-chip power detectors as well as by probing the antenna port and using a commercial D-band ELVA power sensor. The measured power at the antenna port is consistent with the on-chip power sensor readings since 7-8 dB loss between the transmit amplifier output and the antenna pad are expected due the 6-dB coupler and the associated interconnect.

Finally, the on-chip tuner was used to calibrate the reference channel using equation (5.61). Figure 6.37 illustrates the reflection coefficient of several tuner states, without any correction, compared with the corresponding measured states of the standalone tuner. After

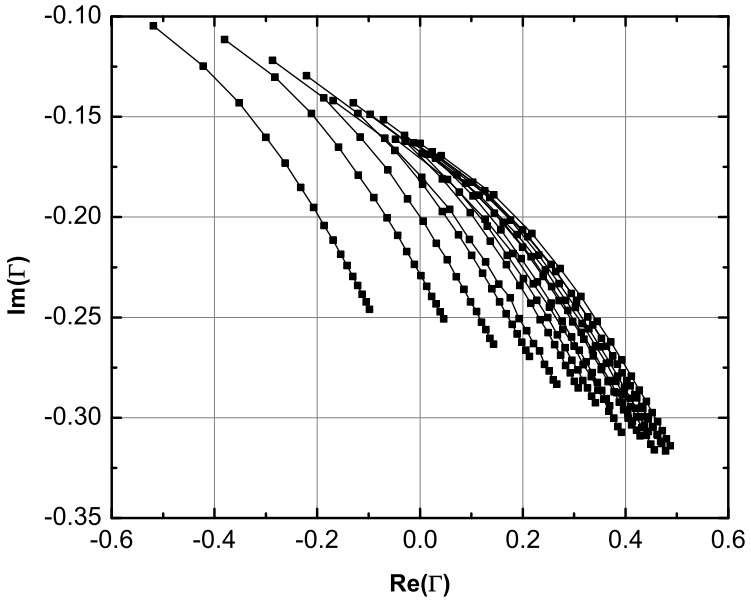


Figure 6.35: Measured Γ_{in} states of the impedance tuner breakout at 146 GHz.

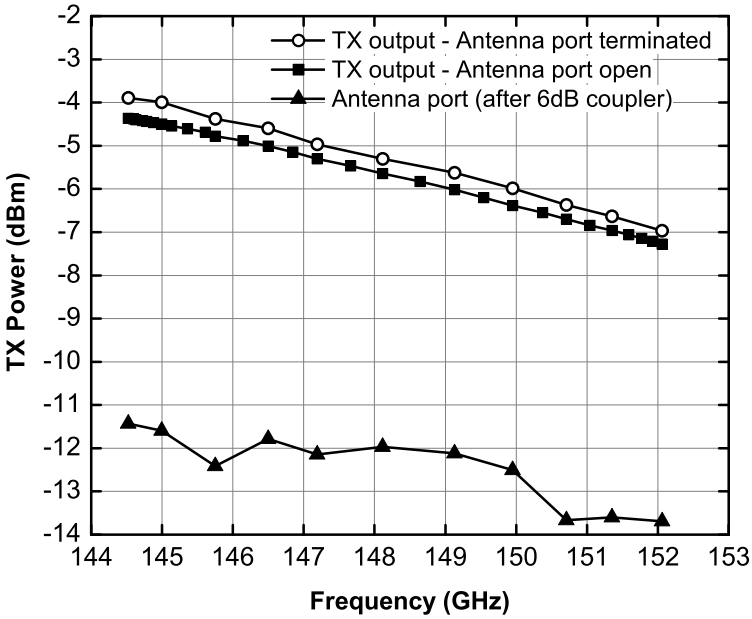


Figure 6.36: Measured output power of the transceiver measured at the transmit amplifier output using the on-chip power detector, and at the antenna port with an external power sensor.

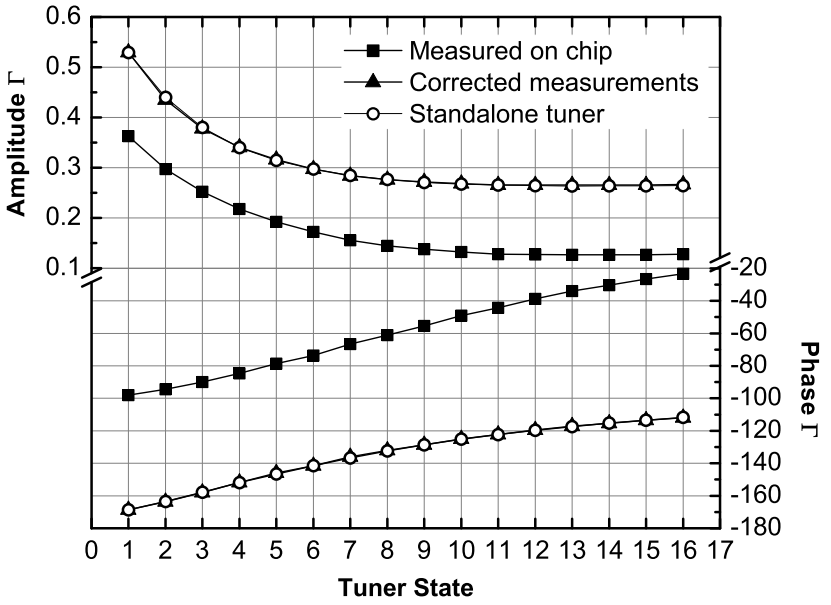


Figure 6.37: Measured reflection coefficients at 146 GHz using the transceiver chip and the tuner before and after calibration.

applying a calibration with three impedance states, the RMS error in the corrected reflection coefficients was less than 1%.

6.7. Summary

Two single-chip, low-power, SiGe BiCMOS transceivers operating at 122 and 145 GHz have been presented. By careful selection of circuit topologies, the power supply voltage was kept below 1.8 V and the power consumption was reduced to less than 900 mW in both cases. The systems employ new architectures that allow for significant improvement in their SNR and systematic errors, while simultaneously focusing on demonstrating low cost systems above 100 GHz.

Specifically, by employing an on-chip antenna feed along with an external patch resonator, no off-chip transitions were necessary at the 122 GHz chip. As a result, the chip was wirebonded in a low-cost, open-lid QFN package, illustrating the feasibility of low cost, highly integrated D-band systems. The 145 GHz transceiver attempted to further expand on the idea of reducing the cost by incorporating power detectors and an impedance tuner that allow for built-in self-test and for the implementation of a novel calibration approach.

Table 6.2 compares the performance of the transceivers presented in this chapter with recently published work above 100 GHz.

Table 6.2: Transceiver performance comparison

	f_0 (GHz)	G_{RX} (dB)	NF (dB)	P1dB (dBm)	P_{out} (dBm)	PN (1MHz) (dBc/Hz)	P_{DC} (W)	f_T/f_{max} (GHz)	Notes
[179]	122	21	11	-44	-	-94	0.37	245/350 SiGe	Receiver 2×60 GHz
[178]	140	30	12	-	-8	-80	1.5	230/280 SiGe	Transceiver 2×70 GHz
[177]	160	-	-	-	1	-89	1.7	260/380 SiGe	Transmitter 9×17.7 GHz
[177]	160	25	14	-	-	-89	1.5	260/380 SiGe	Receiver 9×17.7 GHz
[207]	122	-	-	-	2	-95	0.35	210/240 SiGe	Transmitter 1×122 GHz
[215]	118	-	-	-	-8	-	0.2	65nm CMOS	Transmitter ext. LO
[216]	220	3.5	7.4	-	-	-	0.11	200/300 GaAs	Receiver 4×55 GHz
[217]	123	3	5	-	-	-	0.12	200/300 GaAs	Receiver 4×30 GHz
[214]	160	27	12	-	0	-82	2	230/280 SiGe	Transceiver 2×80 GHz
This Work	122	13	11.5	-20	3.6*	-100	0.9	230/280 SiGe	Transceiver 1×122 GHz
This Work	145	14	10	-20	-12**	-82	0.8	230/280 SiGe	Transceiver 1×145 GHz

* At the transmitter output, not including the antenna coupler and antenna.
** At the antenna port.

7 Conclusions

7.1. Summary and Contributions

This thesis presented new ideas at the circuit and system level for a variety of high frequency applications. After revisiting the theory behind high frequency small and large signal amplifier design in chapters 2 and in the beginning of chapter 3 respectively, a new circuit topology for power amplification in nanoscale CMOS was proposed in chapter 3. The purpose of the proposed power amplifier approach was to boost the achievable efficiency by making use of class-D, switched mode operation and by minimizing the use of passive components.

Proceeding into system level considerations, chapter 4 presented a transceiver for E-band communications that features an alternative transmitter architecture to facilitate direct QPSK modulation, while avoiding the need for separate upconversion and linear power amplification. Lastly, Chapter 5 reviewed the critical system level considerations behind radar systems and proposed two new architectures for implementing a precision distance sensor, the details of which were presented in chapter 6.

7.2. Future Work

The distance sensors presented in chapters 5 and 6 were developed as part of a collaboration between the university of Toronto, Robert Bosch GmbH and several other research partners in the European Union SUCCESS project (`http://www.success-project.eu/`). The developed chips are being packaged by the project partners and a system demonstrator was built. Preliminary results have already been reported in [209], demonstrating the feasibility of the approach.

The SNR formulas and predictions of chapter 6 were only used as a guidance for system design and have not been experimentally verified in this thesis. Although these formulas are

based on the fundamentals of radio systems and noise, their experimental verification would be important in order to investigate if there are other unmodeled sources of noise that would limit the system performance.

The distance sensors could be redesigned in a more advanced SiGe process, such as [27], which would provide room for considerable reduction in the power consumption. Along with a silicon antenna lens with a very small form-factor, such a chip would have the potential to be used in smart-phones or other portable devices, where there is a quest for increasing integration of different sensors in a single device.

The power amplifier of chapter 3 can provide a very good solution for integrating power amplifiers in CMOS transceivers. However two issues need to be resolved. First, a boost converter needs to be implemented in the same technology in order to generate the high supply voltage required by the stacked PA. Such a converter using a stacked architecture has been demonstrated in [218]. Second, the class-D amplifier needs to be linearized to process variable envelope signals. The most promising technique for this is through the use a bandpass $\Delta\Sigma$ modulator [106, 219, 220].

The transistor sizing methodology of chapter 3 is based on optimizing the drain efficiency of only the output power amplifier stage and did not account the driver stages. The main difficulty in doing so is the fact that all the parasitic capacitors were assumed to always fully charge and discharge at every conduction cycle. This is not the case for the drivers whose transistors needs to be sized precisely in order to be able to drive the input capacitance of the following stage. Therefore, a different mathematical procedure is required for optimizing the size the transistors of the driver stages that will minimize their power consumption while being able to drive the final PA stage. With such an approach, it will be possible to optimize the PAE of the multistage PA.

Another obvious application of the class-D CMOS PA is as a broadband, high voltage driver for optical modulators. Ideally, a level shifter that will provide the appropriate signal at the pFET top transistor without the need for a capacitor will be included. This will alleviate the low frequency roll-off and render the amplifier capable of processing signals with low frequency content.

7.3. List of Publications

The work described in this thesis has been presented in the following publications:

- I. Sarkas, J. Hasch, A. Balteanu, and S. Voinigescu, "A Fundamental Frequency 120-GHz SiGe BiCMOS Distance Sensor With Integrated Antenna," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 795–812, Mar. 2012.

- I. Sarkas, S. Nicolson, A. Tomkins, E. Laskin, P. Chevalier, B. Sautreuil, and S. Voinigescu, "An 18-Gb/s, Direct QPSK Modulation SiGe BiCMOS Transceiver for Last Mile Links in the 70-80 GHz Band," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 10, pp. 1968–1980, Oct. 2010.

- I. Sarkas, A. Balteanu, E. Dacquay, A. Tomkins, and S. Voinigescu, "A 45 nm SOI CMOS Class-D mm-Wave PA with >10 Vpp differential swing," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest*, Feb. 2012, pp. 88–90.

- I. Sarkas, M. G. Girma, J. Hasch, T. Zwick, and S. P. Voinigescu, "A Fundamental Frequency 143-152 GHz Radar Transceiver with Built-in Calibration and Self-Test," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2012.

- I. Sarkas and S. P. Voinigescu, "A 1.8 VSiGe BiCMOS Cable Equalizer with 40-dB Peaking Control up to 60 GHz," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2012.

- E. Laskin, A. Tomkins, A. Balteanu, I. Sarkas, and S. Voinigescu, "A 60-GHz RF IQ DAC Transceiver with On-Die at-speed Loopback," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2011.

- I. Sarkas, E. Laskin, J. Hasch, P. Chevalier, and S. Voinigescu, "Second Generation Transceivers for D-band Radar and Data Communication Applications," in *IEEE International Microwave Symposium (IMS) Digest*, May 2010, pp. 1328–1331.

- I. Sarkas, S. Nicolson, A. Tomkins, E. Laskin, P. Chevalier, B. Sautreuil, and S. Voinigescu, "A 12-Gb/s, Direct QPSK Modulation SiGe BiCMOS Transceiver for Last Mile Links in the 70-80 GHz Band," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2009.

Bibliography

[1] K. H. K. Yau, P. Chevalier, A. Chantre, and S. P. Voinigescu, "Characterization of the Noise Parameters of SiGe HBTs in the 70–170-GHz Range," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 8, pp. 1983–2000, Aug. 2011.

[2] K. Yau, "On the Metrology of Nanoscale Silicon Transistors Above 100 GHz," Ph.D. dissertation, University of Toronto, 2011.

[3] W. Deal, X. Mei, K. Leong, V. Radisic, S. Sarkozy, and R. Lai, "THz Monolithic Integrated Circuits Using InP High Electron Mobility Transistors," *IEEE Transactions on Terahertz Science and Technology*, vol. 1, no. 1, pp. 25–32, Sep. 2011.

[4] A. Tessmann, A. Leuther, V. Hurm, I. Kallfass, H. Massler, M. Kuri, M. Riessle, M. Zink, R. Loesch, M. Seelmann-Eggebert, M. Schlechtweg, and O. Ambacher, "Metamorphic HEMT MMICs and Modules Operating Between 300 and 500 GHz," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 10, pp. 2193–2202, Oct. 2011.

[5] M. Seo, M. Urteaga, A. Young, J. Hacker, A. Skalare, R. Lin, and M. Rodwell, "A Single-Chip 630 GHz transmitter with 210 GHz sub-harmonic PLL local oscillator in 130 nm InP HBT," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 2012.

[6] S. Emami, R. Wiser, E. Ali, M. Forbes, M. Gordon, X. Guan, S. Lo, P. McElwee, J. Parker, J. Tani, J. Gilbert, and C. Doan, "A 60GHz CMOS Phased-Array Transceiver Pair for multi-Gb/s Wireless Communications," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2011, pp. 164–166.

[7] H. P. Forstner, H. Knapp, H. Jager, E. Kolmhofer, J. Platz, F. Starzer, M. Treml, A. Schinko, G. Birschkus, J. Bock, K. Aufinger, R. Lachner, T. Meister, H. Schafer, D. Lukashevich, S. Boguth, A. Fischer, F. Reininger, L. Maurer, J. Minichshofer, and

D. Steinbuch, "A 77 GHz 4-Channel Automotive Radar Transceiver in SiGe," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2008, pp. 233–236.

[8] S. Trotta, M. Wintermantel, J. Dixon, U. Moeller, R. Jammers, T. Hauck, A. Samulak, B. Dehlink, K. Shun-Meen, H. Li, A. Ghazinour, Y. Yin, S. Pacheco, R. Reuter, S. Majied, D. Moline, T. Aaron, V. Trivedi, D. Morgan, and J. John, "An RCP Packaged Transceiver Chipset for Automotive LRR and SRR Systems in SiGe BiCMOS Technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 778–794, Mar. 2012.

[9] S. P. Voinigescu, *High-Frequency Integrated Circuits*, 1st ed. Cambridge University Press, 2013.

[10] A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*, 1st ed. Cambridge University Press, 2007.

[11] A. M. Niknejad and H. Hashemi, *mm-Wave Silicon Technology: 60 GHz and Beyond*. Springer, 2008.

[12] I. C.-H. Lai and M. Fujishima, *Design and Modeling of Millimeter-wave CMOS Circuits for Wireless Transceivers*. Springer, 2010.

[13] E. Laskin, P. Chevalier, A. Chantre, B. Sautreuil, and S. P. Voinigescu, "165-GHz Transceiver in SiGe Technology," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 5, pp. 1087–1100, May 2008.

[14] E. Ojefors and U. Pfeiffer, "A 650GHz SiGe Receiver Front-End for Terahertz Imaging Arrays," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2010, pp. 430–431.

[15] E. Ojefors, J. Grzyb, Y. Zhao, B. Heinemann, B. Tillack, and U. Pfeiffer, "A 820GHz SiGe Chipset for Terahertz Active Imaging Applications," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2011, pp. 224–226.

[16] S. Nicolson, A. Tomkins, K. Tang, A. Cathelin, D. Belot, and S. Voinigescu, "A 1.2 V, 140 GHz Receiver with On-Die Antenna in 65 nm CMOS," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2008, pp. 229–232.

[17] M. Seo, B. Jagannathan, J. Pekarik, and M. Rodwell, "A 150 GHz Amplifier With 8 dB Gain and +6 dBm P_{sat} in Digital 65 nm CMOS Using Dummy-Prefilled Microstrip Lines," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 12, pp. 3410–3421, Dec. 2009.

[18] D. Huang, T. LaRocca, M.-C. Chang, L. Samoska, A. Fung, R. Campbell, and M. Andrews, "Terahertz CMOS Frequency Generator Using Linear Superposition Technique," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2730–2738, Dec. 2008.

[19] Y. Tousi, O. Momeni, and E. Afshari, "A 283-to-296 GHz VCO with 0.76 mW Peak Output Power in 65 nm CMOS," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2012, pp. 258–260.

[20] B. Cetinoneri, Y. Atesal, A. Fung, and G. Rebeiz, "W-Band Amplifiers With 6-dB Noise Figure and Milliwatt-Level 170-GHz Doublers in 45-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 692–701, Mar. 2012.

[21] M. Uzunkol, F. Golcuk, B. Cetinoneri, Y. A. Atesal, O. D. Gurbuz, J. M. Edwards, and G. M. Rebeiz, "Millimeter-wave and Terahertz Sources and Imaging Systems Based on 45 nm CMOS Technology," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 2012.

[22] E. Ojefors, U. Pfeiffer, A. Lisauskas, and H. Roskos, "A 0.65 THz Focal-Plane Array in a Quarter-Micron CMOS Process Technology," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 1968–1976, Jul. 2009.

[23] R. Al Hadi, H. Sherry, J. Grzyb, N. Baktash, Y. Zhao, E. Ojefors, A. Kaiser, A. Cathelin, and U. Pfeiffer, "A Broadband 0.6 to 1 THz CMOS Imaging Detector with an Integrated Lens," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 2011.

[24] R. Han, Y. Zhang, Y. Kim, D. Y. Kim, H. Shichijo, E. Afshari, and O. Kenneth, "280GHz and 860GHz Image Sensors Using Schottky-Barrier Diodes in 0.13 μm Digital CMOS," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2012, pp. 254–256.

[25] H. Sherry, J. Grzyb, Y. Zhao, R. Al Hadi, A. Cathelin, A. Kaiser, and U. Pfeiffer, "A 1kpixel CMOS Camera Chip for 25 fps Real-Time Terahertz Imaging Applications," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2012, pp. 252–254.

[26] B. Heinemann, R. Barth, D. Bolze, J. Drews, G. Fischer, A. Fox, O. Fursenko, T. Grabolla, U. Haak, D. Knoll, R. Kurps, M. Lisker, S. Marschmeyer, H. Rucker,

D. Schmidt, J. Schmidt, M. Schubert, B. Tillack, C. Wipf, D. Wolansky, and Y. Yamamoto, "SiGe HBT Technology with f_T/f_{max} of 300 GHz/500 GHz and 2.0 ps CML Gate Delay," in *International Electron Devices Meeting (IEDM) Digest*, Dec. 2010, pp. 30.5.1–30.5.4.

[27] P. Chevalier, F. Pourchon, T. Lacave, G. Avenier, Y. Campidelli, L. Depoyan, G. Troillard, M. Buczko, D. Gloria, D. Celi, C. Gaquiere, and A. Chantre, "A Conventional Double-Polysilicon FSA-SEG Si/SiGe:C HBT reaching 400 GHz f_{max}," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) Digest*, Oct. 2009.

[28] P. VanDerVoorn, M. Agostinelli, S. Choi, G. Curello, H. Deshpande, M. El-Tanani, W. Hafez, U. Jalan, L. Janbay, M. Kang, K. Koh, K. Komeyli, H. Lakdawala, J. Lin, N. Lindert, S. Mudanai, J. Park, K. Phoa, A. Rahman, J. Rizk, L. Rockford, G. Sacks, K. Soumyanath, H. Tashiro, S. Taylor, C. Tsai, H. Xu, J. Xu, L. Yang, I. Young, J. Yeh, J. Yip, P. Bai, and C. Jan, "A 32nm Low Power RF CMOS SOC Technology Featuring High-k/Metal Gate," in *Symposium on VLSI Circuits Digest*, Jun. 2010, pp. 137–138.

[29] C.-H. Jan, M. Agostinelli, H. Deshpande, M. El-Tanani, W. Hafez, U. Jalan, L. Janbay, M. Kang, H. Lakdawala, J. Lin, Y.-L. Lu, S. Mudanai, J. Park, A. Rahman, J. Rizk, W.-K. Shin, K. Soumyanath, H. Tashiro, C. Tsai, P. VanDerVoorn, J.-Y. Yeh, and P. Bai, "RF CMOS Technology Scaling in High-k/Metal gate era for RF SoC (System-on-Chip) Applications," in *International Electron Devices Meeting (IEDM) Digest*, Dec. 2010, pp. 27.2.1–27.2.4.

[30] G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, 2nd ed. Wiley, 2005.

[31] G. D. Vendelin and S.-C. Shin, "Applying f_{max}, f_t, and f_{mag} for Microwave Transistor Designs at Microwave and Millimeter-Wave Frequencies," *IEEE Microwave Magazine*, vol. 8, no. 1, pp. 84–90, Feb. 2007.

[32] F. E. Terman, *Electronic and Radio Engineering*, 4th ed. McGraw-Hill, 1955.

[33] L. C. Peterson, "Equivalent Circuits of Linear Active Four-Terminal Networks," *Bell System Technical Journal*, vol. 27, pp. 593–622, Oct. 1948.

[34] R. Ryder and R. Kircher, "Some Circuit Aspects of the Transistor," *Bell System Technical Journal*, vol. 28, pp. 367–400, Jul. 1949.

[35] R. Wallace and W. Pietenpol, "Some Circuit Properties and Applications of n-p-n Transistors," *Bell System Technical Journal*, vol. 30, pp. 530–563, Jul. 1951.

[36] J. Linvill and L. Schimpf, "The Design of Tetrode Transistor Amplifiers," *Bell System Technical Journal*, vol. 35, pp. 813–840, Jul. 1956.

[37] J. G. Linvill and J. F. Gibbons, *Transistors and Active Circuits*, 1st ed. McGraw-Hill, 1961.

[38] A. Stern, "Stability and Power Gain of Tuned Transistor Amplifiers," *Proceedings of the IRE*, vol. 45, no. 3, pp. 335–343, Mar. 1957.

[39] R. Pritchard, "High-Frequency Power Gain of Junction Transistors," *Proceedings of the IRE*, vol. 43, no. 9, pp. 1075–1085, Sep. 1955.

[40] S. Mason, "Power gain in feedback amplifier," *IRE Transactions on Circuit Theory*, vol. 1, no. 2, pp. 20–25, Jun. 1954.

[41] D. M. Pozar, *Microwave Engineering*, 3rd ed. Wiley, 2004.

[42] H. Fukui, "Available Power Gain, Noise Figure, and Noise Measure of Two-Ports and Their Graphical Representations," *IEEE Transactions on Circuit Theory*, vol. 13, no. 2, pp. 137–142, Jun. 1966.

[43] C. Cheng, "Neutralization and Unilateralization," *IRE Transactions on Circuit Theory*, vol. 2, no. 2, pp. 138–145, Jun. 1955.

[44] R. S. Carson, *High Frequency Amplifiers*, 2nd ed. Wiley, 1982.

[45] R. E. Collin, *Foundations for Microwave Engineering*, 2nd ed. Wiley-IEEE Press, 2000.

[46] P. Chevalier, T. Lacave, E. Canderle, A. Pottrain, Y. Carminati, J. Rosa, F. Pourchon, N. Derrier, G. Avenier, A. Montagné, A. Balteanu, E. Dacquay, I. Sarkas, D. Céli, D. Gloria, C. Gaquière, S. Voinigescu, and A. Chantre, "Scaling of SiGe BiCMOS Technologies for Applications Above 100 GHz," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2012.

[47] A. Stern, C. Aldridge, and W. Chow, "Internal Feedback and Neutralization of Transistor Amplifiers," *Proceedings of the IRE*, vol. 43, no. 7, pp. 838–847, Jul. 1955.

[48] K. Kobayashi, R. Esfandiari, and A. Oki, "A Novel HBT Distributed Amplifier Design Topology Based on Attenuation Compensation Techniques," *IEEE Transactions on Microwave Theory and Techniques*, vol. 42, no. 12, pp. 2583–2589, Dec. 1994.

[49] S. Venkateswaran and A. Boothroyd, "Power Gain and Bandwidth of Tuned Transistor Amplifier Stages," *Proceedings of the IEE - Part B: Electronic and Communication Engineering*, vol. 106, no. 15, pp. 518–529, May 1959.

[50] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 6th ed. Oxford University Press, 2009.

[51] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 2nd ed. Cambridge University Press, 2009.

[52] C. Enz and Y. Cheng, "MOS Transistor Modeling for RF IC Design," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 2, pp. 186–201, Feb. 2000.

[53] G. Avenier, , N. Revil, P. Chevalier, S. Pruvost, J. Bouvier, G. Avenier, G. Troillard, L. Depoyan, M. Buczko, S. Montusclat, A. Margain, S. T. Nicolson, K. H. K. Yau, D. Gloria, A. Chantre, M. Diop, N. Loubet, N. Derrier, C. Leyris, S. Boret, D. Dutartre, and S. P. Voinigescu, "0.13 μm SiGe BiCMOS Technology Fully Dedicated to mm-wave Applications," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2312–2321, Sep. 2009.

[54] S. Narasimha, K. Onishi, H. Nayfeh, A. Waite, M. Weybright, J. Johnson, C. Fonseca, D. Corliss, C. Robinson, M. Crouse, D. Yang, C.-H. Wu, A. Gabor, T. Adam, I. Ahsan, M. Belyansky, L. Black, S. Butt, J. Cheng, A. Chou, G. Costrini, C. Dimitrakopoulos, A. Domenicucci, P. Fisher, A. Frye, S. Gates, S. Greco, S. Grunow, M. Hargrove, J. Holt, S.-J. Jeng, M. Kelling, B. Kim, W. Landers, G. Larosa, D. Lea, M. Lee, X. Liu, N. Lustig, A. McKnight, L. Nicholson, D. Nielsen, K. Nummy, V. Ontalus, C. Ouyang, X. Ouyang, C. Prindle, R. Pal, W. Rausch, D. Restaino, C. Sheraw, J. Sim, A. Simon, T. Standaert, C. Sung, K. Tabakman, C. Tian, R. Van Den Nieuwenhuizen, H. Van Meer, A. Vayshenker, D. Wehella-Gamage, J. Werking, R. Wong, J. Yu, S. Wu, R. Augur, D. Brown, X. Chen, D. Edelstein, A. Grill, M. Khare, Y. Li, S. Luning, J. Norum, S. Sankaran, D. Schepis, R. Wachnik, R. Wise, C. Warm, T. Ivers, and P. Agnello, "High Performance 45-nm SOI Technology with Enhanced Strain, Porous Low-k BEOL, and Immersion Lithography," in *International Electron Devices Meeting (IEDM) Digest*, Dec. 2006.

[55] T. Dickson, K. Yau, T. Chalvatzis, A. Mangan, E. Laskin, R. Beerkens, P. Westergaard, M. Tazlauanu, M.-T. Yang, and S. Voinigescu, "The Invariance of Characteristic Current Densities in Nanoscale MOSFETs and Its Impact on Algorithmic Design Methodologies and Design Porting of Si(Ge) (Bi)CMOS High-Speed Building Blocks," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 8, pp. 1830–1845, Aug. 2006.

[56] M. Khanpour, K. Tang, P. Garcia, and S. Voinigescu, "A Wideband W-Band Receiver Front-End in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 8, pp. 1717–1730, Aug. 2008.

[57] A. Tomkins, P. Garcia, and S. Voinigescu, "A Passive W-Band Imaging Receiver in 65-nm Bulk CMOS," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 10, pp. 1981–1991, Oct. 2010.

[58] L. Zhou, C.-C. Wang, Z. Chen, and P. Heydari, "A W-band CMOS Receiver Chipset for Millimeter-Wave Radiometer Systems," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 2, pp. 378–391, Feb. 2011.

[59] H. Friis, "Noise Figures of Radio Receivers," *Proceedings of the IRE*, vol. 32, no. 7, pp. 419–422, Jul. 1944.

[60] H. Rothe and W. Dahlke, "Theory of Noisy Fourpoles," *Proceedings of the IRE*, vol. 44, no. 6, pp. 811–818, Jun. 1956.

[61] H. Haus and R. Adler, *Circuit theory of linear noisy networks*. Technology Press of Massachusetts Institute of Technology, 1959. [Online]. Available: http://archive.org/details/circuittheoryofl00haus

[62] H. Hillbrand and P. Russer, "An Efficient Method for Computer Aided Noise Analysis of Linear Amplifier Networks," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 4, pp. 235–238, Apr. 1976.

[63] H. Fukui, "The Noise Performance of Microwave Transistors," *IEEE Transactions on Electron Devices*, vol. 13, no. 3, pp. 329–341, Mar. 1966.

[64] S. Voinigescu, M. Maliepaard, J. Showell, G. Babcock, D. Marchesan, M. Schroter, P. Schvan, and D. Harame, "A Scalable High-Frequency Noise Model for Bipolar Transistors with Application to Optimal Transistor Sizing for Low-Noise Amplifier Design," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 9, pp. 1430–1439, Sep. 1997.

[65] G. Niu, J. Cressler, S. Zhang, W. Ansley, C. Webster, and D. Harame, "A Unified Approach to RF and Microwave Noise Parameter Modeling in Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 48, no. 11, pp. 2568–2574, Nov. 2001.

[66] H. Fukui, "Optimal Noise Figure of Microwave GaAs MESFETs," *IEEE Transactions on Electron Devices*, vol. 26, no. 7, pp. 1032 – 1037, Jul. 1979.

[67] M. Pospieszalski, "Modeling of Noise Parameters of MESFETs and MODFETs and Their Frequency and Temperature Dependence," *IEEE Transactions on Electron Devices*, vol. 37, no. 9, pp. 1340–1350, Sep. 1989.

[68] A. Cappy, "Noise Modeling and Measurement Techniques [HEMTs]," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, no. 1, pp. 1–10, Jan. 1988.

[69] A. Scholten, L. Tiemeijer, R. van Langevelde, R. Havens, A. Zegers-van Duijnhoven, and V. Venezia, "Noise Modeling for RF CMOS Circuit Simulation," *IEEE Transactions on Electron Devices*, vol. 50, no. 3, pp. 618–632, Mar. 2003.

[70] H. Haus and R. Adler, "Optimum Noise Performance of Linear Amplifiers," *Proceedings of the IRE*, vol. 46, no. 8, pp. 1517–1533, Aug. 1958.

[71] S. Weinreb, D. Fenstermacher, and R. Harris, "Ultra-Low-Noise 1.2- to 1.7-GHz Cooled GaAsFET Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 30, no. 6, pp. 849–853, Jun. 1982.

[72] K. Hartmann and M. Strutt, "Changes of the Four Noise Parameters due to General Changes of Linear Two-Port Circuits," *IEEE Transactions on Electron Devices*, vol. 20, no. 10, pp. 874–877, Oct. 1973.

[73] B. Razavi, *RF Microelectronics*, 1st ed. Prentice Hall, 1997.

[74] R. Tucker, "Low-noise design of microwave transistor amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 23, no. 8, pp. 697–700, Aug. 1975.

[75] J. Bardin and S. Weinreb, "Experimental Cryogenic Modeling and Noise of SiGe HBTs," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 2008, pp. 459–462.

[76] M. Gordon and S. Voinigescu, "An Inductor-Based 52-GHz 0.18 μm SiGe HBT Cascode LNA with 22 dB Gain," in *European Solid-State Circuits Conference (ESSCIRC) Proceedings*, Sep. 2004, pp. 287–290.

[77] I. Bahl, "High-performance inductors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 4, pp. 654–664, Apr. 2001.

[78] J. Long and M. Copeland, "The Modeling, Characterization, and Design of Monolithic Inductors for Silicon RFICs," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 3, pp. 357–369, Mar. 1997.

[79] T. Dickson and S. Voinigescu, "Low-Power Circuits for a 2.5-V, 10.7-to-86-Gb/s Serial Transmitter in 130-nm SiGe BiCMOS," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 10, pp. 2077–2085, Oct. 2007.

[80] J. Burghartz and B. Rejaei, "On the Design of RF Spiral Inductors on Silicon," *IEEE Transactions on Electron Devices*, vol. 50, no. 3, pp. 718–729, Mar. 2003.

[81] A. Reyes, S. El-Ghazaly, S. Dorn, M. Dydyk, D. Schroder, and H. Patterson, "Coplanar Waveguides and Microwave Inductors on Silicon Substrates," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 9, pp. 2016–2022, Sep. 1995.

[82] T. Dickson, M.-A. LaCroix, S. Boret, D. Gloria, R. Beerkens, and S. Voinigescu, "30-100-GHz Inductors and Transformers for Millimeter-wave (Bi)CMOS Integrated Circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 1, pp. 123–133, Jan. 2005.

[83] K. Yau, I. Sarkas, A. Tomkins, P. Chevalier, and S. Voinigescu, "On-Wafer S-Parameter De-Embedding of Silicon Active and Passive Devices up to 170 GHz," in *IEEE International Microwave Symposium (IMS) Digest*, May 2010, pp. 600–603.

[84] M. Moller, "Challenges in the Cell-Based Design of Very-High-Speed SiGe-Bipolar ICs at 100 Gb/s," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 9, pp. 1877–1888, Sep. 2008.

[85] C. Paul, "Effectiveness of multiple decoupling capacitors," *IEEE Transactions on Electromagnetic Compatibility*, vol. 34, no. 2, pp. 130–133, May 1992.

[86] A. Mangan, S. Voinigescu, M.-T. Yang, and M. Tazlauanu, "De-embedding Transmission Line Measurements for Accurate Modeling of IC Designs," *IEEE Transactions on Electron Devices*, vol. 53, no. 2, pp. 235–241, Feb. 2006.

[87] S. Nicolson, P. Chevalier, B. Sautreuil, and S. Voinigescu, "Single-Chip W-band SiGe HBT Transceivers and Receivers for Doppler Radar and Millimeter-Wave Imaging," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 10, pp. 2206–2217, Oct. 2008.

[88] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Artech House, 2006.

[89] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. Cambridge University Press, 2003.

[90] S. Pornpromlikit, "CMOS RF Power Amplifier Design Approaches for Wireless Communication," Ph.D. dissertation, University of California, San Diego, 2010.

[91] A. Grebennikov, N. O. Sokal, and M. J. Franco, *Switchmode RF and Microwave Power Amplifiers*, 2nd ed. Academic Press, 2012.

[92] D. Chowdhury, P. Reynaert, and A. Niknejad, "A 60 GHz 1 V +12.3 dBm Transformer-Coupled Wideband PA in 90 nm CMOS," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2008, pp. 560–635.

[93] I. Aoki, S. Kee, R. Magoon, R. Aparicio, F. Bohn, J. Zachan, G. Hatcher, D. McClymont, and A. Hajimiri, "A Fully-Integrated Quad-Band GSM/GPRS CMOS Power Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2747–2758, Dec. 2008.

[94] A. Shirvani, D. Su, and B. Wooley, "A CMOS RF Power Amplifier with Parallel Amplification for Efficient Power Control," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 6, pp. 684–693, Jun. 2002.

[95] Y. Atesal, B. Cetinoneri, M. Chang, R. Alhalabi, and G. Rebeiz, "Millimeter-Wave Wafer-Scale Silicon BiCMOS Power Amplifiers Using Free-Space Power Combining," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 4, pp. 954–965, Apr. 2011.

[96] A. Afsahi, A. Behzad, and L. Larson, "A 65 nm CMOS 2.4 GHz 31.5 dBm power amplifier with a distributed LC power-combining network and improved linearization for WLAN applications," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2010, pp. 452–453.

[97] T. Sowlati and D. Leenaerts, "A 2.4-GHz 0.18-μm CMOS self-biased cascode power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 8, pp. 1318–1324, Aug. 2003.

[98] M. Shrivastava, M. Baghini, H. Gossner, and V. Rao, "Part I: Mixed-Signal Performance of Various High-Voltage Drain-Extended MOS Devices," *IEEE Transactions on Electron Devices*, vol. 57, no. 2, pp. 448–457, Feb. 2010.

[99] A. Mai and H. Rucker, "Drain-Extended MOS Transistors Capable for Operation at 10 V and at Radio Frequencies," in *European Solid-State Device Research Conference (ESSDERC) Proceedings*, Sep. 2010, pp. 110–113.

[100] D. Calvillo-Cortes, M. Acar, M. van der Heijden, M. Apostolidou, L. de Vreede, D. Leenaerts, and J. Sonsky, "A 65nm CMOS Pulse-Width-Controlled Driver with 8 Vpp Output Voltage for Switch-Mode RF PAs up to 3.6 GHz," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2011, pp. 58–60.

[101] K. J. Dean, *Transistors, Theory and Circuitry.* McGraw-Hill, 1964.

[102] M. Shifrin, Y. Ayasli, and P. Katzin, "A New Power Amplifier Topology with Series Biasing and Power Combining of Transistors," in *Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest*, Jun. 1992, pp. 39–41.

[103] S. Pornpromlikit, J. Jeong, C. Presti, A. Scuderi, and P. Asbeck, "A Watt-Level Stacked-FET Linear Power Amplifier in Silicon-on-Insulator CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 1, pp. 57–64, Jan. 2010.

[104] A. Ezzeddine and H. Huang, "The High Voltage/High Power FET (HiVP)," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2003, pp. 215–218.

[105] S. Pornpromlikit, H. Dabag, B. Hanafi, J. Kim, L. Larson, J. Buckwalter, and P. Asbeck, "A Q-Band Amplifier Implemented with Stacked 45-nm CMOS FETs," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2011.

[106] T.-P. Hung, J. Rode, L. Larson, and P. Asbeck, "Design of H-Bridge Class-D Power Amplifiers for Digital Pulse Modulation Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2845 –2855, Dec. 2007.

[107] A. Karanicolas, "A 2.7-V 900-MHz CMOS LNA and Mixer," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 12, pp. 1939–1944, Dec. 1996.

[108] B. Serneels, T. Piessens, M. Stepert, and W. Dehaene, "A High-Voltage Output Driver in a Standard 2.5 V 0.25 μm CMOS Technology," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2004.

[109] B. Serneels, T. Piessens, M. Steyaert, and W. Dehaene, "A High-Voltage Output Driver in a 2.5-V 0.25-μm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 3, pp. 576–583, Mar. 2005.

[110] S.-A. El-Hamamsy, "Design of High-Efficiency RF Class-D power Amplifier," *IEEE Transactions on Power Electronics*, vol. 9, no. 3, pp. 297–308, May 1994.

[111] W. Chudobiak and D. Page, "Frequency and Power Limitations of Class-D Transistor Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 4, no. 1, pp. 25–37, Feb. 1969.

[112] S. Ben-Yaakov and M. Evzelman, "Generic and unified model of switched capacitor converters," in *IEEE Energy Conversion Congress and Exposition (ECCE)*, Sep. 2009, pp. 3501–3508.

[113] M. Seeman and S. Sanders, "Analysis and Optimization of Switched-Capacitor DC-DC Converters," *IEEE Transactions on Power Electronics*, vol. 23, no. 2, pp. 841–851, Mar. 2008.

[114] W.-S. Kim, X. Li, and M. Ismail, "A 2.4 GHz CMOS Low Noise Amplifier Using an Inter-stage Matching Inductor," in *Midwest Symposium on Circuits and Systems (MWSCAS)*, vol. 2, 1999, pp. 1040–1043.

[115] T. Suzuki, Y. Nakasha, T. Takahashi, K. Makiyama, K. Imanishi, T. Hirose, and Y. Watanabe, "A 90Gb/s 2:1 Multiplexer IC in InP-Based HEMT Technology," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2002, pp. 192–193.

[116] A. Agah, H. Dabag, B. Hanafi, P. Asbeck, L. Larson, and J. Buckwalter, "A 34% PAE, 18.6 dBm 42-45 GHz Stacked Power Amplifier in 45 nm SOI CMOS," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2012, pp. 57–60.

[117] H. Xu, Y. Palaskas, A. Ravi, M. Sajadieh, M. El-Tanani, and K. Soumyanath, "A Flip-Chip-Packaged 25.3 dBm Class-D Outphasing Power Amplifier in 32 nm CMOS for WLAN Application," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 7, pp. 1596–1605, Jul. 2011.

[118] C. Cao, H. Xu, Y. Su, and K. O, "An 18-GHz, 10.9-dBm Fully-Integrated Power Amplifier with 23.5% PAE in 130-nm CMOS," in *European Solid-State Circuits Conference (ESSCIRC) Proceedings*, Sep. 2005, pp. 137–140.

[119] W.-H. Hung, H.-S. Chen, S.-H. Chou, and L.-H. Lu, "An 18 dBm Transmitter Frontend with 29% PAE for 24GHz FMCW Radar Applications," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2012, pp. 137–140.

[120] J. Wells, "Faster than Fiber: The Future of multi-Gb/s Wireless," *IEEE Microwave Magazine*, vol. 10, no. 3, pp. 104–112, May 2009.

[121] D. Lockie and D. Peck, "High-Data-Rate Millimeter-Wave Radios," *IEEE Microwave Magazine*, vol. 10, no. 5, pp. 75–83, Aug. 2009.

[122] A. Natarajan, S. Reynolds, M.-D. Tsai, S. Nicolson, J.-H. Zhan, D. G. Kam, D. Liu, Y.-L. Huang, A. Valdes-Garcia, and B. Floyd, "A Fully-Integrated 16-Element Phased-Array Receiver in SiGe BiCMOS for 60-GHz Communications," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 5, pp. 1059–1075, May 2011.

[123] A. Valdes-Garcia, S. Nicolson, J.-W. Lai, A. Natarajan, P.-Y. Chen, S. Reynolds, J.-H. C. Zhan, D. Kam, D. Liu, and B. Floyd, "A Fully Integrated 16-Element Phased-Array Transmitter in SiGe BiCMOS for 60-GHz Communications," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2757–2773, Dec. 2010.

[124] A. Tomkins, R. Aroca, T. Yamamoto, S. Nicolson, Y. Doi, and S. Voinigescu, "A Zero-IF 60GHz 65nm CMOS Transceiver With Direct BPSK Modulation Demonstrating up to 6Gb/s Data Rates Over a 2 m Wireless Link," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 8, pp. 2085–2099, Aug. 2009.

[125] J. Lee, Y. Chen, and Y. Huang, "A Low-Power Low-Cost Fully-Integrated 60-GHz Transceiver System With OOK Modulation and On-Board Antenna Assembly," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 2, pp. 264–275, Feb. 2010.

[126] W. Shin, M. Uzunkol, and G. Rebeiz, "Ultra Low Power 60 GHz ASK SiGe Receiver with 3-6 Gb/s Capabilities," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2009.

[127] C. Marcu, D. Chowdhury, C. Thakkar, J.-D. Park, L.-K. Kong, M. Tabesh, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. Zamani, E. Alon, and A. Niknejad, "A 90 nm CMOS Low-Power 60 GHz Transceiver With Integrated Baseband Circuitry," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 12, pp. 3434–3447, Dec. 2009.

[128] V. Dyadyuk, J. Bunton, J. Pathikulangara, R. Kendall, O. Sevimli, L. Stokes, and D. Abbott, "A Multigigabit Millimeter-Wave Communication System With Improved Spectral Efficiency," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2813–2821, Dec. 2007.

[129] O. Katz, R. Ben-Yishay, R. Carmon, B. Sheinman, F. Szenher, D. Papae, and D. Elad, "High-Power High-Linearity SiGe based E-band Transceiver Chipset for Broadband Communication," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2012, pp. 115–118.

[130] S.-J. Huang, Y.-C. Yeh, H. Wang, P.-N. Chen, and J. Lee, "W -Band BPSK and QPSK Transceivers With Costas-Loop Carrier Recovery in 65-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 12, pp. 3033–3046, Dec. 2011.

[131] T. Kosugi, A. Hirata, T. Nagatsuma, and Y. Kado, "mm-wave Long-Range Wireless Systems," *IEEE Microwave Magazine*, vol. 10, no. 2, pp. 68–76, Apr. 2009.

[132] E. Laskin and S. Voinigescu, "A 60 mW per Lane, 4×23-Gb/s 2^7-1 PRBS Generator," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 10, pp. 2198–2208, Oct. 2006.

[133] D. M. Pozar, *Microwave and RF Design of Wireless Systems*, 1st ed. Wiley, 2000.

[134] B. Razavi, "Design Considerations for Direct-Conversion Receivers," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, no. 6, pp. 428–435, Jun. 1997.

[135] T. Dickson, R. Beerkens, and S. Voinigescu, "A 2.5-V, 40-Gb/s Decision Circuit using SiGe BiCMOS Logic," in *Symposium on VLSI Circuits Digest*, Jun. 2004, pp. 206–209.

[136] H.-M. Rein and M. Moller, "Design Considerations for Very-High-Speed Si-Bipolar IC's Operating up to 50 Gb/s," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 8, pp. 1076–1090, Aug. 1996.

[137] R. Frye, S. Kapur, and R. Melville, "A 2-GHz Quadrature Hybrid Implemented in CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 3, pp. 550–555, Mar. 2003.

[138] I. Sarkas, M. Khanpour, A. Tomkins, P. Chevalier, P. Garcia, and S. Voinigescu, "W-band 65-nm CMOS and SiGe BiCMOS transmitter and receiver with lumped I-Q phase shifters," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2009, pp. 441–444.

[139] S. Nicolson, "Low-Voltage and Low-Power Silicon Transceivers and Receivers for W-Band and D-Band Applications," Ph.D. dissertation, University of Toronto, 2008.

[140] "Noise Figure Measurement Accuracy The Y-Factor Method," Agilent Technologies, Tech. Rep. 57-2.

[141] C. Thakkar, L. Kong, K. Jung, A. Frappe, and E. Alon, "A 10 Gb/s 45 mW Adaptive 60 GHz Baseband in 65 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 4, pp. 952–968, Apr. 2012.

[142] I. Sarkas and S. Voinigescu, "A 1.8 VSiGe BiCMOS Cable Equalizer with 40-dB Peaking Control up to 60 GHz," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2012.

[143] E. Laskin, A. Tomkins, A. Balteanu, I. Sarkas, and S. Voinigescu, "A 60-GHz RF IQ DAC Transceiver with On-Die at-speed Loopback," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2011.

[144] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO Radar: An Idea Whose Time Has Come," in *Proceedings of the IEEE Radar Conference (Radarcon)*, Apr. 2004, pp. 71–78.

[145] A. Schicht, K. Huber, A. Ziroff, M. Willsch, and L.-P. Schmidt, "Absolute Phase-Based Distance Measurement for Industrial Monitoring Systems," *IEEE Sensors Journal*, vol. 9, no. 9, pp. 1007–1013, Sep. 2009.

[146] M.-C. Amann, T. Bosch, M. Lescure, R. Myllylä, and M. Rioux, "Laser Ranging: a Critical Review of Usual Techniques for Distance Measurement," *Optical Engineering*, vol. 40, no. 1, pp. 10–19, 2001.

[147] G. Kompa, V. Gorfinkel, J. Sola, A. Stolze, W. Vogt, and F. Volpe, "Powerful Picosecond-Pulsed Laser Radar with Micrometer Ranging Resolution," in *European Microwave Conference Digest*, Sep. 1996, pp. 147–152.

[148] G. Woods, D. Maskell, and M. Mahoney, "A High Accuracy Microwave Ranging System for Industrial Applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 42, no. 4, pp. 812–816, Aug. 1993.

[149] H. Zhang, L. Li, and K. Wu, "Software-Defined Six-Port Radar Technique for Precision Range Measurements," *IEEE Sensors Journal*, vol. 8, no. 10, pp. 1745–1751, Oct. 2008.

[150] A. Stelzer, C. Diskus, K. Lubke, and H. Thim, "A microwave Position Sensor with Submillimeter Accuracy," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 12, pp. 2621–2624, Dec. 1999.

[151] S. Kim and C. Nguyen, "On the Development of a Multifunction Millimeter-Wave Sensor for Displacement Sensing and Low-Velocity Measurement," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 11, pp. 2503–2512, Nov. 2004.

[152] Y. Venot and W. Wiesbeck, "76.5 GHz Radar Sensor for Contact-free Distance Measurement with Micrometer Accuracy," in *IEEE Sensors Journal*, Oct. 2003, pp. 216–221.

[153] K. Haddadi, M. Wang, D. Glay, and T. Lasri, "A 60 GHz Six-Port Distance Measurement System With Sub-Millimeter Accuracy," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 10, pp. 644–646, Oct. 2009.

[154] S. Stuchly, M. Hamid, and A. Andres, "Microwave Surface Level Monitor," *IEEE Transactions on Industrial Electronics and Control Instrumentation*, vol. 18, no. 3, pp. 85–92, Aug. 1971.

[155] M. A. Richards, *Fundamentals of Radar Signal Processing*, 1st ed. McGraw-Hill, 2005.

[156] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed. Wiley, 2005.

[157] R. Feger, C. Wagner, S. Schuster, S. Scheiblhofer, H. Jager, and A. Stelzer, "A 77-GHz FMCW MIMO Radar Based on an SiGe Single-Chip Transceiver," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 5, pp. 1020–1035, May 2009.

[158] M. Skolnik, *Introduction to Radar Systems*, 3rd ed. McGraw-Hill, 2002.

[159] I. Gresham, A. Jenkins, R. Egri, C. Eswarappa, N. Kinayman, N. Jain, R. Anderson, F. Kolak, R. Wohlert, S. Bawell, J. Bennett, and J.-P. Lanteri, "Ultra-Wideband Radar Sensors for Short-range Vehicular Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 9, pp. 2105–2122, Sep. 2004.

[160] A. Carr, L. Cuthbert, and A. Olver, "Digital Signal Processing for Target Detection FMCW radar," *IEE Proceedings on Communications, Radar and Signal Processing*, vol. 128, no. 5, pp. 331–336, Oct. 1981.

[161] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[162] J. A. Scheer and J. L. Kurtz, *Coherent Radar Performance Estimation*. Artech House, 1993.

[163] D. E. Barrick, "FM/CW Radar Signals and Digital Processing," NOAA, Tech. Rep. ERL 283-WPL 26, 1973.

[164] S. Scheiblhofer, S. Schuster, and A. Stelzer, "Signal Model and Linearization for Nonlinear Chirps in FMCW Radar SAW-ID Tag Request," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1477–1483, Jun. 2006.

[165] S. Piper, "Homodyne FMCW Radar Range Resolution Effects with Sinusoidal Nonlinearities in the Frequency Sweep," in *Proceedings of the IEEE Radar Conference (Radarcon)*, May 1995, pp. 563–567.

[166] S. Scheiblhofer, S. Schuster, and A. Stelzer, "Effects of Systematic FMCW Radar Sweep Nonlinearity on Bias and Variance of Target Range Estimation," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 2006, pp. 1418–1421.

[167] K. Iizuka and A. Freundorfer, "Detection of Nonmetallic Buried Objects by a Step Frequency Radar," *Proceedings of the IEEE*, vol. 71, no. 2, pp. 276–279, Feb. 1983.

[168] K. Iizuka, A. P. Freundorfer, K. H. Wu, H. Mori, H. Ogura, and V.-K. Nguyen, "Step-Frequency Radar," *AIP Journal of Applied Physics*, vol. 56, no. 9, pp. 2572–2583, 1984.

[169] L. Robinson, W. Weir, and L. Young, "An RF Time-Domain Reflectometer Not in Real Time," *IEEE Transactions on Microwave Theory and Techniques*, vol. 20, no. 12, pp. 855–857, Dec. 1972.

[170] E. J. Candès and C. Fernandez-Granday, "Towards a Mathematical Theory of Super-Resolution," *Preprint*, 2012. [Online]. Available: http://arxiv.org/abs/1203.5871

[171] I. Sarkas, "Step Frequency Radar Using Compressed Sensing," University of Toronto, Tech. Rep., 2010, JEB1433. [Online]. Available: http://wiki.math.toronto.edu/TorontoMathWiki/images/b/bd/2010_JEB1433_Sarkas_SFR_CS_Report.pdf

[172] T. Mitomo, N. Ono, H. Hoshino, Y. Yoshihara, O. Watanabe, and I. Seto, "A 77 GHz 90 nm CMOS Transceiver for FMCW Radar Applications," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 928–937, Apr. 2010.

[173] S. Nicolson, K. Yau, S. Pruvost, V. Danelon, P. Chevalier, P. Garcia, A. Chantre, B. Sautreuil, and S. Voinigescu, "A Low-Voltage SiGe BiCMOS 77-GHz Automotive Radar Chipset," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 5, pp. 1092–1104, May 2008.

[174] B. Dehlink, H.-D. Wohlmuth, H.-R. Forstner, H. Knapp, S. Trotta, K. Aufinger, T. Meister, J. Bock, and A. Scholtz, "A Highly Linear SiGe Double-Balanced Mixer for 77 GHz Automotive Radar Applications," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2006.

[175] A. Pavio and S. Sutton, "A Microstrip Re-entrant Mode Quadrature Coupler for Hybrid and Monolithic Circuit Applications," in *IEEE International Microwave Symposium (IMS) Digest*, May 1990, pp. 573–576.

[176] S. March, "Phase Velocity Compensation in Parallel-Coupled Microstrip," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 1982, pp. 410–412.

[177] U. Pfeiffer, E. Ojefors, and Y. Zhao, "A SiGe Quadrature Transmitter and Receiver Chipset for Emerging High-Frequency Applications at 160 GHz," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2010, pp. 416–417.

[178] E. Laskin, P. Chevalier, B. Sautreuil, and S. P. Voinigescu, "A 140-GHz double-sideband transceiver with amplitude and frequency modulation operating over a few meters," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) Digest*, Oct. 2009, pp. 178–181.

[179] K. Schmalz, W. Winkler, J. Borngraber, W. Debski, B. Heinemann, and J. Scheytt, "A Subharmonic Receiver in SiGe Technology for 122 GHz Sensor Applications," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1644–1656, Sep. 2010.

[180] I. Sarkas, E. Laskin, J. Hasch, P. Chevalier, and S. Voinigescu, "Second Generation Transceivers for D-band Radar and Data Communication Applications," in *IEEE International Microwave Symposium (IMS) Digest*, May 2010, pp. 1328–1331.

[181] H. So, "Time-Delay Estimation for Sinusoidal Signals," *IEE Proceedings Radar, Sonar and Navigation*, vol. 148, no. 6, pp. 318–324, Dec. 2001.

[182] ——, "A Comparative Study of Two Discrete-Time Phase Delay Estimators," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, pp. 2501–2504, Dec. 2005.

[183] C. Enz and G. Temes, "Circuit Techniques for Reducing the Effects of Op-Amp Imperfections: Autozeroing, Correlated Double Sampling, and Chopper Stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584–1614, Nov. 1996.

[184] D. Noon, D. Longstaff, and G. Stickley, "Correction of I/Q Errors in Homodyne Step Frequency Radar Refocuses Range Profiles," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Digest*, vol. 2, Apr. 1994, pp. 369–372.

[185] J. Budge, M.C. and M. Burt, "Range correlation effects on phase and amplitude noise," in *IEEE Southeastcon Proceedings*, Apr. 1993.

[186] A. Stove, "Linear FMCW Radar Techniques," *IEE Proceedings on Radar and Signal Processing*, vol. 139, no. 5, pp. 343–350, Oct. 1992.

[187] A. Rumiantsev and N. Ridler, "VNA Calibration," *IEEE Microwave Magazine*, vol. 9, no. 3, pp. 86–99, Jun. 2008.

[188] B. Brown. Solving the one-port calibration equations. Agilent Technologies. [Online]. Available: http://na.tm.agilent.com/vnahelp/tip20.html

[189] M. Zhao, J. Shea, S. Hagness, and D. van der Weide, "Calibrated Free-Space Microwave Measurements With an Ultrawideband Reflectometer-Antenna System," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 12, pp. 675–677, Dec. 2006.

[190] C. Alkemade, W. Snelleman, G. Boutilier, B. Pollard, J. Winefordner, T. Chester, and N. Omenetto, "A Review and Tutorial Discussion of Noise and Signal-to-Noise Ratios in Analytical Spectrometry-I. Fundamental Principles of Signal-to-Noise Ratios," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 33, no. 8, pp. 383–399, 1978.

[191] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Prentice Hall, 1999.

[192] M. A. Caloyannides, "Microcycle Spectral Estimates of 1/f Noise in Semiconductors," *AIP Journal of Applied Physics*, vol. 45, no. 1, pp. 307–316, 1974.

[193] S. Mandal, S. Arfin, and R. Sarpeshkar, "Sub-μHz MOSFET 1/f Noise Measurements," *Electronics Letters*, vol. 45, no. 1, pp. 81–82, 2009.

[194] I. Sarkas, M. Khanpour, A. Tomkins, P. Chevalier, P. Garcia, and S. Voinigescu, "W-band 65-nm CMOS and SiGe BiCMOS Transmitter and Receiver with Lumped I-Q Phase Shifters," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2009, pp. 441–444.

[195] L. Dauphinee, M. Copeland, and P. Schvan, "A balanced 1.5 GHz voltage controlled oscillator with an integrated LC resonator," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 1997, pp. 390–391.

[196] S. Voinigescu, D. Marchesan, and M. Copeland, "A family of monolithic inductor-varactor SiGe-HBT VCOs for 20 GHz to 30 GHz LMDS and fiber-optic receiver applications," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, 2000, pp. 173–177.

[197] H. Li and H.-M. Rein, "Millimeter-Wave VCOs with Wide Tuning Range and Low Phase Noise, Fully Integrated in a SiGe Bipolar Production Technology," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 2, pp. 184–191, Feb. 2003.

[198] S. Nicolson, K. Yau, P. Chevalier, A. Chantre, B. Sautreuil, K. Tang, and S. Voinigescu, "Design and Scaling of W-Band SiGe BiCMOS VCOs," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 9, pp. 1821–1833, Sep. 2007.

[199] C. Lee, T. Yao, A. Mangan, K. Yau, M. Copeland, and S. Voinigescu, "SiGe BiCMOS 65-GHz BPSK transmitter and 30 to 122 GHz LC-varactor VCOs with up to 21% tuning range," in *Compound Semiconductor Integrated Circuits Symposium (CSICS) Digest*, Oct. 2004, pp. 179–182.

[200] J. Rogers and C. Plett, *Radio Frequency Integrated Circuit Design*, 1st ed. Artech House, 2003.

[201] S. Trotta, B. Dehlink, A. Ghazinour, D. Morgan, and J. John, "A 77 GHz 3.3 V 4-Channel Transceiver in SiGe BiCMOS Technology," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) Digest*, Oct. 2009, pp. 186–189.

[202] J. May and G. Rebeiz, "A 40-50–GHz SiGe 1:8 Differential Power Divider Using Shielded Broadside-Coupled Striplines," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 7, pp. 1575–1581, Jul. 2008.

[203] R. Meyer, "Low-power Monolithic RF Peak Detector Analysis," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 1, pp. 65–67, Jan. 1995.

[204] J. Hasch, U. Wostradowski, S. Gaier, and T. Hansen, "77 GHz Radar Transceiver with Dual Integrated Antenna Elements," in *German Microwave Conference*, Mar. 2010, pp. 280–283.

[205] H. Knapp, A. Schinko, E. K. ans Stefan Matzinger ans Georg Strasser, R. Lachner, L. Maurer, and J. Minichshofer, "Three-Channel 77 GHz Automotive Radar Transmitter in Plastic Package," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC) Digest*, Jun. 2012, pp. 119–122.

[206] Y.-C. Ou and G. M. Rebeiz, "On-Chip Slot-Ring and High-Gain Horn Antennas for Millimeter-Wave Wafer-Scale Silicon Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 8, pp. 1963–1972, Aug. 2011.

[207] M. Jahn, H. Knapp, and A. Stelzer, "A 122-GHz SiGe-Based Signal-Generation Chip Employing a Fundamental-Wave Oscillator With Capacitive Feedback Frequency-Enhancement," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 9, pp. 2009–2020, Sep. 2011.

[208] E. Ojefors, F. Pourchon, P. Chevalier, and U. Pfeiffer, "A 160-GHz Low-Noise Down-conversion Receiver Front-end in a SiGe HBT Technology," *Inernational Journal of Microwave and Wireless Technologies*, vol. 3, no. 3, pp. 347–353, Mar. 2011.

[209] M. G. Girma, J. Hasch, I. Sarkas, S. Voinigescu, and T. Zwick, "122 GHz Radar Sensor based on a Monostatic SiGe-BiCMOS IC with an On-Chip Antenna," in *European Microwave Conference Digest*, Oct. 2012.

[210] D. Kissinger, B. Laemmle, L. Maurer, and R. Weigel, "Integrated Test for Silicon Front Ends," *IEEE Microwave Magazine*, vol. 11, no. 3, pp. 87–94, May 2010.

[211] O. Inac, D. Shin, and G. Rebeiz, "A Phased Array RFIC With Built-In Self-Test Capabilities," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 1, pp. 139–148, Jan. 2012.

[212] D. Kissinger, H. Knapp, L. Maurer, and R. Weigel, "A 77-GHz Down-Conversion Mixer Architecture with Built-in Test Capability in SiGe Technology," in *Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) Digest*, Oct. 2010, pp. 200–203.

[213] M. Jahn, R. Feger, C. Pfeffer, T. F. Meister, and A. Stelzer, "A SiGe-Based 140-GHz Four-Channel Radar Sensor With Digital Beamforming Capability," in *IEEE International Microwave Symposium (IMS) Digest*, Jun. 2012.

[214] I. Sarkas, E. Laskin, A. Tomkins, A. Balteanu, E. Dacquay, L. Tarnow, S. P. Voinigescu, J. Hasch, P. Chevalier, and B. Sautreuil, "Silicon-based Radar and Imaging Sensors Operating above 120 GHz," in *MIKON Digest*, May 2012, pp. 91–96.

[215] N. Deferm and P. Reynaert, "A 120 GHz 10 Gb/s phase-modulating transmitter in 65 nm LP CMOS," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2011, pp. 290–292.

[216] M. Abbasi, S. Gunnarsson, N. Wadefalk, R. Kozhuharov, J. Svedin, S. Cherednichenko, I. Angelov, I. Kallfass, A. Leuther, and H. Zirath, "Single-Chip 220-GHz Active Heterodyne Receiver and Transmitter MMICs With On-Chip Integrated Antenna," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 2, pp. 466–478, Feb. 2011.

[217] S. Koch, M. Guthoerl, I. Kallfass, A. Leuther, and S. Saito, "A 120 - 145 GHz Heterodyne Receiver Chipset Utilizing the 140 GHz Atmospheric Window for Passive Millimeter-Wave Imaging Applications," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 10, pp. 1961–1967, Oct. 2010.

[218] B. Serneels, E. Geukens, B. De Muer, and T. Piessens, "A 1.5 W 10 V-output Class-D Amplifier Using a Boosted Supply from a Single 3.3 V Input in Standard 1.8 V/3.3 V 0.18 μm CMOS," in *IEEE International Solid-State Circicuits Conference (ISSCC) Digest*, Feb. 2012, pp. 94–96.

[219] T. Chalvatzis, E. Gagnon, M. Repeta, and S. Voinigescu, "A Low-Noise 40-GS/s Continuous-Time Bandpass $\Delta\Sigma$ ADC Centered at 2 GHz for Direct Sampling Receivers," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 1065–1075, May 2007.

[220] P. Ostrovskyy, H. Gustat, M. Ortmanns, and J. Scheytt, "A 5-Gb/s $2.1 - 2.2$-GHz Bandpass $\Delta\Sigma$ Modulator for Switch-Mode Power Amplifier," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 8, pp. 2524–2531, Aug. 2012.